AN ANALYSIS OF ERROR PATTERNS, RATES AND GRADE EQUIVALENT SCORES ON SELECTED READING MEASURES AT THREE LEVELS OF PERFORMANCE

By

MARGERY LOIS BERENDS

Bachelor of Arts
Bachelor of Science
Kansas State Teachers College
Emporia, Kansas
1950

Master of Arts
University of Colorado
Boulder, Colorado
1958

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF EDUCATION
July, 1971

AN ANALYSIS OF ERROR PATTERNS, RATES AND GRADE EQUIVALENT SCORES ON SELECTED READING MEASURES AT THREE LEVELS OF PERFORMANCE

Thesis Approved:

Darrel D. Ray
Thesis Adviser

Bernard R Belden

Robert L Brown

Julia L. McHale

Odella Lohmann

D Dunham
Dean of the Graduate College

ACKNOWLEDGMENTS

The writer wishes to express special appreciation to Dr. Bernard R. Belden under whose direction this graduate program was begun. Without his encouragement and assistance, this dissertation would never have been undertaken.

Special appreciation is extended, also, to Dr. Darrel D. Ray, chairman of the writer's Advisory Committee, for his immeasurable assistance and encouragement throughout this study.

To the other members of the advisory committee, Dr. Bernard R. Belden, Dr. Julia McHale, Dr. Idella Lohmann, and Dr. Robert S. Brown, appreciation is expressed for their assistance.

To Dr. Dale Williams, Bi-State Mental Health Foundation, to curriculum directors, principals, fourth-grade classroom teachers of Kay County, Oklahoma, and to the fourth-grade students who participated in this study, special thanks are extended.

My sincerest appreciation is expressed to my friends and colleagues, especially Dr. Rita Stuever, Dr. Bettie Vanice, and Mrs. Karen McKellips for their patience and assistance during the completing of this study.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

PRESENTATION OF THE PROBLEM

Introduction

In classrooms, remedial reading programs, and clinics, oral reading tests are used, alone or in combination with silent reading measures, to determine the reading grade levels of boys and girls who are having some kind(s) of reading problem(s). The grade scores thus obtained are used, typically, to indicate suitable levels of difficulty of instructional materials. Eligibility for special reading classes frequently is determined by the discrepancy between the reading grade and some expected level of achievement. Rates of oral reading are frequently computed from the same reading measures.

For nearly half a century, leaders in the field of reading have advocated the analysis of oral reading errors to determine the reading strategies utilized by the pupil and/or pinpoint instructional deficiencies. Although the analyzer cannot be certain what has taken place in any one miscue, it is assumed that ". . .the patterns which emerge produce a picture in depth of the reading process in the reader." (Goodman, 1969, p. 19).

Need for the Study

Among the most commonly used instruments for securing reading-grade equivalents, rates of oral reading, and oral reading errors are

the Durrell Analysis of Reading Difficulty, the Gates-McKillop Reading Diagnostic Tests, and informal reading inventories. A different procedure for obtaining oral reading grade scores is utilized by each of the three instruments. In this study, the Standard Reading Inventory, which is a kind of "standardized" informal, is used to represent the informal instrument.

On the Durrell, the grade score is determined by the time required for reading each of the paragraphs in the test on which fewer than seven errors are made. The number of oral reading errors is used only to designate the appropriate paragraphs to be read. Comprehension is checked but does not directly affect the grade score secured.

On the Gates-McKillop, the grade score is based entirely upon the word recognition accuracy. A raw score is obtained on each paragraph dependent upon the number of errors made in reading it. The reader is penalized, point-wise, for inaccurate reading but the number of errors permitted in any one paragraph before testing is terminated is greater for this test than for either of the other two. The grade score, based on the total of the paragraph raw scores, is derived from a table of norms. Comprehension is not checked.

The third type of test is scored for accuracy of oral reading, also (although the comprehension performance normally receives equal weight). The instructional level is considered to be a range, with the maximum level being the level of the highest story in which the word recognition (and comprehension) standards are met. The per cent of oral reading accuracy required for success at a given story level is higher for this instrument than for either the Durrell or the Gates-McKillop.

The literature suggests that disabled readers are more inaccurate

than normal readers of comparable reading level (Monroe, 1928; Packman, 1970). There is some evidence that the rate of oral reading is slower for poor readers than for good readers of the same age or reading ability (Watkins, 1953; Packman, 1970); however, it seems feasible that disabled readers may be less penalized, thus obtaining relatively higher grade scores, on the Durrell which is scored for time than on the Gates-McKillop and the Standard Reading Inventory which are scored for reading accuracy. Furthermore, the tendency of poor readers to make errors on easy words as frequently as on more difficult ones (Packman, 1970) suggests that they may do relatively better on words in isolation tests than on tests of contextual material.

