

## SECONDARY STUDENTS' FREE RECALL AS A MEASURE OF COMPREHENSION AND EQUIVALENCY OF CONTENT AREA PARAGRAPHS

TERRY LOVELACE

University of Southwestern Louisiana

Traditionally, the informal reading inventory (IRI) has been employed as a criterion-referenced measure of students' reading comprehension ability (Farr and Roser, 1979). The most important component of silent reading achievement evaluated is the level of difficulty of text a reader can comprehend, which can be determined more accurately via a well-constructed, well-administered IRI than by a standardized achievement test (Harris and Sipay, 1980). A student's silent reading comprehension achievement is frequently based on his ability to recall exact wording from a passage in order to answer specific literal questions. One of the most commonly-employed IRI's in university reading clinics, *The Durrell Analysis of Reading Difficulties* (Durrell, 1955) goes so far as to ask the reader to verbalize all he/she can recall from the passage. These memories are then used as a measure of silent reading comprehension.

Many reading specialists and clinicians use the IRI to evaluate students' academic achievement during a specific remediation program, utilizing paragraphs deemed "equivalent" by IRI authors to pre and posttest.

This study examines 1) the recall of exact wording from grade level equivalent IRI paragraphs by secondary students; 2) weighting of exact wording recalled by students from three content area paragraphs; and 3) the relationship between subjects (Ss) sex, grade level, serial learning ability, and exact and weighted recall of three paragraphs.

### METHODOLOGY

#### Subjects

Forty secondary students participated in this study: 20 males and 20 females selected from grade levels nine and eleven at Acadiana High School (student population, 2,300). Home-room teachers of students enrolled in the honors curriculum were asked to rank their students by grade point average; the top half of each group were pooled across classrooms. Random selection of 10 ninth grade males, 10 ninth grade females, 10 eleventh grade males, and 10 eleventh grade females was conducted using a table of random numbers. Scores from the *SRA Achievement Test* were available for the ninth grade students, who scored on the following subtests, according to the national percentile 72.11—Vocabulary Subtest, 77.58—Comprehension Subtest, and 76.05—Composite Reading Score.

#### Procedure

Ss were tested in small groups of 10 students; each of the three experimental sessions lasted approximately 45 minutes, with sessions scheduled one week apart. During the first session, students completed a knowledge inventory designed to determine self-assessed knowledge of specific topics listed by the examiner. A five-point Likert-like scale (1=I know nothing about this subject, 5=I know everything there is to know about this subject) was used to gauge the student's attitude toward how much he/she knew about eight topics, three of which were topics of the three paragraphs which the student would read later, five of which were placebo topics. The 40 students averaged 2.85 on the five-point scale for the topic "zoo animals," 1.93 for the topic "lunar exploration," and 2.40 for the topic "trees." Their responses ranged from "I know very little about this subject" to "I know something about this subject."

Ss then read one of three expository passages of 163 words (Forms A, C, and D) selected at random from the *Ekwall Reading Inventory* (Ekwall, 1979). Readability level of the passages was determined to be eighth grade level, based on the

Harris-Jacobson Readability Formula (Harris and Sipay, 1980). Passages were composed by Ekwall, and normed on 40 students. Reliability of passages, based on administering forms within a period of one week, resulted in product moment correlation coefficients of .79 between Forms B and D, and .82 between Forms A and C. The topic of the paragraph labeled Form A is lunar exploration; this paragraph will be henceforth referred to as the *moon paragraph*. The topic of the paragraph labeled Form C is the ostrich, henceforth referred to as the *ostrich paragraph*. The topic of the paragraph labeled Form D is the rubber tree, henceforth referred to as the *rubber paragraph*.

Students were presented with one of the three paragraphs, and told to read the paragraph carefully; they were cautioned to remember as much of the paragraph as possible, since they would be asked to recite all they could remember from the paragraph into a cassette tape recorder later. Students were allowed as much time as they desired to read and to study the paragraph. They were then administered a serial learning task comprised of 12 neutral words presented in four trials with words exposed for two-second intervals; the task is one developed by Bloomer (1978) as part of a learning process test battery, and has been carefully normed and validated. Students attempted to memorize the words, and then wrote down the words they remembered for each of the four trials during a 20-minute interval.

Ss then recited all they remembered from the passage into a cassette recorder. Prompts were then provided in order to cue further recall. This procedure was repeated twice at one week intervals, until all paragraphs had been read.

### Scoring

Protocols were scored twice, once for retention of exact wording, and then for retention of weighted wording. Specifically, a computer program was developed by USL computer programmer Wayne Vicknair to match exact wording from an original or source paragraph with wording from a protocol (transcribed tape recording of a Ss' recitation of what he/she had remembered from a passage). For each word in the protocol which matches the original, a score of one is indicated, while a score of zero is indicated if the computer is unable to make that match. The program prints a list of one's and zero's for each word in the source paragraph, and totals the correct matches. A cumulative frequency distribution of matches for each word in the source paragraph is also printed.

