DOCUMENT RESUME

ED 039 108 RE 002 728

AUTHOR Guszak, Frank J.
TITLE A Comparative Study of the Validity of the Cloze Test and Metropolitan Achievement Test (Reading Comprehension Subtest) for Making Judgments of Instructional Levels.
INSTITUTION Texas Univ., Austin.
PUB DATE Aug 69
NOTE 36p.

EDRS PRICE EDRS Price MF-$0.25 HC-$1.90
DESCRIPTORS *Cloze Procedure, *Elementary School Students, Grade 4, Grade 5, Grade 6, Informal Reading Inventory, *Reading Assignments, *Reading Difficulty, *Reading Level, Reading Research, Standardized Tests

ABSTRACT

The use of the cloze procedure to determine instructional reading levels for pupils in grades 4, 5, and 6 was investigated. For each grade, 50 subjects were selected who, on the basis of teacher judgment, were reading at their respective grade levels. The Botel Reading Inventory, Form A (Word Opposites) served as the criterion for instructional level. Compared with the Botel Test scores were scores from three cloze tests for each grade level (below grade level, at grade level, and above grade level) and the scores of the Reading Comprehension subtest of the Metropolitan Achievement Tests. Findings revealed relatively low correlations (.11 to .18) between the Botel and the cloze tests. The Botel and the Metropolitan tests correlated fairly high at grades 4 and 5 (.49, .21), and showed a statistically significant difference at grade 6. No significant relationships were noted between the cloze test instructional levels and the Metropolitan Test scores. It was suggested that different means of assessing comprehension may have accounted for low correlations between the cloze and the Botel tests. Tables and a bibliography are included. (WB)

A Comparative Study of the Validity of the Cloze Test and Metropolitan Achievement Test (Reading Comprehension Subtest) for Making Judgments of Instructional Levels

by

Frank J. Guszak

The University of Texas at Austin

August, 1969

TABLE OF CONTENTS

INTRODUCTION

Common observation and research (Harris, 1961; Botel, 1968; Bond and Tinker, 1967) indicate that many elementary grade children are asked to read books and materials that are too difficult for them. Although the condition is largely the result of the graded system which generates assumptions that all children learn all things at virtually the same time, it seems imperative that teachers place students in materials which are commensurate with the students' reading skills.

Seemingly, the first teaching task must be one of determining the appropriateness of reading materials for the various students. To some extent, the standardized achievement tests which are frequently offered at least once a school year in most school systems, provide such information. However, as many teachers note, the results of such tests do not provide a reliable index of reading success in various materials. The reasons for this are basically the following:

1) Achievement tests are based on limited samples and cannot predict achievement accurately in specific materials which draw on varied concepts, sentence patterns, etc.
2) Achievement tests are most reliable in the middle ranges of achievement and are consequently often very misleading in measuring the achievement of those in the lower ranges (Harris, 1962; Bond and Tinker, 1967).

Because of the limited value of standardized tests for determining the suitability of given reading materials for given students, many reading authorities (Betts, 1946; Johnson and Kress, 1967; Harris, 1962; Barbe, 1961) suggest informal tests of the involved materials. That is,

the best test of reading skill resides in the student's ability or inability to read the given material. Thus if a sixth grade teacher wishes to determine which students can profitably read the sixth grade geography text, the teacher must:

1. Direct each student to read a specified portion of the text.
2. Direct the student to demonstrate some degree of understanding (whatever is deemed basic by the teacher). Such understanding is normally done by answering questions about the selection.

Such testing in the materials is generally called informal reading inventory testing." In most instances the label is equated with the task of finding pupil's reading levels by asking them to read a series of increasingly difficult selections (followed by comprehension questions). Students in the earlier stages of reading development read the various materials both orally and silently while higher level students normally read only silently before answering the questions.

Although potentially valuable, informal reading inventory testing involves many qualitative decisions on the part of the teacher such as:

Oral Reading
What are oral reading errors?
What are the maximum number of oral reading errors that can be permitted?
How fluent should the oral reading be?
How do you determine fluency?

Silent Reading
What is a reasonable amount of time to read the given selection?

Comprehension
What are the most important elements that the student should remember about the selection?
To what extent are the questions relevant to the main elements of the selection?

It must be evident that the quality of teacher judgments in inventory reading assessment is dependent upon very sophisticated judgments. In

fact, the judgments can be so sophisticated that certain reading theoreticians suggest that teachers may make completely inappropriate judgments if they use the prevailing error marking systems (Goodman, 1967; Hunt, 1969; Powell, 1968).

At this point the question may well be asked, "If teachers cannot depend upon achievement tests or their own observations for determining the suitability of reading materials for different children, what, then can they use?" The response to this question has been made in two very different ways. One prominent means has been the effort of several diagnostic reading test authors to develop tests that can more accurately predict the proper instructional level. Spache (1963), Botel (1968), and others have presented data to indicate that their special instruments will predict more accurately than achievement tests. Another prominent means has been seen in the "cloze technique" procedure as developed by Bormuth (1967 a , 1968). In the cloze procedure, students are asked to restore omitted words (usually every fifth word) in a reading passage. On the basis of correct restorations, Bormuth (1967 a , 1968) and Coleman (1969) indicate that rather accurate determinations of comprehension can be made.

Because the tests of Botel and Spache (as any other standardized instruments) suffer from the same limitations as achievement tests, it appears that their power in determining the appropriateness of reading material is somewhat limited. Devoid of such restraints and geared to the exact material, the cloze test procedure offers a most valuable means of determining the readability of any selection for any student. Thus, the current study was designed to seek further information about the

predictive value of the cloze test in the elementary grades. Such determination was to be made by comparing the results of individual pupil's scores on cloze tests as well as other types of reading measures.