To determine the relationship between these measures, this study will compare the reading grade equivalents of disabled readers on the oral reading passages of the Durrell, the Gates-McKillop, and the Standard Reading Inventory. Comparisons will also be made with flashed and untimed word-list scores and with an isolated-word test in which the words are written horizontally, more nearly like contextual material.

Oral reading rate appears to vary according to the difficulty of the material (McCracken and Mullen, 1970). Since a greater error ratio within the range of acceptable performance is permitted on one test than another, the rate of oral reading obtained for a given child may depend upon the instrument being used as well as the level of performance being evaluated. This supposition will be investigated in the study by comparing the reading rates on the three reading instruments listed above. Reading rates at three levels of reading performance will also be compared.

For an error analysis to be valid, the assumption must be made that

the pupil's oral reading error pattern reflects his individual reading strengths and weaknesses, and that, within a given piece of material, the probability of one error-type's being made is equal to that of another. There is some evidence, however, that error patterns may be affected by the difficulty of the material (Schale, 1964; Christenson, 1966), the sentence structure of the material and its similarity to the reader's speech patterns (Goodman, 1969; Nurss, 1970), and the opportunity for making certain kinds of errors (Gates, 1947; Bennett, 1942; Payne, 1930).

The effect of materials and difficulty-level on the error patterns obtained by disabled readers will be explored in this study through an analysis of the errors made on the oral reading passages of each of the reading instruments at three levels of reading performance by means of an error analysis and a subsequent comparison of error patterns between tests and between levels of performance.

The major studies comparing the reading levels, rates, and patterns of errors on oral reading tests have been done with randomly-selected groups or total-school populations. Since diagnostic reading tests are used primarily with children who have reading problems, a study of the comparability of diagnostic information from representative oral reading instruments for a disabled-reader population appears to be warranted.

Statement of the Problem

The purpose of this study was to examine the reading performance of disabled fourth grade readers to determine: (1) if there are significant differences in the grade-equivalent scores obtained by disabled

fourth grade readers on nine oral reading measures; and (2) if there are significant differences in the rates of reading and prevailing error patterns of disabled fourth grade readers on three oral reading tests and at three levels of reading performance on a single test.

Hypotheses

The hypotheses to be tested are stated in the null form as:

1. There are no significant differences among the mean grade-equivalent scores on the nine oral reading measures.

2. The mean oral reading rates do not differ significantly between tests or between levels of oral reading performance on a single test. (This hypothesis will be examined separately for intertest and intratest comparisons.)

3. The mean proportion of each error-type or category does not vary significantly between tests or between levels of oral reading performance on a single test. (This hypothesis will be examined separately for each of the following categories or sub-categories on both intertest and intratest comparisons: visual perception errors, visual-auditory errors, directional confusion errors, syllabic division errors, structural errors, corrections, repetitions, and omissions-additions-words aided.)

4. There is no significant relation among the rankings of error-types or categories by the three oral tests at INSTRUCTIONAL level.

5. There is no significant relation among the rankings of error-types or categories by oral reading passages at the INSTRUCTIONAL, FRUSTRATION 1, and FRUSTRATION 2 levels of performance.

Definition of Terms

Average-or-above intelligence is defined in this study as a full-scale IQ of 90 or above on the Wechsler Intelligence Scale for Children.

Disabled reader is a reader whose reading level is significantly below his mental-grade level.

Levels of performance and/or levels of reading performance refer to the INSTRUCTIONAL, FRUSTRATION 1, and FRUSTRATION 2 levels defined below. The levels of performance are defined in terms of the word recognition standards of specific tests and have no other connotation of acceptable or unacceptable performance.

INSTRUCTIONAL level (I), when in caps, is used in this study to designate the paragraphs which are not error-free but in which the number of oral reading errors falls within the range of acceptability as defined by the scoring standards of a specific test. For the Durrell it includes the first paragraph on which two or more errors are made through the paragraph immediately preceding the one with seven or more errors; for the Gates, it means from the first paragraph with two or more errors through the paragraph preceding the one with eleven or more errors; for the Standard Reading Inventory (SRI), it includes the stories with error scores falling in the definite or questionable instructional columns on the SRI scoring sheet.

FRUSTRATION 1 level (F^1) is defined as the first paragraph on the Durrell in which seven or more errors are made; on the Gates, it is the first paragraph with eleven or more errors; and on the SRI, it is the first story in which the total or word recognition score falls in the frustration column on the SRI scoring sheet (an error ratio in excess of one error in eleven running words), providing that the next higher

paragraph or story does not meet the requirement for INSTRUCTIONAL level.

FRUSTRATION 2 level (F^2) is the next higher-numbered paragraph or story above the FRUSTRATION 1 level.

Error or miscue refers to any oral response which deviates from the written stimuli in oral reading. The terms are used interchangeably and imply no judgement of "wrongness" or "badness."

B-S-R Error Analysis is an error classification system utilizing twenty error-types and six major error categories; it is described fully in Chapter III.

