

PROBLEMS IN CLOZE TESTING

by Roger K. Williams

Many authors and teachers have hailed cloze tests as the easiest and best method of testing the reading skill. No doubt this type of test *can* be the easiest. A teacher chooses a passage from a book, types it on a page deleting every sixth, seventh, or *n*th word, and requires his students to fill in the correct answers. He may give credit for only those answers which match the original, or he may give partial or full credit for "acceptable" answers, those which fit the slot semantically.

Whether cloze tests are the best method of testing the reading skill is another matter. Some cloze tests can be very good. Others may contain flaws which affect the validity and reliability of the tests. Teachers who are considering using this method should be aware of several problems. In avoiding these problems, they may find that good cloze tests are about as easy to construct as good comprehension (paragraph and questions) tests.

The first problem is that cloze tests only partially parallel the reading process. Reading is primarily the decoding of written symbols into meaning. Filling in blanks requires encoding or production. In reading regular prose, both native and non-native readers, of course, often encounter unfamiliar words or phrases, and must employ inferential encoding procedures to arrive at a meaning, but if there are too many unfamiliar words or phrases encoding or filling in blanks correctly becomes quite difficult and frustrating. Therefore, a cloze test may approximate a situation in which students are reading material far beyond their level more than a situation in which they are reading material on their level.

Because cloze tests require production, scoring can become a problem if the teacher chooses to give credit or partial credit for answers which, though different from the original, are "acceptable". It is quite possible that students may write words which are acceptable semantically but not grammatically. Consider these: "within

one year *between* each other," "stuck *in* home a lot," and "the older twins are being *eaten* at the same time the younger . . ." The students who made these responses may have understood the passage perfectly, but

Roger Williams has taught ESL for the past two years at Pahlavi University, Iran. Previous to that time he taught in the English Language Institute at the University of Guam.

how much credit should each response be given? Is this a grammar test or a reading comprehension test or both?

One author suggests that a cloze test should be given to native students first and that non-natives would then be graded according to the native responses.¹ This is certainly a possibility when testing very advanced non-natives, but the "errors" cited above would likely not be made by natives and yet are in a sense acceptable. So this matter of grading responses is not easily resolved.

The second problem is that there seems to be a special talent involved in being successful on a cloze test. Carroll maintains that because natives vary widely in their ability to perform well on cloze tests and because this ability correlates with their ability in second-language cloze tests, their scores in the latter should be adjusted when used in comparison with the scores of other students.² He further states that

cloze procedure tests depend to a considerable extent upon cognitive ability variables which are completely extraneous to foreign language success. That is to say, even an individual who has good mastery of a foreign language may not be able to demonstrate this mastery on a cloze procedure test if he lacks certain other intellectual qualities such as reasoning ability and ideational fluency.³

Carroll concludes that because success in the usual types of cloze tests is independent of

language proficiency and more dependent on intellectual functioning, cloze tests should be modified if used. The blanks should be, he suggests, linguistic cues rather than crucial semantic cues, or they could be multiple choice items or items with part of the word present.⁴

This brings us to the third problem, the blanks themselves. One reason that natives vary in their ability to fill in the blanks is that when every *n*th word is deleted, a high number of crucial content words may be deleted. Often these cannot be supplied by looking at the context. Sometimes a student who is quite familiar with the subject matter can fill in these crucial content words while those lacking his special background cannot.

Deleting every *n*th word often causes no problem in an easy passage with high redundancy, but other passages should perhaps receive a different treatment. It might be better for test constructors to delete function and content words which should be apparent from context. For example, suppose a teacher is selecting a passage to use as a cloze test and he finds an article on the speed of sound. It is on or slightly below his students' reading level and is long enough to have sufficient blanks for an accurate measure. Part of the passage reads:

The speed of sound in air at ordinary temperatures is about 1,100 feet per second, which is about one mile in five seconds or about 700 miles per hour. The speed of sound increases slightly with a rise in temperature and falls with a decrease in temperature. It is not affected by the pressure of the air.⁵

What problems can be anticipated? First, there is the problem of subject matter. Students who have not previously been exposed to these concepts may do poorly. Secondly a good physics student may be able to supply the numbers, whereas other students might find them or any calculations difficult. If it is determined that this subject is no problem for this particular group and it is decided to leave in the numbers, which words could be omitted? The phrase "the speed of sound" is repeated, and students should reasonably be expected to supply any one of those four words by comparison to the other phrase. The phrases "increases . . . with a rise in temperature" and "falls with a decrease in temperature" are parallel contrasting ideas. The second *with* and

temperature should be easy to supply, as should one of the four words, *increases, rise, falls, decrease*. Of course, one must expect equivalent content words in this case--*drops, decreases, or declines*, instead of *falls*, for example. Perhaps the third *about* could be deleted.

It would seem that native speakers should be able to get a high or perfect score on a cloze test designed for non-natives. If they do not, perhaps factors besides reading ability are being tested. Here is another example which is part of a longer passage. Natives can fill in the blanks readily; intermediate non-natives might vary according to reading ability.

As the _____ circled over the airport, everyone sensed _____ something was wrong. The plane _____ moving unsteadily through the air, and although the passengers had fastened their seat belts _____ were suddenly thrown forward. At that moment, the air-hostess appeared. _____ looked very pale, but was quite calm. Speaking quickly but almost in a whisper, she informed everyone that the pilot had fainted _____ asked if any _____ the passengers knew anything _____ machines--or at least how _____ drive a car. After a moment's hesitation, a man got up and followed the _____ into the pilot's cabin.⁶

In short, passages should be carefully chosen according to passage level and content, and blanks might be chosen according to the ease with which they can be filled by looking at the context or the structure patterns.

Although this method of testing reading has yet to be perfected, cloze tests can be a useful supplement to vocabulary tests, tests of grammatical structures common to written English, and passages with questions. The teacher who experiments with cloze tests should find them an exciting and inventive way of testing the structural aspect of reading.

¹ Donald K. Darnell, "The Development of an English Language Proficiency Test of Foreign Students Using a Clozentropy Procedure," Final Report, Research Project No. 7-H-010 (Boulder, Colorado: University of Colorado, October 1968), pp. 1-10.

²John B. Carroll, *An Investigation of Cloze Items in the Measurement of Achievement in Foreign Languages*, College Entrance Examination Board Research and Development Reports (Cambridge, Mass.: Harvard University, April 1959), p. 52.

³*Ibid.*, p. 114.

⁴*Ibid.*, p. 117.

⁵G.C. Thornley, *Easier Scientific English Practice*. (London: Longmans, Green and Co., 1964), p. 68.

⁶L.G. Alexander, *A First Book in Comprehension, Precis, and Composition*. (London: Longmans, Green and Co., 1965), p. 61.