Researchers in the area of discourse analysis suggest that recall should be weighted, that a Ss' recall of a determiner (*a*, *an*, or *the*) should not be weighted as heavily, because of the frequency of occurrence due to the syntax of the English language, as verbs, subject nouns, direct objects, etc. which carry the brunt of the author's message. The Dolch 220, a list of basic sight words, comprises approximately 51% of text printed in English; due to their high frequency of appearance, they should probably not be weighted as heavily, since what characterizes the difference in meaning from one text to another is the content, which generally contains words of a more technical nature not contained in the Dolch list. The investigator, interested in weighting the protocols, simplistically decided to weight subject nouns and verbs as 6; auxiliary verbs, 5; direct objects, indirect objects, predicate nominatives, and predicate adjectives, 4; infinitives, conjunctions, prepositions, objects of prepositions, 3; adjectives and adverbs, 2; determiners, 1. While recognizably an over-simplistic method of handing weighting of exact wording, eventually hopes to develop a system for grading protocols which could be used by classroom teachers, who do not have the time necessary for the elaborate methods employed by researchers in discourse analysis for grading protocols. Vicknair's computer program provides a method of weighting individual words; the entire scoring cycle is then repeated, resulting in a score reflecting the examiner's

weighting of words in the original paragraph matched with Ss' protocols.

### Analyses

A packaged statistical program (SPSS, Nie, Hull, Jenkins, Steinbrenner, and Bent, 1970) was used to compute frequency distributions, descriptive statistics, Pearson product moment correlation coefficients, *t*-tests, and analyses of variance.

### RESULTS

On the average, Ss retained 88.45 (54.25%) of the 163 words from the rubber paragraph, 71.96 (44.14%) of the words from the ostrich paragraph, and 73.27 words (44.95%) from the moon paragraph. The mean number of correct responses to the five comprehension questions was 4.14 (rubber paragraph), 4.42 (ostrich paragraph), and 4.13 (moon paragraph). Total scores on the three repetitions of the serial learning task (possible raw score=48) were 36.25, trial one; 36.50, trial two; and 36.75, trial three. Ss recalled, on the average, 6.34 of the 12 words presented in the first serial learning task one week after the initial presentation of the words, and 9.21 of the 12 words one week following the second presentation of the words. When the exact recall of wording from paragraphs was weighted as described earlier, Ss' mean scores were 294.29, rubber paragraph; 235.36, moon paragraph; and 253.43, ostrich paragraph.

Computed Pearson product moment correlation coefficients revealed low to moderate positive correlations between Ss' scores from the three serial learning tasks and the reading achievement scores from the SRA *Achievement Test, Level F* (see Table 1). Correlations computed between Ss' SRA scores and recall of paragraphs and weighted recall of paragraphs revealed mild, statistically non-significant positive correlations (see Table 2).

Correlations between Ss' self-assessed knowledge of topics associated with the main ideas of the paragraphs and their exact recall of wording were .42, .23, .02 with moon, rubber, and ostrich paragraphs, respectively. Little correlation was noted between self-assessed knowledge of the three topics by the 40 students (.27, ostrich; .07 rubber and .28, moon).

Ss' exact recall of wording from the three paragraphs was compared via *t*-tests, which revealed statistically significant differences between recall from the rubber paragraph (mean word recalled=86.72) and the moon paragraph (mean words recalled =74.21). No statistically significant differences on computed *t*-tests were noted between recall scores from the paragraphs on the moon and the ostrich, or between recall from the paragraphs on the ostrich and rubber. These results were replicated with weighted recall of the three passages. Correlation coefficients ranging between 0.54 and 0.65 were noted between the exact recall of wording from the three paragraphs and between 0.51 and 0.67 between weighted recall from the three passages (see Table 3).

Examination of correlation coefficients revealed low to moderate positive correlations between exact recall of the three paragraphs and the serial learning tasks, weighted recall of the three paragraphs and the serial learning task, and students' correct responses to the comprehension questions (prompts) and their scores on the serial learning task (see Table 4).

Table 1.

**PEARSON PRODUCT MOMENT CORRELATION  
COEFFICIENTS BETWEEN SERIAL LEARNING AND  
SRA READING ACHIEVEMENT TEST SCORES (N=19)**

SRA Subtests	Serial Learning Task		
	Trial 1	Trial 2	Trial 3
Vocabulary	0.27	0.03	0.45*
Comprehension	0.16	0.16	0.59**
Composite	0.24	0.09	0.55**

\*p &lt; .05

\*\*p &lt; .01

Table 2.