Error-type means a specific kind of error (e.g., word omission) and is a subdivision of an error category. The twenty error-types used in this study are defined more fully in the description of the B-S-R Error Analysis in Chapter III.

Error category refers to a class or grouping of error-types. The six major categories on the B-S-R Error Analysis are: visual perception --word parts, directional confusion, visual-auditory, syllabic division, structure, and behavorial characteristics. The last category is divided into three sub-categories for the statistical analyses: corrections, repetitions, and omissions-additions-words aided. The categories are described in the B-S-R Error Analysis description in Chapter III.

Word recognition errors on the SRI include substitutions, mispronunciations, word omissions, word additions, and words pronounced by the examiner.

Total errors on the SRI include the word recognition errors listed above plus corrections and repetitions of a word, group of words, or

word part.

Informal Reading Inventory refers to an informal reading test consisting of graded oral and silent reading selections followed by comprehension questions at each grade or basal reader level. Three levels --independent, instructional, and frustration--are identified by comprehension and oral reading accuracy criteria.

Betts' criteria (1946) for the informal reading inventory are:

Independent level: 99 per cent oral reading accuracy and
90 per cent comprehension accuracy;

Instructional level: 95 per cent oral reading accuracy and
75 per cent comprehension accuracy;

Frustration level: 90 per cent or less oral reading accuracy;
50 per cent or less comprehension accuracy.

Intratest comparisons refer to comparisons between the levels of performance--INSTRUCTIONAL, FRUSTRATION 1, and FRUSTRATION 2 levels--of a single test.

Intertest comparisons refer to comparisons among the oral reading passages on the Durrell, Gates-McKillop, and SRI.

Delimitations

Scope of the Study

This study includes an analysis of the oral reading errors made by disabled fourth grade pupils at the INSTRUCTIONAL, FRUSTRATION 1, and FRUSTRATION 2 levels on each of three standardized oral reading tests. Comparisons of the resulting error patterns were made between tests and between levels of performance. Similar comparisons were made of oral rates of reading. Grade-equivalent scores were compared for five

measures of contextual reading and four measures of words in isolation.

The subjects in this study included all of the fourth-grade children in one county in northern Oklahoma who were average or above in intelligence and were reading below grade level. The final sample included 77 children from twenty-eight of the thirty-four public and parochial schools in the county.

Limitations of the Study

This study is limited to disabled fourth grade pupils from one county in northern Oklahoma.

The oral reading tests used in this study were only a sample of the measures which might have been used. Other tests might have yielded different results.

The reading selections at FRUSTRATION 1 and FRUSTRATION 2 levels were probably not of the same relative difficulty for all subjects in the sample.

Assumptions

It is assumed that the instruments used in this study actually measure the factors they are designed to measure and are pertinent to the study.

It is assumed that the use of oral reading errors to establish levels of reading performance is valid and that the number of errors made by a child is indicative of the relative difficulty of the material for him.

It is assumed that each word in a story will afford to a given child an opportunity to make any one of several types of errors and that the errors are a random sample of reading behavior for an individual reader.

It is assumed that the classification of reading errors is valid and that the particular analysis system used in this investigation is appropriate for this purpose.

It is assumed that the uncontrolled variables are randomly assigned.

Organization of the Study

Chapter I has presented a need for the study, a statement of the problem to be explored and the hypotheses to be tested, the definition of terms as used in the study, and the delimitations of the study.

Chapter II will review the literature which is related to the problem being studied.

Chapter III will describe the population studied, the instruments used for the collection of the data, the testing procedures, and the statistical techniques used to test the hypotheses.

Chapter IV will contain a statistical analysis of the data. It will contain the treatment of the data, the analysis of the results, and indications of the degree to which the hypotheses were found to be correct.

Chapter V will present a summary of the study and discussion of the results including conclusions and recommendations.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

The literature related to oral reading error patterns and to the comparison of instructional levels or grade-equivalent scores among various kinds of reading tests is voluminous indeed. For convenience in presentation, this review of the literature will be organized into three sections: (1) grade-equivalent score comparisons of oral reading tests (2) comparisons of oral reading rates; and (3) oral reading error patterns. Subdivisions under sections two and three will include the influence of the material or test used, the effect of the relative-difficulty level of the passages, and the influence of the reading proficiency of the sample. Literature related to the significance of various oral reading errors and error classification systems will also be examined.

Comparison of Grade-Equivalent Scores

Harris (1970) and Gates (1947) have suggested that differences in grade scores between oral reading tests and standardized silent reading measures may yield valuable diagnostic information concerning a child's reading strengths and weaknesses. A survey of the literature indicates, however, that grade levels identified by informal measures often differ from those obtained on silent reading tests, perhaps because of

differences in the types of reading behaviors sampled (Leibert, 1965). Likewise, grade scores on one silent reading test may be significantly higher or lower than the grade scores on another (Pflieger, 1949; Taylor and Crandall, 1962; Davis, 1968; Williams, 1963; Sipay, 1961; and Brown, 1963).