PROBLEM

Students are, as previously indicated, often asked to read materials which are beyond their reading capabilities. In order to make a valid judgment of whether or not a student can read given materials, it seems imperative to test the student directly in the materials. Taylor (1957), Bormuth (1967 a), Jenkinson (1957), and Coleman (1962) have revealed research support to indicate that cloze tests (involving the restoration of omitted words) can be used as valid and reliable predictors of reading comprehension. The current research was designed to further study some additional aspects of the cloze tests' ability to determine reading comprehension.

In order to determine how well cloze tests predict readability it was essential to determine readability by some other means. Thus, the current study sought to make comparisons of the cloze test with the Botel Reading Inventory and the reading subtest (Comprehension) of the Metropolitan Achievement Test.

Botel (1968) reported that his reading inventory correlated very highly with the placements made by a group of study teachers using carefully prescribed informal reading inventory techniques. (The correlations ranged from as high as .95 in Grade 2 to the low of .73 in Grade 6). In making similar comparisons between various standardized achievement test scores and the teachers' findings of instructional levels, he found that the standardized tests faired considerably poorer than his instrument. From his research, he concluded that the Botel Inventory was more closely related to the criterion than the standardized silent reading tests used.

Because the Botel correlations were consistently higher in the intermediate grades than those of the standardized achievement tests, the decision was made to use the Botel Reading Inventory, Form A. (Word Opposites) as the criterion of instructional level. Against this criterion score for each child would be placed his instructional score on the cloze test and his Metropolitan Achievement Test (Reading Comprehension) score. Specifically, the study was addressed to the following hypotheses:

Hypothesis 1
There is no significant relationship between the cloze test instructional levels and those of the Botel Reading Inventory Form A (Word Opposites) in Grades 4, 5, and 6.

Hypothesis 2
There is no significant relationship between the Metropolitan Achievement Test (Reading Comprehension) scores and those of the Botel Reading Inventory Form A (Word Opposites) in Grades 4, 5, and 6.

Hypothesis 3
There is no significant relationship between the cloze test instructional levels and the Metropolitan Achievement Test (Reading Comprehension) levels in Grades 4, 5, and 6.

Hypothesis 4
There is no significant difference between the relationship of the cloze test instructional levels and the criterion and those of the Metropolitan Achievement Test (Reading Comprehension) and the criterion in Grades 4, 5, and 6.

BACKGROUND LITERATURE

A study of the project title "A Comparative Study of the Validity of the Cloze Test and Metropolitan Achievement Test (Reading Comprehension Subtest) for Making Judgments of Instructional Levels' suggests the presence of a valid criterion of instructional level. Actually, such validity is in question because the validity in this study is based upon research by Botel (1968), wherein he found high, positive correlations between his reading test and teacher judgments of instructional levels in a certain school. Conceivably, what was reported as validity may have been only reliability between the two sets of scores. At issue are the basic determinants of the "instructional level." In this review the discussion will center upon the concept of the so-called "instructional level" or "optimum level of difficulty" and the means relative to the determination of such.

The Instructional Level

Beldin (1969), traces the beginnings of the concept of an instructional level (or level of difficulty which appears optimum for instruction) well beyond Killgallon (1942), and Betts (1946), and found the concept in the early writings of Gray (1925), Thorndike (1934), and others.

What the optimum determinants of reading difficulty happens to be remains a subject of great debate. On the one side can be found individuals of the Betts persuasion who appear to accept the following criteria:

Levels of Reading Difficulty

	Word Recognition	Comprehension
Independent	99%	90%
Instructional	95%	75%
Frustrational	90% or less	50% or less

On the opposite side we find individuals such as Powell (1968), Hunt (1969), Spache (1969), and others who find the Killgallon--Betts criteria to be arbitrarily fashioned and not commensurate with reality. Both Spache and Powell indicate studies that have demonstrated that comprehension (at the Betts' level) can be obtained with significantly lower word recognition skill than outlined by the criteria. Powell found that first and second graders could on the average comprehend at the 70% level with 85% word recognition.

If word recognition criteria are dropped and the focus is placed entirely upon the comprehension factor there remains disagreement as to the minimum level of acceptable comprehension. Whereas the Killgallon-Betts criterion was 75%, Spache feels that a comprehension of 60% is acceptable. Frequent modifications of the initial I.R.I. concepts frequently imply a 70% criterion because of the pattern of using a ten question format.

When the multitude of variables which surround the informal reading inventory concepts are taken into consideration, it becomes very difficult to determine what optimum functioning really is. So difficult indeed is the process, that McCracken (1964), deemed it useful to add a rate dimension that would further identify the nature of the instructional task, i.e. indicate a plodding reader who might be both accurate in word recognition and comprehension.

Conceivably, much research remains before we can fix specific percentage criteria for determining optimum degrees of instruction. Such research must certainly describe the effects of varying types of instruction upon varying types of pupils (emotional makeup, etc.) using varying kinds of materials.

Achievement Tests and Instructional Levels

As indicated in the introduction, achievement tests have been criticized as useful measures of instructional reading levels. The problem of using standardized reading tests for making judgments about instructional levels is very clearly illustrated by McDonald (as cited by Emans, Urbas, and Dumett, 1966) in discussing a research project in a high school that administered different reading tests to the same students with differing results. Botel (1968), revealed evidence which indicated that achievement test placements were very far off when low group children were concerned. His results revealed discrepancies to the extent that more children were improperly placed by the standardized tests than were properly placed. Reports of others (Harris, 1961, Bond and Tinker, 1967), reveal similar pictures, especially with regard to children who deviate from the mean reading levels tested by the various tests.

It should be pointed out that achievement tests are basically instruments designed for broad achievement assessment and not the assessment of an individual's ability to read a given text on a given day in a given situation. Thus, it is not highly reasonable to anticipate book placement on the basis of such instruments.

