A COMPARISON OF CLOZE ABILITY IN DEFICIENT AND
NON-DEFICIENT READERS MATCHED ACCORDING TO
VERBAL ABILITY

THESIS

Presented to the Graduate Council of the
North Texas State University in Partial
Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

BY

Helen V. Berrier, B. S.
Denton, Texas
May, 1974
Berrier, Helen V., A Comparison of Cloze Ability in Deficient and Non-deficient Readers Matched According to Verbal Ability. Master of Science (School Psychology), May, 1974, 20 pp., 3 tables, 3 illustrations, references, 18 titles.

The present study was designed to investigate whether a good reader, by the fifth grade, will have attained sufficient knowledge of the language structure to enable him to more exactly and more appropriately reconstruct mutilated texts than a poor reader, matched for verbal intelligence level. Four 250-word cloze-treated passages were administered to twelve deficient and twelve non-deficient sixth grade readers, matched according to sex and the verbal portion of the Wechsler Intelligence Scale for Children. Analyses of variance failed to show any significant differences between good and poor readers except for a weak indication that good readers produced more exact replacements.
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Means and Standard Deviations (SD's) of Cloze Passages Scored for Incorrect Replacements</td>
<td>9</td>
</tr>
<tr>
<td>2. Means and Standard Deviations (SD's) of Cloze Passages Scored for Items Left Blank</td>
<td>11</td>
</tr>
<tr>
<td>3. Means and Standard Deviations (SD's) of Cloze Passages Scored for Exact and Synonym Replacements</td>
<td>12</td>
</tr>
</tbody>
</table>
# LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Interaction between Experimental Groups and Difficulty Level of Passages</td>
<td>10</td>
</tr>
<tr>
<td>2.</td>
<td>Interaction between Scoring Methods and Difficulty Level of Passages</td>
<td>13</td>
</tr>
<tr>
<td>3.</td>
<td>Interaction between Experimental Groups and Scoring Methods</td>
<td>15</td>
</tr>
</tbody>
</table>
A COMPARISON OF CLOZE ABILITY IN DEFICIENT AND NON-DEFICIENT READERS MATCHED ACCORDING TO VERBAL ABILITY

Taylor (1953) introduced the cloze procedure as a means of investigating Luria's (1966) "guessing reading"--the ability to make use of the contextual constraints of language. Closure is a tendency to complete a familiar but not-quite-finished pattern. A cloze unit is any single occurrence of a successful attempt to reproduce accurately a part deleted from a "message" (any language product) by deciding from the context which remains what the missing part should be. The cloze procedure is a method of intercepting a message from a "transmitter" (writer or speaker), mutilating its language parts, and so administering it to "receivers" (readers or listeners) that their attempts to make the patterns whole again yield a number of cloze units.

Much of the research on the cloze procedure has involved methodological considerations and the degree to which performance on cloze tests is related to performance on tests of certain cognitive abilities. The following issues were the main concerns of investigations dealing with the methodology of the cloze procedure:

(1.) As for the construction of cloze passages, random and every-nth deletion systems are, in the long run, equivalent (Taylor, 1953), with an every-fifth word deletion
system spacing words adequately (Bickley, Ellington, & Bickley, 1970). An any-word (or mechanical) deletion of words is superior to a rational deletion of words based on certain word classes (Taylor, 1953, 1957).

(2.) The scoring of exact replacements of deleted words is effectively equivalent to the scoring of synonyms (Rankin, 1959). However, alternating scoring methods (Bickley et al., 1970) or including both measures (Hargis, 1972) can lend insight in certain situations involving creative performance (Hargis, 1972).

(3.) With regard to the number and length of cloze passages required to construct an adequate test, the number of subjects being tested must be considered (Bormuth, 1965). Generally, a fifty-item test provides a sufficient sample for stable scores (Taylor, 1956).

(4.) Elementary age children have been used in cloze research (Bormuth, 1963) because before about the fourth or fifth grade level, children's prose is generally not written encompassing the redundancy characteristics of "average" adult prose with regard to the structural elements of the language (Carterette & Jones, 1963).

Research dealing with the various aforementioned methodological issues was accompanied by investigations concerned with the relationships between scores on cloze tests and a variety of cognitive abilities. In a study by Froese (1971) 366 sixth graders were administered the Canadian Lorge-Thorndike Intelligence Test and the Canadian Test of Basic
Skills, vocabulary and comprehension subtests. Passages from twelve science textbooks were subjected to the cloze technique and administered to subjects. The data yielded high correlations between cloze scores and the intelligence measure, between vocabulary subtest scores and cloze scores, and between comprehension subtest scores and cloze scores.