Standardized Silent Reading Tests--Informal Reading Inventories

Significant differences have been found between the grade scores on standardized silent reading tests and the instructional levels obtained from informal reading inventories. Sipay (1961) compared the grade equivalents secured by 202 fourth-grade pupils on three standardized measures with their instructional levels on an informal reading inventory scored at two word-accuracy levels, 96 per cent (Criteria 96) and 90 per cent (Criteria 90). Even when the less stringent of the two standards was used, the Metropolitan overestimated the instructional level by one or more grades about two times in ten and the Gates in slightly less than three times in ten. The California was one or more grade levels higher in nearly half the cases.

The standardized test scores reported by Williams (1963) averaged one to four years above the informal reading inventory instructional levels of her fourth, fifth, and sixth grade sample when the scores of individual pupils were compared. However, the Metropolitan Reading Tests and the Gates Reading Survey mean grade scores were similar to the mean instructional level on the informal reading inventory which was based on the reading series used in the local school system.

Although the mean grade scores on standardized silent reading tests tend to be higher than the instructional levels on informals, the scores for individual students are less predictable. Schiffman (1963) reported

higher standardized test results than informal reading inventory instructional levels for all of the elementary and three-fourths of the secondary retarded readers in his sample (N = 697).

Leibert (1965) found a low correlation between scores of seventy second grade pupils on the Gates Advanced Primary Reading Tests and reading levels on an informal reading inventory. In no instance was the informal reading level as high as the standardized test grade score. All of Robeck's (1963) disabled readers made "frustration-level error ratios" when placed in material comparable in difficult to their grade placement scores on Monroe's Reading Aptitude Test, Gates Advanced Reading Tests, and/or Durrell-Sullivan Reading Capacity and Achievement Tests.

Daniel (1962) found a difference of two years between grade scores obtained by thirty-five third grade pupils on the Gates Advanced Primary Reading Tests and instructional levels identified by a modified criterion scoring (90-97 per cent word recognition; 70-79 per cent comprehension) of an informal reading inventory. He concluded that instructional levels could be identified by subtracting the constant 2.0 from the Gates score.

A study reported by McCracken during the same year gave some support to Daniel's conclusions. McCracken (1962) administered the Iowa Tests of Basic Skills and an informal reading inventory to fifty-six sixth grade pupils. Like Daniel, McCracken concluded that standardized tests overrated immediate instructional levels on informal inventories by approximately two years. If the pupils in his study were instructed at levels two years below their standardized test scores, 21 per cent would still be in books uncomfortably difficult, 4 per cent of which were at

frustration level, and 7 per cent would be reading books which were too easy.

In a later study utilizing 971 pupils in grades two through six, McCracken (1964) reported that 20 per cent of the pupils were rated at the same level by the Stanford Reading Test and an informal reading inventory, and 8 per cent of the pupils were rated higher by the informal measure. Furthermore, McCracken concluded,

> ...an individual score on either test would be of little value in predicting an individual score on the other. For example, scores on the Stanford falling into the interval 4.0-4.9 were associated with informal instructional levels of primer through sixth reader inclusively, and informal instructional reading level ratings of fourth reader were associated with Stanford scores ranging from 3.0 to 8.1 (McCracken, 1964, p. 359).

Standardized Silent Tests--Oral Reading Tests

Similar findings were reported by Botel and Davis. Botel (1969) compared instructional levels on the Botel Reading Inventory with grade scores on the California Reading Test, the Iowa Tests of Basic Skills, or the STEP Reading Test. Although the standardized tests tended to overplace the pupils from one to three levels, Botel concluded that 20 per cent of the pupils at grade three, 30 per cent at grade four, 12 per cent at grade five, and 37 per cent at grade six were underplaced one or two reader levels by the standardized test score.

Davis (1964) secured grade-equivalents on the Gates Reading Survey, the Stanford Achievement Test, and four non-standardized measures (Temple Informal Reading Inventory, Botel Reading Placement Test, teachers' evaluations of pupil reading levels, and pupils' ratings of their levels) for fifty randomly-selected fourth graders. While the mean standardized test grade scores ranged from 3.2 grades below the levels of the non-standardized measures to 2.7 grades above, the

standardized-test performance of individual pupils ranged from 7.5 grades below the non-standardized rating to 5.8 grades above. Davis concluded that standardized tests were not accurate for defining any level of reading achievement.

Long (1959), on the other hand, concluded that the Reading Comprehension subtest of the Iowa Tests of Basic Skills was a good predictor of instructional reading levels, as measured by the Oral Reading subtest of the Durrell Analysis of Reading Difficulty, for retarded readers in grades two, four, and six. The mean grade-equivalent score on the Durrell Oral Reading Test was one-third year higher than the mean grade score on the Reading Comprehension section of the Iowa which had been administered two to four months earlier. Five second and five fourth grade pupils who had scores falling in the lowest 10 per cent on city-wide standardized tests were rated at grade level or above by the Durrell.