Cloze Tests and Instructional Level

Wilson L. Taylor (1953) introduced the term "cloze procedure" in 1953 and initiated considerable research relative to the value of closure

tasks as predictors of reading comprehension.

Basic to the procedure is the idea of closure wherein the reader must use the surrounding context in order to restore omitted words. Comprehension of the total unit and its available parts (including the emerging cloze write-ins) is essential to the task. As the task has been described by Taylor, Bormuth (1968), and others, cloze tests have generally involved the following administration and scoring protocols:

Administration

1. Every nth word (usually every 5th word, Bormuth (1968), is omitted and in its place a blank of sufficient length allowed for the student to write in the answer.
2. The student is instructed to write only one word in each blank and to try to fill in every blank.
3. Guessing is encouraged.
4. Students are advised that spellings will not be counted as errors.

Scoring

1. In most instances, the exact word must be restored (Bormuth, 1968).
2. Misspellings are counted as correct when the response is deemed correct in a meaning sense.

The validity of the cloze test as a measure of readability and comprehension has been most interesting because of (1) the ways in which it has been accomplished and (2) the almost universal finding of high correlations between cloze and other prediction instruments.

Initially Taylor (1953) compared cloze score rankings of passages of varying difficulty with readability rankings of the same passages by two common readability formulas (Dale-Chall, 1948, Flesch, 1949). The passages were similarly rank ordered by each technique. Superiority for the cloze procedure was demonstrated when very difficult passages could be more readily assessed as such by the cloze procedure.

Another way in which the cloze test has been validated has been by comparing results between cloze tests and standardized tests of reading achievement. Potter (1968) reveals in Table 1 the following correlations of a number of such studies:

Table 1

Correlations Between Cloze Readability Tests and Standardized Tests of Reading Achievement

Study	Subjects	Tests	Correlations
Jenkinson (1957)	High School	Cooperative Reading C2	
		Vocabulary	.78
		Level of Comprehension	.73
Rankin (1957)	College	Diagnostic Survey	
		Story Comprehension	.29
		Vocabulary	.68
		Paragraph	.60
Fletcher (1959)	College	Cooperative Reading C2	
		Vocabulary	.63
		Level of Comprehension	.55
		Speed of Comprehension	.57
		Dvorak-Van Wagenen	
		Rate of Comprehension	.59
Hafner (1963)	College	Michigan Vocabulary Profile	.56
Ruddell (1963) (5 cloze tests)	Elementary	Stanford Achievement	
		Paragraph Meaning	.61-.74
Weaver and Kingston (1963, 2 cloze tests)	College	Davis Reading	.25-.51
Green (1964)	College	Diagnostic Reading Survey	.51
		Total Comprehension	

Bormuth (1967 a) demonstrated a relationship between cloze and multiple-choice scores of subjects and further determined that a 43 percent cloze score was the equivalent of a 75 percent multiple-choice test score (with corrections made for guessing). Thus, the link between cloze testing and the concepts of instructional level determination was made (using the Betts-Killgallon comprehension criterion of 75%). It appears from the research, that Bormuth has equated cloze performance with the criterion (regardless of the accuracy of the criterion).

As suggested earlier, it is interesting to note the consistency with which the cloze test correlates with other measures of comprehension. The only notable exception appears to be a study by Kingston and Weaver (1963) wherein the researchers measured comparably low correlations of .25 and .51 between cloze test scores and scores of college students on the Davis Reading Test.

Although most of the studies of cloze have taken place from the fourth grade and up, two studies are reported for the lower grades. Gallant's (1965) data suggested that the technique was appropriate for first, second and third grade children. Changes such as the insertion of three words for each blank for choice purposes (a multiple-choice task) and the lengthening of certain sentences to obtain higher Spache Readability levels raise questions about the study. Deutsch (1964) utilized a cloze task via an auditory means (by asking students to fill in the word as the voice model paused) and found that cloze tests were highly related to I.Q.

To date, the background literature suggests that the concept of an instructional or "optimum level of reading instruction" is a concept that

is widely discussed, but which does not possess a necessarily common set of characteristics.

Research relative to the comparative effectiveness of (a) standardized achievement tests and (b) Cloze tests, indicates that the latter practice has greater predictive value. Thus, this study sought to further explore the comparative value by comparing Metropolitan and Cloze test scores.

RESEARCH PROCEDURES

Sample Selection

Prior experience with cloze testing of pupils below fourth grade reading skill had indicated to the researcher (although not to others, i.e., Jenkinson, 1957) that the nature of the task was rather difficult. This experience suggested that the study might best be conducted with pupils reading 4th grade level and above.

Arrangements were made to test pupils in the Brentwood Elementary School in Austin, Texas, who were reading at their respective grade levels (fourth, fifth, sixth) in the judgments of the teachers involved. In essence, this meant that fifty students were selected from each class of approximately one hundred and fifty students.

The initial samples of fifty students were picked on the basis of their proximity to grade level expectation (as based on teachers' judgments). Inevitably, intact reading groups were in the sample.

Because of the unique planning arrangements, the final sample didn't look quite as anticipated in that actual book placements were under those of the standardized achievement tests. For example, of the 50 fourth graders, only 17 were reading a fourth grade level text. The balance were reading third grade level texts. Even more striking was the situation in the sixth grade where only eleven were reading a sixth grade material at the time of testing.

Test Preparation

In order to develop a cloze test that would be a representative measurement of each grade level the following criteria were established.

1. Each selection used would have to meet Bormuth's specifications of lenght (250 words or 50 cloze items as a minimum).
2. Each test should be contained within a single page so as to facilitate close scrutiny by the student.
3. Each selection should be representative of the difficulty of its purported level as determined by the Dale-Chall or Spache Readability Formulas (from book other than reader).
4. Lines should be typed and double spaced so as to permit the student to easily manage the visual difficulties that might occur. Blanks were ten spaces in length so as to permit a reasonable space for pupil write-in answers.