**PEARSON PRODUCT MOMENT CORRELATION  
COEFFICIENTS BETWEEN SRA ACHIEVEMENT TEST  
SCORES IN READING AND EXACT RECALL OF  
PARAGRAPHS AND WEIGHTED RECALL OF  
PARAGRAPHS**

Exact Recall of	SRA Reading Subtests		
	Vocabulary	Compre- hension	Composite
Moon Paragraph	0.28 (N=17)	0.38 (N=17)	0.29 (N=17)
Rubber Paragraph	0.11 (N=17)	0.36 (N=17)	0.24 (N=17)
Ostrich Paragraph	0.28 (N=8)	0.04 (N=8)	0.15 (N=8)
<b>Weighted Recall of</b>			
Rubber Paragraph	0.10 (N=17)	.039 (N=17)	0.24 (N=17)
Moon Paragraph	0.25 (N=17)	0.32 (N=17)	0.24 (N=17)
Ostrich Paragraph	0.24 (N=8)	0.02 (N=8)	0.18 (N=8)

*Note: All of the above correlation coefficients are statistically significant at the .05 level.*

Table 3.

**PEARSON PRODUCT MOMENT CORRELATION  
COEFFICIENTS BETWEEN EXACT RECALL OF  
WORDING FROM THREE PARAGRAPHS AND  
WEIGHTED RECALL FROM THREE PARAGRAPHS**

Exact Recall of Wording from:	Verbatim Recall of Paragraph Wording		
	Moon Paragraph	Rubber Paragraph	Ostrich Paragraph
Moon Paragraph		0.70** (N=29)	0.65** (N=18)
Rubber Paragraph			0.54** (N=19)
Weighted Recall of Wording from:			
Rubber Paragraph	0.67** (N=29)		
Moon Paragraph			0.61** (N=18)
Ostrich Paragraph		0.52** (N=19)	

\* &lt; .05

\*\*P &lt; .01

Table 4.

**PEARSON PRODUCT MOMENT CORRELATION  
COEFFICIENTS BETWEEN Ss' EXACT RECALL,  
COMPREHENSION, AND WEIGHTED RECALL OF  
THREE PARAGRAPHS AND SCORES FROM  
CORRESPONDING SERIAL LEARNING TASKS**

Raw Scores from Serial Learning Tasks Correlated with:	Pearson r	N of Ss	Level of Significance
Recall of moon paragraph	0.29	33	.05
Recall of rubber paragraph	.057	35	.01
Recall of ostrich paragraph	0.05	23	.42
Comprehension of moon paragraph	.009	32	.32
Comprehension of rubber paragraph	0.43	35	.01
Comprehension of ostrich paragraph	0.27	24	.10
Weighted recall of moon paragraph	0.28	33	.06
Weighted recall of rubber paragraph	0.55	35	.01
Weighted recall of ostrich paragraph	0.04	23	.43

Two-way analyses of Ss' raw scores from the serial learning task (trials one, two and three) by sex by grade reveal statistically significant differences ( $p < .05$ ) by grade level on trials one and two, with means higher for ninth grade students. No statistically significant differences were noted between sex and serial learning scores, or between grade level and serial learning scores from trial three.

Pearson product moment correlation coefficients between exact recall of paragraphs, weighted recall of paragraphs, and students' comprehension scores for the paragraphs are presented in Table 5. The correlation coefficients are positive, but low to moderate between students' exact recall and their comprehension of the questions, and show moderate to high statistically-significant relationships between comprehension scores and weighted recall.

One-way analyses of variance revealed no statistically significant differences between sex of Ss and Ss':

- 1) comprehension scores from the three paragraphs;
- 2) exact recall of wording from the three paragraphs;
- 3) weighted recall of wording from the three paragraphs;

However statistically significant differences exist between ninth and eleventh grade students:

- 1) comprehension scores from the three paragraphs;
- 2) exact recall of wording from the three paragraphs;
- 3) weighted recall of wording from the three paragraphs.

Grade level differences favored ninth grade students; their mean scores were higher in all of the above three analyses.

Table 5.