Attea's (1966) randomly-selected third grade pupils made significantly lower scores on the Durrell Oral Reading Test than on the Reading Comprehension section of the Iowa Tests of Basic Skills. The mean score on the oral subtest of the Diagnostic Reading Scales was significantly higher than the Iowa, while the mean grade score on the Gates-McKillop Oral Paragraphs did not differ significantly from the Iowa comprehension mean.

Oral Reading Tests

Patty (1965) compared mean grade scores on the Gilmore Oral Reading Test and the Gray Oral Reading Tests with the mean instructional levels on the oral portion of an informal reading inventory, the silent sections of an informal, and the total informal inventory (as scored by

Betts' criteria), to twenty-five boys and a girl who had been referred to a University Child Study Clinic. The two tests indicated the same oral instructional level (± one year) in 22 out of the 26 cases.

Sipay (1961) and Brown (1963) found close agreement between the mean instructional levels obtained on alternate forms of informal reading inventories constructed from the same basal reader series. Williams (1963), however, found significant differences in instructional levels among three informal reading inventories based on different basal-reader series. She concluded that instructional levels based on one series may differ from levels based on another.

Attea (1966), in a previously-cited study, compared grade levels on comparable subtests of the Durrell Analysis of Reading Difficulty, the Gates-McKillop Reading Diagnostic Tests, and the Spache Diagnostic Reading Scales. The oral reading grade scores and the word-analysis grade-equivalents differed significantly among the three tests. The Gates oral reading test was "consistently higher" than the Durrell. (The mean difference was .33 grades.) The Spache oral reading and word analysis scores were significantly higher than the corresponding Gates scores.

Attea also noted "operationally unequal readabilities" within sets of paragraphs that were supposedly of equal difficulty on the Diagnostic Reading Scales: the twenty-three subjects who attempted Paragraph 3A were successful, but only one of them was successful on Paragraph 3B, although both are purported to be of 3.3 grade level difficulty. All sixteen subjects attempting Paragraph 8A were successful; none succeeded on 8B although the Manual indicates that both have a grade-equivalency of 8.5.

McCracken and Mullen (1970) reported a correlation of .95 between midpoints of the instructional ranges obtained by fourth-grade pupils on the Botel Reading Inventory and the Standard Reading Inventory.

Botel, Bradley, and Kashuba's (1970) fourth-graders obtained very similar mean scores on the Botel and the Standard Reading Inventory although the correlation between the two tests was somewhat lower than in the McCracken and Mullen (1970) study. The mean grade level on the Spache Diagnostic Reading Scales was approximately one year higher; however, the Spache correlated more highly with the SRI than did the Botel (on which there was no difference in grade score means).

Spache (1950) concluded that the median reading level (2.8) of twenty-three remedial pupils on the Oral Reading Test of the original Durrell Analysis of Reading Difficulty was not changed by the number of paragraphs on which the results were based; moreover, it did not matter whether the paragraphs were at grade level and above, or included paragraphs below the reading level. The number and relative levels of paragraphs read did affect the mean, however. The use of three paragraphs, two of which were above the reading level, produced the highest mean (3.2); the use of three paragraphs, one below and one above the reading level, produced the next-highest mean (3.1); the use of two paragraphs, one at grade level and one below, produced the third highest mean; and a single paragraph at the reading level produced the lowest mean (2.7). The Median grade score obtained on Gray's Oral Check Tests by the same group was 2.7 and the mean score was 2.9.

Words in Isolation--Paragraph Reading Tests

Several research studies have compared the grade scores on words-in-isolation tests with grade levels on other reading measures. Their

findings are somewhat equivocal. Garlock, Dollarhide, and Hopkins (1965) found nonsignificant differences between the mean grade placement score on the reading section of the Wide Range Achievement Test and the Accuracy and Comprehension mean grade equivalents on the Gilmore Oral Reading Test for 180 students in grades one through twelve. However, their subjects were somewhat atypical; thirty-five of the group were in classes for the Educable Mentally Retarded.

Attea (1966) concluded from a pilot study that the Word Recognition Test of the Durrell Analysis of Reading Difficulty yielded "much higher" grade equivalent scores for a group of third-grade children than did the Oral Reading Test. In the major study, she found that the mean grade score on the Durrell Oral Reading Test was significantly below the mean grade level on the Reading Comprehension section of the Iowa Tests of Basic Skills, whereas the Word Recognition mean grade-equivalent was significantly above. It appears from adding together Attea's figures that the mean difference between the Durrell Oral Reading Test and the Word Recognition score was .81 grades and the difference between the Oral Reading Test and the Word Analysis score, .74 grades.