Information pertinent to the samples chosen is presented in the table below (Table 2).

Table 2

Cloze Samples, Their Sources, and Their Readability Levels as Determined by The Spache or Dale-Chall Readability Formulas

Grade	Cloze Sample Title	Publisher	Recom. Level	Readability
2	David's Silver Dollar			2.4 (Spache)
3	The Valentine Box		3rd	3.3 (Spache)
4	The Brave Little Tailor			4.5 (Dale-Chall)
5	Station in Space			5.4 (Dale-Chall)
6	The Squeak of Leather	Ginn	6th	6.1 (Dale-Chall)
7	The River of the Wolves	Ginn	7th	7.1 (Dale-Chall)

After the cloze tests were prepared, testing sets for the three grade levels were made up. Each packet contained the following:

Botel Word Opposites Test.

Cloze Test. One Grade Below Presumed Grade Level.

Cloze Test. On Presumed Grade Level.

Cloze Test. One Grade Above Presumed Grade Level.

Thus, after completing the Botel test, each student would begin working on a cloze test which was presumably one grade level below his current reading level. He was then supposed to progress through the other tests. Students who did not pass the lowest tests or who passed the highest tests were given further tests to determine their reading levels.

Test Administration

Arrangements were made for the selected sample group in each grade to meet in a large space (4th graders in a given room, 5th graders in another room, etc.). The following instructions were given to the examiners:

1. ASK EACH CHILD TO WRITE HIS/HER NAME ON THE TEST. Check to see that each child has written his/her name on the first page of the test.

2. EXPLAIN THAT THERE ARE TWO KINDS OF TESTS IN THE PACKET.
   A. READ THE DIRECTIONS FOR THE FIRST TYPE OF TEST (PICK A WORD IN EACH LINE WHICH MEANS THE OPPOSITE OF THE NUMBERED WORD). READ THE EXAMPLE AND BE SURE THAT EVERYBODY UNDERSTANDS THE TASK. PAGE THROUGH THE FIRST FIVE PAGES AND SHOW THEM THAT THESE PAGES COMPRISE THE FIRST TYPE OF TEST.

   B. EXPLAIN THAT THE SECOND TYPE OF TEST IS A "FILL IN THE BLANK" TYPE OF TEST AND THAT THEY ARE TO INSERT THE WORDS THAT SEEM TO BE NEEDED IN ORDER FOR THE SELECTION TO MAKE SENSE. SHOW THEM THAT THE LAST THREE PAGES OF THE TEST PACKET ARE TESTS OF THIS TYPE. FURTHER INDICATE:
      1. It is well to read through the whole "Fill in the Blank" page first before writing in words.
      2. Spelling errors will not be counted.
      3. They may erase and change words if they like at any time.
      4. After they have finished each page they should read through it again to see if they are satisfied with their answers.
      5. If they simply can't think of an answer for a space to go past it and come back to it later.

STRESS THAT THE TASK WILL BE DIFFICULT AND THAT THEY SHOULD TRY NOT TO BE TOO DISCOURAGED IF THEY DON'T KNOW WHAT ANSWERS TO PUT IN. FURTHER EXPLAIN THAT WE HAVE SOME OF THE SAME PROBLEMS THAT THEY HAVE WHEN WE TAKE TESTS LIKE THIS.

Notes

1. Circulate to determine whether the students understand the test tasks.
2. Try to allow as much time as each individual needs to complete the tasks.
3. If this appears to be too much for one sitting, please make provisions to stop at some point and complete the other tasks at another sitting.

In the fourth and fifth grade testing sessions, the author was present and gave the initial instructions. He was not present in the sixth grade examination.

Time for the examination varied rather sharply, with a variance of 45 to 88 minutes in the fourth and fifth grades.

It was noted by the examiner that the cloze test task in the fourth and fifth grades appeared quite difficult to the children. Many suggested that it was "too hard."

Statistical Procedures

Comparisons of the power of the two tests with regard to making instructional level decisions were made by (1) correlation and (2) graphic matching, using the Botel word opposites test (Form A) scores as the criterion scores.

The various reading tests (Metropolitan, cloze, Botel) were correlated in paired fashion with the Pearson Product-moment correlation technique (Bruning & Kintz, 1968). Initially, the cloze results in the three grades were correlated with the criterion results (as obtained by the Botel). Next, the Metropolitan Achievement Test (Reading Comprehension) results in the same grades were correlated with the criterion results (as obtained by the Botel). Then, the results of the cloze and Metropolitan tests were correlated.

The relationships of the cloze and Metropolitan test results to the

criterion were subsequently measured by a "t" test (Bruning and Kintz, 1968).

For a determination of the relative matching of the cloze and Metropolitan tests with the criterion, a chart was developed wherein the placement agreements and disagreements (underplacement and overplacement) could be registered.

RESEARCH FINDINGS

It was the intent of the study to determine the validity of the cloze and Metropolitan tests for making judgments of instructional levels by testing the following hypothesis:

1. There is no significant relationship between the cloze test instructional levels and those of the Botel Reading Inventory Form A (Word Opposites) in Grades 4, 5, and 6.

2. There is no significant relationship between the Metropolitan Achievement Test (Reading Comprehension) scores and those of the Botel Reading Inventory Form A (Word Opposites) in Grades 4, 5, and 6.

3. There is no significant relationship between the cloze test instructional levels and the Metropolitan Achievement Test (Reading Comprehension) levels in Grades 4, 5, and 6.

4. There is no significant difference between the relationship of the cloze test instructional levels and the criterion and those of the Metropolitan Achievement Test (Reading Comprehension) and the criterion in Grades 4, 5, and 6.

Each of these sets of findings is treated in the following material.

Hypothesis 1. There is no significant relationship between the cloze test instructional levels and those of the Botel Reading Inventory Form A (Word Opposites) in Grades 4, 5, and 6.

