

# PROBLEMS IN CLOZE TESTING RE-EXAMINED

## A Reply to Roger K. Williams by Kenneth G. Aitken

Cloze tests are deceptively simple reading tests that merit greater consideration by ESL teachers. The purpose of this article is to reexamine three issues pertinent to the use of cloze procedure that were recently raised by Williams (1974). In order to familiarize the reader with the subject, I will first describe the construction of a cloze test, then I will review each of the three issues in question and propose explanations to clarify the issues.

Cloze tests, pioneered by Wilson Taylor (1953), are constructed by simply randomly deleting every n-th word from a prose passage. The subject must then supply an appropriate fill-in by guessing from the remaining context. Pack (1973) suggests an every-fifth-word deletion pattern over a 250 word passage. However any n-th word pattern will do. A fifty deletion test seems to be a convenient length and provide a reasonable sample of the passage.

Williams (1974) raises three crucial issues concerning the validity, hence, utility of cloze tests in ESL:

- (1). Cloze tests only partially parallel the reading process.
- (2). There seems to be a special talent involved in being successful on a cloze test.
- (3). Crucial content words are often omitted which cannot be supplied by examining the context.

Do cloze tests only partially parallel the reading process? Williams argues that reading and listening are decoding process; that is to say they are skills of reception rather than production. Filling in blanks, then, would require production skills as it is an encoding process. Recently, however

psycholinguists have begun to question the assumption that encoding and decoding skills are autonomous from each other. Holtzman (1967) has suggested that there is an integrative factor of language use which he calls "linguistic perceptual expectancy". This factor, which Spolsky (1968) calls "redundancy utilization," underlies both the so-called encoding skills of speaking and writing, and the decoding skills of listening and reading. Goodman (1969:82) makes this point regarding the reading process:

Research has demonstrated that the reader does not process print sequentially, but rather in a manner which reflects his use of language at every opportunity. Expectancies about syntax and semantics within context lead to hypotheses which can be confirmed (or disconfirmed) with only a small portion of the cues in the text.

---

**Kenneth G. Aitken, a former BATESL major at Church College of Hawaii (now Brigham Young University-Hawaii Campus) is an ESL instructor at Vancouver Community College in Vancouver, B.C., Canada, where he is working on the testing and placement of ESL students.**

---

Oller (1973) states that Goodman's observation applies "in the case of every observable aspect of language usage." Assuming the Goodman hypothesis is correct, then there is an encoding or production skill as well as an inseparable decoding or receptive skill in every act of reading.

Consider what happens when a subject  
(continued on page 16)

# CLOZE TESTING

(continued from page 9)

responds to the items on a cloze test. The subject must guess a word to fill a blank and complete a sequence, on the basis of incomplete information. This draws on both the productive and receptive skills previously discussed. The information he has available in the context allows him by analysis, to synthesize, or guess whole. Simultaneously, that synthesis, or guess, may become part of the next analysis required to produce some future synthesis. Oller (1973:114) explains the process as follows:

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. By sampling subsequent sequences, he either confirms or disconfirms these expectations. If the expectations are disconfirmed they must be revised and new hypotheses must be formed."

If the above assumptions concerning the reading process and the process of writing a cloze test are valid, then we can say that cloze tests do, indeed, parallel the reading process.

Williams second issue was that Carroll et al (1959) suggest that there seems to be a special talent involved in being successful on a cloze test. Two major limitations to the Carroll et al study that detract from the above generalization were based on the results from a very small number of subjects. The second limitation was that they marked correct only exact word replacements then justified their generalization by correlating the test results with the Modern Language Aptitude Test (MLAT). Oller (1972) has found that the any-acceptable word replacement scoring method is the most reliable and convenient way of scoring cloze tests of ESL students. One might also reasonably question Carroll et al's procedure of correlating test scores with a language aptitude test. It would probably be better to use a proficiency

criterion for validation (Oller 1973).

Darnell (1968) used a scoring method which drew upon a previously determined domain of responses from native speakers for each cloze item. This technique yielded a high (.83) correlation with the *Test of English as a Foreign Language*. Oller (1972) found similarly high correlations between cloze tests marked by the any acceptable word method, and the *UCLA ESL Placement Exam*. Stubbs and Tucker (1974) also report high correlations (.76) with the *English Entrance Exam* used at the American University of Beirut. Williams reports that the Carroll study found that success in cloze tests of foreign language is independent of foreign language proficiency, but it does not appear that Carroll used proficiency as a criterion. It seems, then, that Carroll's conclusions were hastily drawn.

The third issue that Williams raises in his article is that crucial content words are often omitted which cannot be supplied by examining the context. He suggests that rather than a random deletion system that "it might be better for test constructors to delete function and context words which should be apparent from context." His reasoning is that:

- (1). Students who have not previously encountered the concepts in the passage may do poorly.
- (2). Native speakers should be able to get a high or perfect score on a cloze test for non-native speakers.

Each of Williams two reasons were discussed in light of his recommendation. His first reason raises the question, what is the purpose of the cloze test? If cloze tests are used as a pre-assessment of knowledge of the subject matter of a passage then we certainly do not want to penalize the student whose entering competencies into the task are greater than the average. If a cloze is used as a post-assessment of reading comprehension of an assigned reading then it is just those crucial content words that we are most likely to be interested in. If a cloze test is used to determine the readability of a text to

ascertain its suitability for a class, we certainly do not want to discriminate against the experienced individual because it may well be his familiarity with the subject, not his reading skill (whatever you define that as!) that aids him in reading any passage. Essentially Williams' argument, as I see it, is that language skills must be measured separately from one's experience. I contend that this is impossible: language cannot be separated from its use in communication, or from the individual's experience bank of interactions with his perceived world. If, then, a cloze test is used as a language proficiency measure, it would not be inappropriate to delete words that cause the student to probe his experience bank for a response.