**PEARSON PRODUCT MOMENT CORRELATION  
COEFFICIENTS BETWEEN Ss' RESPONSES TO  
RELATED COMPREHENSION QUESTIONS AND  
THEIR EXACT RECALL AND WEIGHTED RECALL OF  
THREE PARAGRAPHS**

Related Comprehension Scores Correlated with:	Pearson $r$	N of Ss	Level of Significance
<b>Exact Recall of:</b>			
Moon Paragraph	0.62	32	.01
Rubber Paragraph	0.32	35	.03
Ostrich Paragraph	0.28	23	.10
<b>Weighted Recall of:</b>			
Moon Paragraph	0.99	33	.01
Rubber Paragraph	0.99	35	.01
Ostrich Paragraph	0.98	23	.01

## CONCLUSIONS

Secondary students retained between 44 and 54% of the exact wording from three eighth grade level paragraphs, a much higher percentage of retention than that reported by McKoon and Keenan (reported in Kintsch, 1974; McKoon, 1977), whose Ss recalled only 25% of the exact wording of passages read. Lovelace (1980), employing the same passages used in this study, noted that adults aged 20-39 retained a mean of 37.4 (25%) of the exact wording of the paragraphs. Statistically significant differences in recall of exact wording were based on the higher mean scores of ninth grade students who also scored significantly better than the eleventh grade students on two of the three serial learning task trials. Further investigations might help determine if older students and adults develop the ability to increase memory for the gist of a paragraph with a resulting decline in verbatim recall, or whether they

develop a cautiousness factor, similar to that noted in adult cognitive studies, which makes them more reluctant to vocalize information if they are not absolutely certain that the information is appropriate.

While Ekwall reported a reliability coefficient of .79 between the rubber and moon paragraphs, statistically significant differences were noted in Ss' ability to recite exact wording from the passages (mean recall of wording from rubber paragraph = 88.43, mean recall of wording from moon paragraph = 73.27). Significant differences were also noted in Ss' comprehension of the prompts provided after the free recall. Again, ninth grade Ss' mean scores appear to be the source of the difference.

This study suggests that the three paragraphs, equivalent in readability level according to the Harris-Jacobson formula, are not equivalent for ninth grade Ss, as far as verbatim recall of wording, weighted recall of wording, and comprehension scores are concerned. Pre- and post-testing using these equivalent paragraphs are recommended as a means of determining these students' gains in reading achievement over a period of time.

Serial learning apparently is not a primary factor in reading achievement as measured by standardized reading tests, which may be due to the differing demands made on the reader by the two tasks. Standardized reading achievement tests generally allow the reader to search the passage for the correct answer to a multiple choice item, whereas recall of wording demands that the reader search long-term memory and discriminate among pieces of information in order to retrieve the correct information. Evidence from this study adds credence to the existence of differing demands placed on the reader by the two types of tasks.

Correlation coefficients between what students thought they knew about the paragraph topics and their actual recall are low, perhaps indicating an element of cautiousness on the part of the Ss, who may not have been willing to self-assess their knowledge, or who may have been unsure of just what particular concepts they were being asked to evaluate. A semantic differential technique will be employed in the replication of this study.

Weighted verbatim recall of passages correlated between .98 and .99 with Ss' comprehension of passages as measured by answers to comprehension questions (prompts), whereas exact recall correlated with comprehension resulted in coefficients of .28 to .62. Weighting of exact recall of wording appears to be a more accurate means of measuring students' understanding of prose.

## REFERENCE NOTES

- Mathews, II, S. R. & Campbell, K. Comprehension research: Beyond single dependent measures. Paper presented at the American Reading Conference, Sarasota, Florida, 1980.
- Lovelace, T. Adult Recall as a Measure of Comprehension. Paper presented at the American Reading Conference, Sarasota, Florida, 1980.

## REFERENCES

- Anderson, J. R. & Paulson, R. Representation and retention of verbatim information. *Journal of Verbal Learning and Verbal Behavior*, 1977, 16, 439-452.
- Bloomer, R. *Learning Test*. New York: Brador, 1979.
- Durrell, D. D. *Durrell Analysis of Reading Difficulty*. New York: Harcourt, Brace, Jovanovitch, 1955.
- Ekwall, E. E. *Ekwall Reading Inventory*. Boston: Allyn and Bacon, Inc., 1979.
- Farr, R., & Roser, N. *Teaching a child to read*. New York: Harcourt, Brace, Jovanovitch, Inc., 1979.
- Gagne, E. D. Long term retention of information following learning from prose. *Review of Educational Research*, 1978, 48, 629-665.
- Harris, A. J., & Sipay, E. R. *How to increase reading ability, 7th Ed*. New York: David McKay Co., Inc., 1980.

Kintsch, W. On modeling comprehension. *Educational Psychologist*, 1979, 14, 3-14.

Kirsh, I., & Guthrie, J. T. The concept and measurement of functional literacy. *Reading Research Quarterly*, 1977-1978, 13, 485-507.

Klatzky, R. L. *Human Memory: Structures and Processes*, 2nd Ed. San Francisco: W. H. Freeman and Company, 1980.

McKoon, G. Organization of information in text memory. *Journal of Verbal Learning and Verbal Behavior*, 1977, 16, 247-260.