This hypothesis was not rejected by the data from the correlation of the cloze and Botel Scores in each of the three grades. Results of the correlations indicated very low but positive correlations as follows:

4th Grade - r. .11

5th Grade - r. .17

6th Grade - r. .18

Hypothesis 2. There is no significant relationship between the Metropolitan Achievement Test (Reading Comprehension) scores and those of the Botel Reading Inventory Form A (Word Opposites) in Grades 4, 5, and 6.

This hypothesis was rejected in the fourth and sixth grades but not in

the fifth grade. Results are as follows:

4th Grade - r. .49

5th Grade - r. .21

6th Grade - r. .49*

*significant at .05

The failure to reject the hypothesis at the fourth and sixth grade levels indicates a fairly high correlation between these two standardized instruments at these levels. Seemingly, it would be anticipated that significant, positive correlations would exist at all three levels.

Hypothesis 3. There is no significant relationship between the cloze test instructional levels and the Metropolitan Achievement Test (Reading Comprehension) levels in Grades 4, 5, and 6.

This hypothesis was not rejected in the 4th and 5th grades. However, in the 6th grade there was a negative correlation (-.57) which was significant at the .05 level of significance. Seemingly, part of the difficulty seems to be accounted for by the difficulty of the seventh grade cloze test. Not a single sixth grader passed the seventh grade cloze selection, even though many scored much higher than grade level on the Metropolitan (and the Botel as well).

It is curious to note that the 7th grade selection "The River of the Wolves" checked out as 7.1 on the Dale-Chall Readability Formula. After the testing had been completed the Fry Readability Formula came to the attention of the author and the selection was tested with this instrument. The Fry formula revealed a difficulty of 5th grade.

Hypothesis 4. There is no significant difference between the relationship of the cloze test instructional levels and the criterion and those of the Metropolitan Achievement Test (Reading Comprehension) and the criterion in Grades 4, 5, and 6.

This hypothesis was tested statistically by a "t"* test which revealed a significant relationship at the .05 level in the fourth grade (Table 3). Thus, the hypothesis was rejected at this grade level. The hypothesis was not rejected in Grades 5 and 6.

Table 3

Comparison of r's Between (1) Metropolitan Achievement Test (Reading Comprehension) and (2) Cloze Tests and CRITERION

Grade	Metropolitan	Cloze	t
4	.49	.11	2.08
5	.17	.17	.05
6	.18	.18	.93

* "Test for Difference Between Dependent Correlations" from Computational Handbook of Statistics by Janes L. Bruning and B. L. Kintz, Glenview, Illinois: Scott, Foresman, 1968, p. 193.

Differences in the relationships of the two tests to the criterion were illustrated by comparison with the criterion (Table 4).

In the 5th and 7th grade facility levels as determined by the criterion test (Botel), the Metropolitan Achievement Test placed more students on criterion level than did the cloze test. In the 6th grade, there was little difference with only one more child being placed correctly via the cloze.

Underplacement appeared to be the result of most of the cloze test results. With the exception of fourteen students overplaced one year in the fifth grade and one overplaced two years in the fourth, some 98 students were underplaced from one to four years.

Whereas underplacement appeared to be the result of most cloze test

Table 4

PERCENTAGE OF PUPILS PLACED CORRECTLY, UNDERPLACED, AND OVERPLACED BY THE METROPOLITAN ACHIEVEMENT TEST (READING COMPREHENSION) AND CLOZE TESTS IN GRADES 4, 5, AND 6

		UNDERPLACED				BOTEL	OVERPLACED			
Grade	N	-4	-3	-2	-1	Grade Level	+1	+2	+3	+4
4th Grade Level Metro. Placement	5					2	2	1		
4th Grade Level Cloze Test	5			1		3		1		
5th Grade Level Metro. Placement	61				6	29	20	4	2	
5th Grade Level Cloze Test	61			2	39	6	14			
6th Grade Level Metro. Placement	35			3	4	15	8	4	1	
6th Grade Level Cloze Test	35			14	5	16				
7th* Grade Level Metro. Placement	37		2	10	8	17				
7th* Grade Level Cloze Test	37	1	11	4	20	1				
9th** Grade Level Metro. Placement	1					1				
9th** Grade Level Cloze Test	1	1								

* represents Grades 7-8 on Botel
** represents Grades 9-12 on Botel

placement, overplacement appeared to be the effect of the Metropolitan Achievement Test in the 5th and 6th grades. In the 7th grade both cloze and Metropolitan tests appeared to overplace.

CONCLUSIONS

The very low correlations between the cloze score instructional levels and the instructional levels of the Botel Reading Inventory Form A (Word Opposites) seems rather strange in light of the Bormuth (1967 a), and Botel (1968), studies which indicated the strength of each instrument in assessing instructional levels. Seemingly, the results suggest one or more of the following possibilities:

- The Botel Reading Inventory is not a valid measure of reading comprehension difficulty.
- The cloze test technique (or the use of a single cloze test for assessment) is not a valid measure of reading comprehension difficulty.
- The cloze test difficulty rankings by the readability formulas were not accurate.

It should be pointed out that two tests are very different with regard to their means of assessing comprehension. The Botel (Word Opposites) Test determines comprehension on the basis of the student's ability to read a word and select an opposite word from a group of alternative words. The cloze test requires the reader to establish clozure on a certain percentage of omitted words in a selection of at least 250 words. Seemingly, one might be able to perform the Botel task on the basis of individual word attack skill plus an understanding of key vocabulary. In the cloze task, the larger concerns of syntax and lexical meaning come into play. Thus, it seems possible that a student might be able to perform the individual vocabulary task without being able to do the latter successfully. Thus, it is highly conceivable that the cloze test might be the more valid predictor of the ability of a pupil to read a given selection. Of course, the acceptance of such a generalization must rest heavily upon the validity of the research of Bormuth (1968), and

others who have demonstrated cloze validity by comparisons with other comprehension checks, notably multiple choice.