Williams, of course, is right when he cautions teachers to select passages for cloze that are appropriate to the class or contextual domain they wish to test. Cloze tests so chosen and marked by the any-appropriate-word system will in all likelihood be valid.

The second reason Williams gives for deleting only function words and content words apparent from the context is that he feels that native speakers should be able to make high or even perfect scores on a cloze test designed for second language speaker. It seems that this would be more of a language proficiency test than a reading test. This is an attractive possibility, but is this what Williams wants to use cloze tests for? The every n-th word system is the simplest method and is still quite satisfactory. We have already found that Darnell (1968) and Oller (1972) using the usual method of construction have obtained high correlations with generally accepted measures of ESL proficiency. By changing the deletion system to make a cloze test easier for a native speaker, one would suspect that the internal consistency of the items and the test reliability for ESL students might be upset. Using an every n-th word deletion system to construct cloze tests, Oller (1972) found that out of 150 items only seven failed to discriminate significantly at the .05 level by a *t* test between the top 50% and bottom

50% of students. Other studies (Oller 1973) have yielded similar results. All these studies using an every n-th word deletion ratio had Kuder Richardson reliabilities between .80 and .90.

The simplicity of test construction is one of the most attractive features of cloze tests. Rather than complicate the construction process, it would be more practical in most circumstances to use an easier reading selection to construct the test from. Oller (1972) found that cloze tests which were rated "very easy" by the Flesch formula (1948) and on the Fourth grade by the Dale and Chall (1948) system yielded correlations between .63 and .80 with the various parts of the UCLA ESL Placement Exam. The subjects were 398 foreign students in UCLA of which 210 were graduate students. Not only would it be more practical to use an easier reading selection than change the deletion system, but also it would be practically as valid as using a more difficult passage.

The construction of cloze tests have been reviewed with the three criticisms of them raised by Williams. The claim that cloze tests only partially parallel the reading process, has been refuted and it has been shown that current psycholinguistic theory suggests an integrative language factor that does not separate reception from production in language use. I have challenged the notion that a special talent is needed in being successful on a cloze test, citing Oller's (1973) criticism of Carroll's procedures and argued that changing the deletion procedure from every n-th word to a system that would make all fill-ins for deletions apparent from the context is unsound and impractical.

#### BIBLIOGRAPHY

- Carroll, J.B., A.S. Carton, and C.P. Wilds. (1959). "An Investigation of Cloze Items in Measurement of Achievement in Foreign Languages." *College Entrance Examination Board Research and Development Reports*. Cambridge, Mass.: Laboratory for Research in Instruction, Harvard University. (continued on Page 18)

## CLOZE TESTING RE-EXAMINED

(continued from page 17)

Dale, E. and J.S. Chall. (1948). "Formula for Predicting Readability." *Education Research Bulletin*, vol. 27, pp. 11-20, 37-54.

Darnell, Donald K. (1968). "The Development of an English Language Proficiency Test of Foreign Students Using a Cloze-tropy Procedure". Final Report, Research Project No. 7-H-010, Boulder, Colorado: University of Colorado ERIC ED 024 039.

Flesch, R.F. (1948). "A New Readability Yardstick." *Journal of Applied Psychology*, vol. 37, pp. 221-223.

Goodman, Kenneth S. (1969). "Analysis of Oral Reading Miscues: Applied Psycholinguistics". *Reading Research Quarterly*, vol. 5, pp. 9-30.

Holtzman, Paul D. (1967). "English Language Proficiency Testing and the Individual." *Selected Conference Papers of the Association of Teachers of English as a Second Language*, pp. 76-84. Los Altos, Calif.: Language Research Associates Press.

Oller, John W. Jr. (1973). "Cloze Tests of Second Language Proficiency and What

They Measure." *Language Learning*, vol. 23, No. 1, pp. 105-118.

Oller, John W. Jr. (1972). "Scoring Methods, and Difficulty Levels for Cloze Tests of ESL Proficiency." *Modern Language Journal*, vol. 56, pp. 151-158.

Pack, Alice C. (1973). "Cloze Testing and Procedure," *TESL Reporter*, vol. 6, no.2, pp. 1-2.

Spolsky, Bernard, Bengt Sigurd, Masahito Sato, Edward Walker, and Catherine Aterburn. (1968). "Preliminary Studies in the Development of Techniques for Testing Overall Second Language Proficiency". *Language Learning*, Special Issue, no. 3, pp. 79-101.

Stubbs, Joseph Bartow and G. Richard Tucker. "The Cloze Test as a Measure of English Proficiency". *Modern Language Journal*, vol. 58, no. 5-6 (September-October 1974), pp. 239-241.

Taylor, Wilson L. (1953). "Cloze Procedure: A New Tool for Measuring Readability." *Journalism Quarterly*, vol. 30, pp. 414-438

Williams, Roger K. (1974). "Problems in Cloze Testing." *TESL Reporter*, vol. 7, no. 4. pp. 7-9.