Lisman (1971) divided fifty-seven seventh and sixty eighth graders into three reading ability groups. All were given the Wechsler Intelligence Scale for Children and the Gates-MacGinitie Reading Test, Survey E for grades seven through nine immediately after the experimental procedure. A practice sample was given before the test on prepared vowel deletion and cloze passages. The data yielded relatively strong correlations (.46 to .60) between vocabulary and cloze scores, between comprehension and cloze scores, between full scale IQ and cloze scores, and between verbal IQ and cloze scores.

Hafner (1963) obtained a correlation of .73 between an any-word deletion test and the Otis Quick Scoring Mental Ability Test and a correlation of .56 between the same cloze test and the information subtest of the Wechsler-Bellevue Intelligence Scale. On the other hand, Weaver and Kingston (1963) have suggested that scores are affected by a special aptitude or ability for utilizing redundancy in a passage and supplying missing elements, independent of verbal ability.

The cloze procedure has also been used to investigate differential reading ability, especially among elementary
age children. The purpose of a study by Rankin and Overholser (1969) was to determine the relationship between the utilization of each of thirteen different types of contextual clues and reading ability (lower third, middle third, and upper third) within each of three grade levels. Subjects were 208 fourth, fifth, and sixth graders. A context test was constructed using the cloze procedure and deleting nouns, adjectives, adverbs, and verbs from contexts. Passages were taken from two current fourth grade basal readers or based on modification of passages from the basal readers. Exact words or acceptable synonyms (correct tense and grammatical structure) were scored as correct. The results indicated that reading ability was predictive of the ability to use each of the thirteen clues, with some clues more effective than others.

Swalm (1972) investigated the effect of mode of presentation on cloze scores for children with varying reading abilities. Three hundred twenty-four students (108 from the second, third, and fourth grades) were assigned to the three testing methods randomly (oral reading, silent reading, and listening to tape) with 36 subjects in each testing group from each grade (12 each from above-average, average, and below-average reading ability groups). Materials used included a separate story for each grade level which had been subjected to a cloze random deletion procedure (ten per cent), using only lexical words. The results supported the author's conclusion that when the reading level of the individual is above the difficulty level of the article, he will comprehend
more by reading the article himself rather than listening to someone else read it. When the difficulty level is approximately the same as the individual's reading level, the three methods are equal in effectiveness. But when the difficulty level of the article surpasses the subject's reading level, listening is the preferred mode of presentation, especially over silent reading.

Hargis (1972) used fifteen mentally retarded and fifteen non-mentally retarded children between the ages of ten and twelve years to compare the abilities of the children to use context in reading. Materials were 200-word cloze-treated passages with every-fifth word deleted and replaced with a standard (fourteen pica spaces) sized blank. The difficulty level of each passage corresponded to the instructional reading level of each subject which was determined by means of an informal reading inventory. The instructional reading level was defined as the earliest grade level at which the individual made no more than two to four per cent pronunciation errors on oral reading and attained at least seventy-five per cent comprehension. Passages were matched for readability with the Spache Readability Formula (pre-primer through third grade level) and the Dale-Chall Readability Formula (fourth grade through ninth grade level). No time limit was imposed and subjects who did not understand the instructions were given a demonstration of the task. Scoring methods included counting the number of appropriate words subjects were able to supply for those deleted, appropriate words plus exact
words supplied, and exact words only. Non-mentally retarded children were significantly more able to use context in reading, and the ability to supply grammatically correct synonyms was found to be a function of intelligence level. The mentally retarded children, however, were not deficient in the ability to supply exact words.

Evidence has been cited which links cloze ability to both cognitive and reading abilities. If cloze ability is a cognitive ability related to intelligence, then differences found between good and poor readers are reflections of intellectual differences and not at all indicative of differential reading ability. Closure would become no more than another indicator of intelligence. But if closure is a specific reading ability, then good and poor readers of the same intellectual-verbal ability should perform differently on cloze tests. Hence, it is unclear as to whether deficient and non-deficient readers matched on the basis of verbal intelligence level will perform similarly in reconstructing mutilated texts. The present study will investigate whether a good reader, by the fifth grade, will have attained sufficient knowledge of the structural elements of language to enable him to more exactly and more appropriately reconstruct mutilated texts, the most difficult of which matches or exceeds his instructional reading level, than a poor reader of the same verbal intelligence level.
Method

Subjects

Twelve deficient and twelve non-deficient readers enrolled in a regular sixth grade class and a combination fifth and sixth grade class served as subjects. The subjects were matched according to sex and the verbal portion of the Wechsler Intelligence Scale for Children (range: from 84 to 118). Reading ability level was determined by the individual teachers, one of whom used the Stanford Diagnostic Reading Scales, while the other used an informal reading inventory. The range in reading level varied from second grade, eighth month to seventh grade, eighth month. Subjects who were reading below the fifth grade level were classified as deficient readers (mean reading level: third grade, fourth month), while subjects who were reading above the fifth grade level were classified as non-deficient readers (mean reading level: fifth grade, seventh month). Experimental groups included seven girls and five boys each. The mean age of the deficient readers was eleven years, seven months, and the mean age of the non-deficient readers was eleven years, six months.