It was anticipated that high, positive correlations would exist between the Botel Word Opposites Test and the Metropolitan Reading Comprehension Test because of the fact that the standardization process, through which both have been placed, demands such. Thus, the significant, positive correlations in the fourth and sixth grades were anticipated. The low, non-significant correlations at the fifth grade was not anticipated. In studying the comparative results of the three grade groups in terms of their scores the following information was found with regard to the relationship of Metropolitan and Botel scores:

Table 5

Comparability of Metropolitan and Botel Grade Placements in Grades 4, 5, and 6

Grade	No. taking Both Tests	Same Score/ Both Tests	Higher Score/ Botel	Higher Score/ Metropolitan
4th		19	12	13
5th		23	15	8
6th		17	5	22

While the magnitudes of difference which were reflected in the correlations are not evident in Table 5, it is apparent that the variance is rather marked between the three grades. Beginning with the fourth where there are about an equal number of higher scorers in each test, we see the Botel reflecting much higher scores in the 5th and the reverse occurring in the 6th grade.

Explanations of the wide differences between the 5th and 6th grades are difficult. Possibly, the tests administered (either) to one or the other of the groups were not properly administered. Conceivably, one grade might have encountered a different difficulty level in test taking. All possibilities are speculative.

The failure to obtain significant correlations between the cloze and Metropolitan tests (Hypothesis 3) in the 4th and 5th grades was not unexpected. However, the significant, negative correlation of -.57 at the 6th grade was surprising. As indicated the result seemed to be in large part, test difficulty on the part of the cloze test because no sixth graders passed the seventh grade cloze test, even though many scored higher on both Botel and Metropolitan tests. Seemingly, the 7th grade selection "The River of the Wolves" should be submitted to a similar group of sixth graders as a complete story followed by a set of ten questions for comprehension testing. If the story appeared too difficult for most or all of such a group as based on a criterion of seven answers correct, the readability of the selection as determined by the Dale-Chall and Fry formulas would be further questioned. If however, comprehension proved obtainable, the researcher would have to question either (a) the validity of the cloze method for such selections or (b) the testing procedures under which the examinations were given.

IMPLICATIONS FOR FURTHER STUDY

Comparisons of cloze test results with the results of global comprehension measures as involved in such tests as the Botel Reading Inventory and Metropolitan Achievement Test (Reading Comprehension) do not appear to be the most important kinds of comparisons. Rather, comparisons should seek to further validate the cloze comprehension criteria by comparing cloze results with other measures of passage understanding such as multiple-choice questions, open response questions, tell-back techniques, etc. While much of this research has been done, it seems that further validation is needed.

BIBLIOGRAPHY

Aborn, M. Rubenstein, H., & Sterling, T. D. Sources of contextual constraint upon words in sentences. Journal of Experimental Psychology, 1959, 57 (3), 171.

Barbe, Walter B. Educator's Guide to Personalized Reading Instruction. Englewood Cliff, New Jersey: Prentice Hall, Inc., 1961, 241pp.

Beldin, H. O. Informal Reading Testing: Historical Review and Review of Research. Symposium Paper. International Reading Association, Kansas City, Missouri, 1969.

Bennett, S., Semmel, M. I., & Barrett, L. S. In Lane, H. L. & Zale, E. M. (Eds.) Studies in language and language behavior. Center for Research on Language and Language Behavior, University of Michigan, 1965, 414-428.

Betts, E. A. Foundations of reading instruction. New York: American Book Company, 1946.

Bloomer, R. H. The cloze procedure as a remedial reading exercise. Journal of Developmental Reading, 1962, 5, 173-181.

Bloomer, R. H., Heitzman, A. J. Pretesting and the efficiency of paragraph reading. Journal of Reading, 1965, 8 (4), 219.

Bloomer, R. H., Louthan, V., & Heitzman, A. J. Non-overt reinforced cloze procedure. U.S. Office of Education Project Report #2245, University of Connecticut, 1966.

Blumenfeld, J. P. & Miller, G. R. Improving reading through teaching grammatical constraints. Elementary English, 1966, 43, 752.

Bond, Guy L. and Tinker, Miles A. Reading Difficulties--Their Diagnosis and Correction. New York: Appleton-Century-Crofts, 1967, 564pp.

Bormuth, J. R. Cloze tests as measures of readability and comprehension ability. Unpublished doctoral dissertation, University of Indiana, 1962.

Bormuth, J. R. Experimental applications of cloze tests. International Reading Association Conference Proceedings, 1964, 9, 303. (a)

Bormuth, J. R. Mean word depth as a predictor of comprehension difficulty. California Journal of Educational Research, 1964, 15 (5), 226. (b)

Bormuth, J. R. Relationships between selected language variables and comprehension ability and difficulty. Cooperative Research Project #2082, U. S. Office of Education, 1964. (c)

Bormuth, J. R. Optimum sample size and cloze test length in readability measurement. Journal of Educational Measurement, 1965, 2 (1), 111. (a)

Bormuth, J. R. Validities of grammatical and semantic classifications of cloze scores. International Reading Association Conference Proceedings, 1965, 10, 283. (b)

Bormuth, J. R. & MacDonald, O. L. Cloze tests as a measure of ability to detect literary style. In J. Allen Figurel (Ed.), Reading and Inquiry, International Reading Association Conference Proceedings, 1965, 10, 287-290. (c)

Bormuth, J. R. Readability: A new approach. Reading Research Quarterly, 1966, 1 (3), 79. (a)

Bormuth, J. R. Design of readability research. International Reading Association Conference Proceedings, 1966, 11, 485-489. (b)

Bormuth, J. R. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 1967, 10, 291-299. (a)

Bormuth, J. R. The implications and use of cloze procedure in the evaluation of instructional programs. Center for the Evaluation of Instructional Programs, University of California, Los Angeles, 1967. (b)

Bormuth, J. R. Cloze readability: Criterion reference scores. Journal of Educational Measurement, 1968 (in press).