Herlin (1963) found a difference of approximately a year between mean scores on the Durrell Oral Reading Test (4.17 1.09) and scores on the Word Recognition (5.17 1.45) and Word Analysis (5.12 1.43) Tests. The Word Recognition Tests correlated highly with both the Durrell Oral Reading Test and the California Reading Test.

McCracken and Mullen (1970) examined scores made by 171 pupils in grades one through six on each of the nine subtests of the Standard Reading Inventory to determine if the performance of individual pupils was better at lower levels than their performance on the same subtest at

successively higher levels. The pronunciation-of-words-in-isolation test most consistently reflected increasing levels of difficulty. The investigators concluded that this means the words in isolation test was the single most sensitive test on the SRI for indicating instructional level. The results of the sign test demonstrated the sensitivity of the word test for indicating changes in levels of difficulty rather than desirable instructional levels, however.

Differences in Performance Standards

Undoubtedly, a part of the inconsistency in research findings pertaining to standardized test-informal-reading-inventory grade placement comparisons, which is apparent in the studies cited above and in Table I, may be attributed to differences in the samples and the standardized measures used. An equally important cause of variation, however, may be the differing standards used by investigators to identify instructional and/or other functional reading levels with informal instruments. Since there are no universally-accepted standards of performance which may be used to define the various levels on the informal reading inventory, the criteria used have varied from researcher to researcher (Beldin, 1970). See Table II. Differences include variations in the minimum level of word recognition accuracy and percentage of comprehension considered satisfactory and in the kinds of oral reading miscues to be counted as errors when computing word-accuracy percentages.

Effect of Differing Standards

Sipay (1961) computed instructional levels for 202 fourth grade pupils using two sets of criteria: a 90 per cent minimum word accuracy level (Criteria 90) and a 96 per cent minimum word accuracy level (Criteria 96). Fifty-two per cent of the pupils had the same instruc-

Table I

RESULTS OF STUDIES COMPARING ORAL AND SILENT READING TEST GRADE LEVELS

		N	Grade Level	Kind of Sample	Silent Test	Oral Test	Mean Grade Scores Difference	
Killgallon	(1942)	41	4th	Random	Gates Survey	IRI	1.44	
Schiffman	(1963)	697	Elem.	Retarded Rdrs.	Not Specified	IRI	1.52	
			Sec.	Retarded Rdrs.	Not Specified	IRI	1.26	
McCracken	(1964)	971	2-6	40 classrooms	Stanford Form KM	IRI	0.72	
McCracken	(1962)	56	6th	Total classrooms	Iowa Test of Basic Skills	IRI	1.3 2.3	Maximum Instruct. Minimum Instruct.
Daniel	(1962)	35	3rd	One classroom	Gates Advanced Primary	IRI	2.2 2.7	Modified Criteria (90% wd accuracy) Betts Criteria
Sipay	(1961)	202	4th	8 classrooms	Metropolitan	IRI	.79 No diff.	Criteria 96 Criteria 90
					Gates Survey	IRI	1.0 .29	Criteria 96 Criteria 90
					California	IRI	1.70 1+	Criteria 96 Criteria 90
Leibert	(1965)	70	2nd		Gates Advanced Primary	IRI	.4 2.6	to
Williams	(1963)	73	4-6	3 classrooms	Metrop. Rdg. Gates Survey California	IRI* IRI* IRI*	Averaged 1 to 4 yrs.** " "	

Table I (Continued)

Brown	(1963)	153	4th	One county	California Metropolitan Stanford Iowa Test of Basic Skills Gates Survey	IRI	Significant at .01 level
Long	(1959)	153	2,4, 6	Retarded Rdrs.	Iowa Test of Basic Skills	Durrell Oral Par	No Difference
Attea	(1966)	101	3rd	Random	Iowa Test of Basic Skills	Diag. Rdg. Scales Oral	DRG Sign. higher - .05 level
					Iowa Test of Basic Skills	Gates-McKillop Oral Rdg.	NS (G-M higher)
					Iowa Test of Basic Skills	Durrel Oral Par	Sign. Iowa .01 level higher
					Iowa Test of Basic Skills	DRG word analysis	DRG sign. higher - .01 level
					Iowa Test of Basic Skills	G-M word analysis	NS (G-M higher)
					Iowa Test of Basic Skills	Durrell Word analysis	Dur. sign. higher - .01 level

* Used Word Accuracy Level of 90%

** Individual Comparisons

Table II

INFORMAL READING INVENTORY CRITERIA

	Betts	Kill-gallon	Kress and John-son	Patty	Cooper	Sipay	Wil-liams	O'Brien	Daniel	McCracken	Powell	Dunkeld	Malm-quist
Words Pro-nounced by Examiner	X	X	X	X*	X	X	not specified	(5sec.) X	X	X	X	X	(15 sec.) X
Mispronunci-ations	X	X	X	X	X	X			X	X	X	X	X
Substitu-tions	X	X	X	X	X	X		X	X	X	X	X	X
Omissions		X	X	X	X	X		X	X	X	X	X	X
Additions or Insertions	X	X	X	X				X	X	X	X	X	X
Repetitions	X	X								X			X
Self-corrections										X			
SCORING CRITERIA													
Accuracy in word Recog-nition (low-er limits)	95%	95%**	95%	95%	Prim. 98% 94%+	Inter. 96% 90%+ 96% 90%	90%	95%	Modified Criteria 95% 90%	Question-able In-struc-tional 95% 91%	Gr. Gr. 1-2 3-5 85% 91-94%	Gr. 2 85.5% Gr. 3 91.5% Gr. 4-6 89%	