Materials

Four 250-word cloze-treated passages (every-fifth word deleted and replaced with a ten-space blank), tested for readability with the Spache Readability Formula for grades one through three (Spache, 1970) and the Dale-Chall Readability Formula for grades four through seven (Dale & Chall, 1948),
were constructed from a variety of basal readers and library materials. Passages at the second, fourth, fifth, and seventh grade levels were chosen. The passages were arranged in order of increasing difficulty.

Procedure

All subjects except one were administered the test passages in a group during a one-hour session. The subject who was absent on the day of the group administration was administered the test passages on the following day. The following instructions were read to the subjects by the examiner:

This test is to see how well you can fill in words that have been left out of stories. I want you to fill in the blanks you find with the word which you think fits best and makes the story sound right. If you need to write in a number or the name of something, make up a number or name and write it in. Do the best you can. If you don't know how to spell a word, try the best you can. You won't be marked wrong if I can tell which word you mean. If you don't know what a word is, just skip over it and go on. Try to fill in all the blanks, but if you can't think of a word to put in a particular blank, leave it empty and go on and try to finish the story. Please write or print each word plainly so that I'll be able to read it.

Two sample sentences were presented to the group at this point. The instructions then continued as follows:

Every so often I'll let you know that you need to start working on the next story if you haven't already started on it. If you finish a story before I give you the time, go on. Don't wait for me. You will have about fifty minutes to finish all four stories. If you finish the whole test before the time is up, go back and try to fill in blanks you may have left empty. In any case, please sit quietly and doodle or draw pictures on the blank paper in your booklet.

Are there any questions? When I say, "Begin," open your test booklets and start to work. I'll let you know when the time is up. Ready? Begin.
Scoring

Four scores were determined for each subject: number of incorrect replacements (not members of the same word-form class), number of exact replacements, number of grammatically correct synonyms (members of the same word-form class), and number of items left blank.

Results

Incorrect Replacements

The number of incorrect replacements for the deficient and non-deficient readers at each level of passage difficulty were analyzed by a two-by-four analysis of variance. The factors of the analysis were deficient versus non-deficient readers and the four levels of passage difficulty (repeated measures on the second factor). Table 1 reports the means and standard deviations of each passage for the deficient and non-deficient readers.

TABLE 1
MEANS AND STANDARD DEVIATIONS (SD'S) OF CLOZE PASSAGES SCORED FOR INCORRECT REPLACEMENTS

<table>
<thead>
<tr>
<th>Passages (Grade level equivalent)</th>
<th>Deficient Readers</th>
<th>Non-deficient Readers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Means</td>
<td>SD's</td>
</tr>
<tr>
<td>I (Second grade)</td>
<td>1.4167</td>
<td>.9962</td>
</tr>
<tr>
<td>II (Fourth grade)</td>
<td>6.6667</td>
<td>7.0108</td>
</tr>
<tr>
<td>III (Fifth grade)</td>
<td>11.8333</td>
<td>5.4076</td>
</tr>
<tr>
<td>IV (Seventh grade)</td>
<td>15.9167</td>
<td>7.5373</td>
</tr>
</tbody>
</table>
The analysis indicated that the deficient and non-deficient readers were not significantly different in the number of incorrect replacements, $F(\text{df}=1/22)=2.12470$, $p=0.15907$. Difficulty level of the passages significantly affected the number of incorrect replacements, $F(\text{df}=3/66)=78.30249$, $p<.001$. Newman-Keuls analyses of the mean scores of the passages revealed significant differences among all passages ($p<.01$). Although the interaction of experimental groups and difficulty level of the passages did not reach statistical significance, $F(\text{df}=3/66)=2.25518$, $p=0.09011$, a tendency toward significance appeared to be present, which is reflected in Figure 1.

Fig. 1--Graph denoting slight, but nonsignificant interaction between experimental groups and difficulty level of passages scored for mean number of incorrect replacements.
The pattern depicted in Figure 1 indicates that at the lowest level of difficulty non-deficient readers gave approximately the same number of incorrect replacements as deficient readers, while at the higher levels of difficulty non-deficient readers gave fewer incorrect replacements than the deficient readers.