Botel, Morton. A Comparative Study of the Validity of the Botel Reading Inventory and Selected Standardized Tests. Research Report delivered at the Thirteenth Annual Convention, International Reading Association, Boston, Massachusetts, April 26, 1968, 1-15.

Bruning, J. L. and Kintz, B. L. Computational Handbook of Statistics. Glenview, Illinois: Scott, Foresman and Company, 1968, 193-194.

Carterette, E. C. & Jones, M. H. Redundancy in children's tests. Science, 1963, 140, 1309-1311.

Coleman, E. B. Developing a technology of written instruction: Some determiners of the complexity of prose. Paper read at symposium on Verbal Learning and Written Instruction, New York, March, 1966.

Coleman, E. B. Improving comprehension by shortening sentences. Journal of Applied Psychology, 1962, 46, 131-134.

Coleman, E. B. & Blumenfeld, J. P. Cloze scores of nominalizations and their grammatical transformations using active verbs. Psychological Reports, 1963, 13, 651-654.

Coleman, E. B. & Miller, G. R. A measure of information gained during prose learning. Quarterly Journal of Reading, in press.

Dale, E. and Chall, J. A formula for predicting readability. Educational Research Bulletin, January, 1948, 27, 11-28.

Dale, E. & Seels, B. Readability and reading. International Reading Association Conference Proceedings, 1966.

Darnell, D. K. The relation between sentence order and comprehension. Speech Monographs, 1963, 30, 97.

Deutsch, M., Cherry, E., Maliver, H., & Brown, R. Communication for information in the elementary classroom, Institute for Developmental Studies, New York University, 1964.

DeVito, J. A. Comprehension factors in oral and written discourse of skilled communicators. Speech Monographs, 1965, 32, 124.

Dickens, M. & Williams, F. An experimental application of cloze procedure and attitude measures to listening comprehension. Speech Monographs, 1964, 31, 103-108.

Emans, Robert, Urbas, Raymond, and Dummett, Marjorie, The meaning of reading tests, The Reading Teacher, May, 1966, 406-409.

Epstein, W. The influence of syntactical structure on learning. American Journal of Psychology, 1961, 74, 80-85.

Ervin, S. M. Changes with age in the verbal determinants of word association. American Journal of Psychology, 1961, 74, 361-372.

Fillenbaum, S., Jones, L. V., & Rapoport, A. The predictability of words and their grammatical classes as a function of rate of deletion from a speech transcript. Journal of Verbal Learning and Verbal Behavior, 1964, 2, 186-194.

Flesch, R. The art of readable writing. Harper & Row, 1949.

Fletcher, J. E. A study of the relationships between ability to use context as an aid in reading and other verbal abilities. Unpublished doctoral dissertation, University of Washington, 1959.

Gallant, R. Use of cloze tests as a measure of readability in the primary grades. Proceeding of the International Reading Association Convention, 1965, 10, 286-287.

Goodman, Kenneth. Reading: A Cycle Linguistics Reading Game. Paper presented at American Educational Research Association, Chicago, Illinois, 1967.

Gray, William S. The Value of Informal Tests on Reading Achievement, Journal of Educational Research, (January 1920), 103-111.

Greene, F. P. A modified cloze procedure for assessing adult reading comprehension. Unpublished doctoral dissertation, University of Michigan, 1964.

Greene, F. P. Modification of the cloze procedure and changes in reading test performances. Journal of Educational Measurement, 1965, 2 (2), 213-217.

Hafner, L. E. Relationships of various measures of the cloze. In E. L. Thurston & L. E. Hafner (Eds.) Thirteenth Yearbook of the National Reading Conference. Milwaukee, Wisconsin, National Reading Conference, Inc., 1963, 135-145.

Hafner, L. E. Implications of cloze. In E. L. Thurston & L. E. Hafner (Eds.) The philosophical and sociological bases of reading. Fourteenth Yearbook of the National Reading Conference, Milwaukee, Wisconsin, The National Reading Conference, Inc., 1965, 151-158.

Hafner, L. E. Cloze procedure. Journal of Reading, 1966, 9, 415-421.

Harris, A. J. Effective teaching of reading. New York: David McKay, 1962.

Hunt, Lyman C. The effect of self-selection, interest, and motivation upon independent, instructional, and frustrational levels of reading. Paper presented at the Fourteenth Annual Conference International Reading Association, Kansas City, Missouri, May 1 and 2, 1969, 1-8.

Jenkinson, M. E. Selected processes and difficulties in reading comprehension. Unpublished doctoral dissertation, University of Chicago, 1957.

Johnson, Marjorie S. and Kress, Roy A. Informal Reading Inventories. Newark, Delaware: International Reading Association, 1965, 1-44.

Kender, Joseph P. How Useful Are Informal Reading Tests? Journal of Reading, 11, Feburary, 1969, 337-342.

Kerfoot, J. F. Reading in the elementary school. Review of Educational Research, 1967, 37, 120.

Killgallon, Patsy A. A Study of Relationships Among Certain Pupil Adjustments in Language Situations. Unpublished doctoral dissertation, Pennsylvania State College, 1942.

Kingston, A. J. & Weaver, W. W. Recent development in readability appraisal. Journal of Reading, 1967, 11, 44.

Klare, G. R. The Measurement of Readability. Ames, Iowa: Iowa State University Press, 1963.

Lee, W. D. What does research in readability tell the classroom teacher? Journal of Reading, 1964, 8, 141-144.

Jouthan, V. Some systematic grammatical deletions and their effects on reading comprehension. English Journal, 1965, 54, 295.