Table II (Continued)

	Betts	Kill-gallon	Kress and John-son	Patty	Cooper	Sipay	Wil-liams	O'Brien	Daniel	McCracken	Powell	Dunkell	Malm-quist
Comprehen-sion	75%	75%**	75%	75%	70% 60%	60% 60%	70%	---	70% 70%	70% 60%	70% 70%	70% 70% 70%	---

\+ Questionable Suitability

* "Hesitations"

** In original study, the word accuracy criteria was "less than one error in fourteen words or less than one error in fifteen words when accompanied by other symptoms." Comprehension, 50% or 75% when accompanied by other symptoms.

tional level using either Criteria 90 or Criteria 96; thirty per cent differed by one grade level; sixteen per cent by two grade levels; and two per cent by three grade levels. Daniel (1962) found five-tenths grade difference between the mean instructional level identified by a 95 per cent word accuracy criterion and that identified by a 90 per cent criterion. The correlation between the two criterion levels was .95.

One of the problems that researchers have faced in attempting to validate informal reading inventory criteria, or "instructional levels" on any instrument, has been the absence of generally-accepted measures of "second-grade reading level," "third grade reading level," and so forth, which has been evidenced by the studies reviewed in this section.

Oral Reading Rate

Several investigators have found surprisingly similar average reading rates. Cooper (1952) reported an average oral rate of 115 words per minute for twenty-two classrooms of second and third grade pupils "in materials properly adjusted to their abilities"; intermediate-grade students read 150 words per minute. Patty's (1965) primary pupils tended to read at the rate of 115-120 words per minute and the intermediate-grade pupils, at 140-150 words per minute. The median oral reading rate for the slightly-accelerated third-graders in Duffy and Durrell's (1935) study was 122 words per minute. Gilmore (1947) on the other hand, reported an average oral reading rate on the Gilmore Oral Reading Test of 88 words per minute for his third grade subjects and 96 words per minute for his fourth grade group.

Effect of Material on Reading Rate

The material read in the stories cited above ranged from basal reader paragraphs extracted from books in which the subjects were being instructed, as in the Cooper (1952) study, to standardized oral reading test passages (Duffy and Durrell, 1935; Gilmore, 1947). Rates in the Patty study were averages of the words read per minute on the Gilmore Oral Reading Test, the Gray Oral Reading Tests, and informal reading inventories (basal reader material).

The mean reading rates of primary pupils in Patty's (1965) study tended to be similar on the standardized and informal measures, although the Gilmore mean rate was significantly lower than the rates on informal measures for first grade pupils and were lower than rates on the informals and the Gray at all grade levels. Spache (1950) concluded that if lower level paragraphs were used, as he suggested, the oral passages from the Durrell Analysis of Reading Difficulty (original edition) and Gray's Oral Reading Checks could be used interchangeably in the first six grades for successive testing of rate although Durrell's norms ran somewhat lower than Gray's in the second, third, and sixth grades. It would appear that the type of material used in the various instruments was not sufficiently different to affect the oral reading rate in the studies reviewed.

Studies by Cooper (1952), Schummers (1956), McCracken (1961), and McCracken and Mullen (1970) indicate that rate of oral reading is affected by the relative difficulty of the material. In these studies, the average rate of reading decreased significantly as the level of the passage read increased. McCracken and Mullen (1970) found that Speed of Oral Reading was the second most sensitive test on the Standard Reading

Inventory for detecting changes in difficulty levels. Gilmore (1947) concluded that rate of oral reading was significantly related to both oral and silent reading comprehension.

Packman (1970) found that the oral reading rate on the Standard Reading Inventory of a stratified random sampling of fourth-grade pupils tended to decrease as the level of comprehension decreased. Significant differences were found between the mean reading rate at the 91-100 per cent comprehension level and rates of comprehension levels of 70 per cent and below. Mean reading rates found at the 71-90 per cent comprehension levels differed significantly from the rate at the 50 per cent and below comprehension level. The relationship between level of comprehension and rate of oral reading was more consistent for poor readers across the six levels of reading comprehension than it was for the average and good reader groups.

Kasdon (1970) found approximately a 30 per cent decrease in the reading speed of ninth grade students between the last passage passed on the Gray Oral Reading Tests and the first passage failed. There was not so great a change in reading rate between preceding passages.