**Items Left Blank**

A two-by-four analysis of variance (deficient versus non-deficient readers X four levels of passage difficulty with repeated measures on the second factor) revealed that neither reading ability, $F(df=1/22)=0.93178$, $p=0.34490$, nor difficulty level of the passage, $F(df=3/66)=0.73877$, $p=0.53269$, significantly influenced the number of items left blank by the subjects. Table 2 reports the means and standard deviations of each passage for the deficient and non-deficient readers.

**TABLE 2**

**MEANS AND STANDARD DEVIATIONS (SD'S) OF CLOZE PASSAGES SCORED FOR ITEMS LEFT BLANK**

<table>
<thead>
<tr>
<th>Passages</th>
<th>Deficient Readers</th>
<th>Non-deficient Readers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Means</td>
<td>SD's</td>
</tr>
<tr>
<td>I (Second grade)</td>
<td>4.1667</td>
<td>6.0578</td>
</tr>
<tr>
<td>II (Fourth grade)</td>
<td>2.1667</td>
<td>3.3530</td>
</tr>
<tr>
<td>III (Fifth grade)</td>
<td>1.7500</td>
<td>2.4909</td>
</tr>
<tr>
<td>IV (Seventh grade)</td>
<td>3.5833</td>
<td>4.6409</td>
</tr>
</tbody>
</table>
The interaction of experimental groups and difficulty level of the passages was also not significant, \( F(df=3/66)=1.24639, \) \( p=0.30003. \)

**Exact Replacements and Synonym Replacements**

The number of exact replacements and the number of grammatically correct synonyms for the deficient and non-deficient readers at each level of passage difficulty were analyzed by a two-by-two-by-four analysis of variance. Table 3 reports the means and standard deviations derived from the analysis.

**TABLE 3**

**MEANS AND STANDARD DEVIATIONS (SD'S) OF CLOZE PASSAGES SCORED FOR EXACT AND SYNONYM REPLACEMENTS**

<table>
<thead>
<tr>
<th>Scoring Method</th>
<th>Deficient Readers</th>
<th>Non-deficient Readers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Means SD's</td>
<td>Means SD's</td>
</tr>
<tr>
<td><strong>Passage I: Second Grade Level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exact</td>
<td>29.3333 6.4854</td>
<td>31.9167 3.9418</td>
</tr>
<tr>
<td>Synonym</td>
<td>15.0833 2.4310</td>
<td>15.1667 3.7823</td>
</tr>
<tr>
<td><strong>Passage II: Fourth Grade Level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exact</td>
<td>18.1667 5.0960</td>
<td>22.5000 6.1570</td>
</tr>
<tr>
<td>Synonym</td>
<td>23.0000 3.4156</td>
<td>21.3333 3.4924</td>
</tr>
<tr>
<td><strong>Passage III: Fifth Grade Level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exact</td>
<td>14.5833 4.5817</td>
<td>18.6667 5.8672</td>
</tr>
<tr>
<td>Synonym</td>
<td>21.8333 1.7240</td>
<td>21.5833 4.0095</td>
</tr>
<tr>
<td><strong>Passage IV: Seventh Grade Level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exact</td>
<td>13.6667 4.6188</td>
<td>18.6667 6.0503</td>
</tr>
<tr>
<td>Synonym</td>
<td>16.8333 3.6931</td>
<td>17.0000 2.8866</td>
</tr>
</tbody>
</table>
The factors of the analysis were deficient versus non-deficient readers, exact versus synonym replacements (repeated measures), and the four levels of passage difficulty (repeated measures). The interaction between exact replacements versus synonym replacements and the difficulty level of the passages was significant, $F(df=3/66)=101.7410, p<.001$. Figure 2 depicts the nature of the interaction.

![Graph](image)

**Fig. 2**—Graph denoting interaction between scoring methods and difficulty level of passages.

Figure 2 shows that more exact replacements than synonyms were provided at the lowest level of difficulty. As the difficulty level increased, more synonyms than exact replacements
were given. At the highest level of difficulty the number of synonym replacements declined with incorrect replacements (see Figure 1) representing the bulk of the replacements.

Simple effects analyses of the means of the two scoring methods at each level of difficulty revealed significant differences at both the second grade level of difficulty (Passage I), $F(df=1/236)=113.12617$, $p < .01$, and the fifth grade level of difficulty (Passage III), $F(df=1/236)=8.050613$, $p < .01$. Significant differences were not found at the fourth grade level of difficulty (Passage II), $F(df=1/236)=1.582508$, $p > .25$, or at the seventh grade level of difficulty (Passage IV), $F(df=1/236)=0.2648635$, $p > .25$.