Luke, M. Form class and cloze procedure. Unpublished manuscript, Language Development Program, Center for Human Growth and Development, University of Michigan, 1964.

MacGinitie, W. H. Contextual constraint in English prose paragraphs. Journal of Psychology, 1961, 51, 121-130.

Marks, M. R. & Taylor, W. L. The influence of contextual and goal contextual and goal constrants on the meaningfulness of "automatic sentences." Journal of Social Psychology, 1954, 40, 43-51.

McCracken, R. A. The Informal Reading Inventory as a Means of Improving Instruction. The Evaluation of Children's Reading Achievement, 1967, 79-96.

McLeod, J. GAP reading comprehension test. Melbourne, Australia: Heinemannn, 1965.

McLeod, J. & Anderson, J. Readability assessment and word redundancy of printed English. Psychological Reports, 1966, 18, 35-38.

Miller, G. A. & Friedman, E. A. The reconstruction of mutilated English texts. Information and Control., 1957, 1, 38-55.

Miller, G. A. & Selfridge, J. Verbal context and the recall of meaningful material. American Journal of Psychology, 1950, 63, 176-185.

Miller, G. R. & Coleman, E. B. A set of 36 passages calibrated for comprehensibility. Inglewood, California: Southwest Regional Laboratory for Educational Research and Development, 1966.

Musgrave, B. The effect of who and what context on cloze and commonality scores. Journal of Social Psychology, 1963, 59, 185-192.

Osgood, C. E. The nature and measurement of meaning. Psychological Bulletin, 1962, 49 (3), 197-237.

Parker, D. H. SRA Reading Laboratory: IIa. Science Research Associates, Inc., Chicago, Illinois, 1958.

Parker, D. H. SRA Reading Laboratory: IIb. Science Research Associates, Inc., Chicago, Illinois, 1960.

Parker, D. H. SRA Reading Laboratory: IIb. Science Research Associates, Inc., Chicago, Illinois, 1963.

Parker, D. H. SRA Reading Laboratory: IIc. Science Research Associates, Inc., Chicago, Illinois, 1960.

Parker, D. H. SRA Reading Laboratory: IIIa. Science Research Associates, Inc., Chicago, Illinois, 1964.

Parker, D. H. SRA Reading Laboratory: IVa. Science Research Associates, Inc., Chicago, Illinois, 1959.

Potter, Thomas C. A Taxonomy of Cloze Research, Part I: Readability and Reading Comprehension. Southwest Regional Laboratory.

Powell, William R. "Reappraising the Criteria for Interpreting Informal Inventories." Unpublished paper, Thirteenth Annual Convention, I.R.A., Boston, Massachusetts, 1968.

Predicting readability. Teacher's College Record, 1944, 45, 404-419.

Rankin, E. F. An evaluation of the cloze procedure as a technique for measuring reading comprehension. Unpublished doctoral dissertation, University of Michigan, 1957.

Rankin, E. F. Uses of the cloze procedure in the reading clinic. In J. A. Figurel (Ed.) IRA Conference Proceedings, New York. Scholastic Magazine, 1959, 4, 228-232.

Rankin, E. F. The cloze procedure--Its validity and utility. In O. S. Causey and W. Eller (Eds.), Eighth Yearbook of National Reading Conference, National Reading Conference, Inc., 1959, 8, 131-144.

Rankin, E. F. The cloze procedure--A survey of research. In E. L. Thurston and L. E. Hafner (Eds.) The philosophical and sociological bases of reading. Fourteenth Yearbook of the National Reading Conference, Milwaukee, Wisconsin, The National Reading Conference, Inc., 1965, 133-150.

Rankin, E. F. Residual gain as a measure of individual differences in reading improvement. Journal of Reading, 1965, 10, 224-233.

Rankin, E. F. Research design and the cloze procedure. Proceedings of the International Reading Association Convention, 1966, 2 (1), 489-491.

Ruddell, R. B. The effect of oral and written patterns of language structure on reading comprehension. Unpublished doctoral dissertation, University of Indiana, 1963.

Ruddell, R. B. A study of the cloze comprehension technique in relation to structurally controlled reading material. In J. A. Figurel (Ed.) Improvement of reading through classroom practice. International Reading Association Conference Proceedings, 1964, 9, 298-306.

Ruddell, R. B. The effect of oral and written patterns of language structure on reading comprehension. Reading Teacher, 1965, 18, 270.

Salzinger, K., Portnoy, S., & Feldman, R. S. The effect of order of approximation to the statistical structure of English on the emission of verbal responses. Journal of Experimental Psychology, 1962, 64, 52-57.

Schneyer, J. W. Use of the cloze procedure for improving reading comprehension. Reading Teacher, 1965, 19, 174-179.

Spache, G. D. A new readability formula for primary grade materials. Elementary School Journal, 1952, 53, (7), 410-413.

Spache, G. D. Diagnostic Reading Scales, Harcourt, Brace & World.

Spache, G. D. Reading in the Elementary School. Boston: Allyn and Bacon, Inc., 1969, 516pp.

Strickland, R. G. The language of elementary school children: Its relationship to the language of reading textbooks and the quality of reading of selected children. Bulletin of the School of Education, University of Indiana, 1962, 38 (4).

Taylor, W. L. Cloze procedure: A new tool for measuring readability. Journalism Quarterly, 1953, 30, 414-438.

Taylor, W. L. Recent developments in the use of the cloze procedure. Journalism Quarterly, 1956, 33, 42-48.

Taylor, W. L. Cloze readability scores as indices of individual differences in comprehension and aptitude. Journal of Applied Psychology, 1957, 41, 19-26.

Thorndike, E. L. Reading and Reasoning: A study of mistakes in paragraph reading. Journal of Educational Psychology, 1917, 8, 323-332.