Simple effects analyses of the means of each scoring method across passages yielded significant differences for both exact replacements, $F(df=3/236)=153.23861$, $p < .01$, and synonym replacements, $F(df=3/236)=36.949269$, $p < .01$. Newman-Keuls analyses of the mean scores for exact replacements at each level of passage difficulty revealed the following differences among passages. The mean score of Passage I for exact replacements was significantly different from the mean scores of Passages II, III, and IV ($p < .01$). Passage II was likewise significantly different from Passages III and IV ($p < .01$). The mean scores of the two most difficult passages (fifth and seventh grade levels) did not differ significantly ($p > .05$). Newman-Keuls analyses of the mean scores for synonym replacements at each level of passage difficulty revealed the following differences among passages.
The mean score of Passage I for synonym replacements was significantly different from the mean scores of Passages II, III, and IV \((p < .01)\). Passage II was different from Passage IV \((p < .01)\), but not from Passage III \((p > .05)\). The two most difficult passages (III and IV) were significantly different \((p < .01)\).

Although the interaction between exact replacements versus synonym replacements and deficient versus non-deficient readers did not reach statistical significance, \(F(\text{df}=1/22)=3.2556, p=0.08489\), a tendency toward significance appeared to be present, which is reflected in Figure 3.

![Graph](image_url)

**Fig. 3**--Graph denoting slight, but nonsignificant interaction between experimental groups and scoring methods.
The pattern depicted in Figure 3 suggests that non-deficient readers were more able to give exact replacements than deficient readers. However, the number of synonym replacements was approximately the same for both experimental groups.

The analysis of variance revealed a slight, but nonsignificant difference between the two experimental groups, $F(\text{df}=1/22)=3.1170, p=0.09150$. Subjects were not found to differ in their tendencies to supply exact versus synonym replacements, $F(\text{df}=1/22)=2.5602, p=0.12385$. The difficulty level of the passages, though, significantly influenced the subjects' abilities to reconstruct the mutilated texts, $F(\text{df}=3/66)=86.8065, p<.001$. Neither the interaction between passages and experimental groups, $F(\text{df}=3/66)=1.0259, p=0.38688$, nor the three-way interaction of experimental groups, scoring methods, and passages, $F(\text{df}=3/66)=0.6353, p=0.59490$, was significant.

**Discussion**

Apparently, closure cannot be investigated as a specific reading ability without taking verbal ability into consideration. Both good and poor readers of similar verbal abilities are able to reconstruct mutilated texts with the same degree of proficiency as long as that reconstruction involves a grammatical rather than a semantic reconstruction. Closure—the awareness of perceptual units in communication—for many poor readers, is not part of their problem with learning to read better. Many poor readers, as evidenced by the present
study, are aware of the redundancies in language structure. Hence, instruction designed to improve grammatical cloze abilities would probably fail to improve reading abilities.

The small sample size used in the present study encourages some caution in the interpretation of results. Bormuth (1965) estimated the standard error of the mean for various combinations of test length and numbers of subjects. Bormuth's data confirm the proposition offered by Taylor (1956) that a fifty-item test provides a sufficient sample for stable scores with the chances for the occurrence of easy and hard words canceling out, especially when administered to a sample of forty or fifty subjects. In the light of such findings, the possibilities of committing a number of Type II errors based on the data at hand are quite considerable. (The present study's samples of deficient and non-deficient readers numbered only twelve subjects for each experimental group--twenty-four subjects in all--due to the difficulty in obtaining sufficient numbers from the available school.)

However, a distinction between good and poor readers appears to exist based on the ability to guess the intent, or exact meaning of the author, as suggested by the near significant difference between the deficient and non-deficient readers found when exact as opposed to synonym replacements were compared (see Figure 3). This possible difference between the deficient and non-deficient readers also disputes Rahnkin's (1959) position that the scoring of exact replacements is effectively equivalent to the scoring of synonyms. The
scoring of exact replacements appears to discriminate between good and poor readers, while the scoring of synonyms does not. Such a distinction can be thought of as tapping a comprehension ability. Instruction, then, which sought to improve semantic cloze abilities could possibly be expected to have a positive effect on reading ability.
References


Swalm, J. E. *Comparison of oral reading, silent reading, and listening comprehension assessed by cloze.* (Doctoral dissertation, Rutgers, The State University of New Jersey) Ann Arbor, Mich.: University Microfilms, 1972, No. 72-1102. *(Dissertation Abstracts, 1972, 32 (7-A), 3578-3579)*