Differences in Reading Rate Between Good and Poor Readers

In the Kasdon (1970) study cited above, two groups of subjects were randomly selected from two secondary schools in the ghetto area of New York City. One group of twenty-three subjects read passages from the Gray Oral Reading Tests orally at sight; the other group read the passages silently before reading them aloud. Rate of reading was calculated in words per minute from the passage preceding the two on which the pupil made seven or more errors. Unlike the achieving fourth-graders in an earlier study (Kasdon, 1967), the group who read the

passage silently before reading it aloud did not read at a significantly faster rate than the group reading orally at sight (silent-first rate: 111.8 ± 4.75 words per minute; orally-at-sight rate: 111.0 ± 5.31 words per minute.) Fourth-grade accelerated readers read orally at sight at the rate of 111 words per minute on selections from the Diagnostic Reading Scales; the mean rate on comparable passages read silently before being read aloud was 126 words per minute, a difference significant at the .01 level (Kasdon, 1967).

Packman (1970) found that the oral reading rates of the poor readers in her sample were significantly lower than the rates of the average and good readers even when the reading comprehension levels were similar. The shapes of the reading rate profiles across six performance levels of comprehension were not significantly different for the three reader-groups, however. Watkins (1953) found that intermediate-grade disabled readers read more slowly than younger children of similar reading ability and intelligence who were progressing normally in reading achievement.

Speed of oral reading seemed to discriminate among the good, average, and poor readers in several additional studies (McCracken, 1961; Schummers, 1956; Madden and Pratt, 1941). However, since both good and poor readers read the same selections in these investigations, the greater relative difficulty of the material for the poorer reader may have accounted in part for the difference in reading rate.

Oral Reading Error Patterns

Many investigations have been made of oral reading errors; as has been noted by previous reviewers (Courtney, 1964; Schale, 1964: Weber,

1968), however, it is extremely difficult to identify patterns of oral reading errors from study to study, level to level, or reader type to reader type. There has been little consistency in the method of reporting errors. An error may have been reported in terms of the percentage or ratio of its occurrence to the number of errors, by the percentage or number of subjects making the error, by the average number of errors of that type made per subject, or by its rank order (as determined by its frequency of occurrence or the number of subjects making the error).

The Error Classification Systems used have varied in the number and range of categories and the unit classifiable within the category (word, letter, both word and letter) (Stuever, 1969; Weber, 1968).

The same reading behavior may be classified by a number of different categories "and the items that fall into identically labeled categories may vary" (Weber, 1968, p. 107). A term may be inadequately defined so that it is unclear to which of its possible meanings an investigator has applied the label.

> ...Substitution may refer to any erroneous response to a given written word in one system while referring to a response with absolutely no sound-letter correspondences in another (Weber, 1968, p. 101).

In other instances, an erroneous response which makes a "real word" is intended.

Mispronunciation may refer to any erroneous response to a given written word or it may designate an unsuccessful pronunciation-attempt resulting in a pseudo-word.

Reversal may refer only to a "full reversal" or the use of the final letter in the initial position (i.e., was-saw) as in Christenson's (1966) study; or may, in addition, include one or more of the following: partial reversal (i.e., are-ear), letter rotation (i.e., p-d, b-d,

and/or word-sequence transposition.

Hesitation may refer to a pause of a specified duration before pronouncing a word (e.g., 2 seconds, Gilmore, 1950), or may include (or be limited to) "a word to which the child did not respond in five or more seconds" as in Schummers' (1956) classification; thus, what is classified as a hesitation in one study may be categorized in another as aid or word aided, word pronounced, word refused, refusal, or omission.

Omission may refer only to omissions of whole words or may also include the omission of one or more of the following: letter, syllable, prefix, suffix, inflectional ending, and/or punctuation mark. Gates and McKillop (1962, p. 62) place in this category "failures to respond in five seconds and also words skipped over or 'refused' by the child."

Addition or Insertion may refer only to insertions of a whole word or group of words, or may include the addition of any or all of the following: letter, syllable, prefix, suffix, inflectional ending, and/or punctuation mark.

Repetition or Regression may mean any one of the following to a particular investigator: a word or group of words repeated; a word, part of a word, or group of words repeated; or, one or more words repeated except when the regression was made to correct an error, in which case it is not classified as an error. Regressions are considered by some investigators to be a part of the self-correction or verification process, or a form of hesitation--a filled pause--and are not considered errors (Goodman, 1965; Y. Goodman, 1967; Burke, 1969; Weber, 1970; Allen, 1969; Nurss, 1970).

Error classifications may be overlapping or not mutually exclusive. Bennett (1942) for example, had a separate category for Medial Vowel

Errors although the errors classified in this category logically belonged to one of her Beginning and Ending Correct groups. The results of analyses based on such overlapping categories are difficult to interpret (Hill, 1936; Weber, 1968). A single-word error may be classified under all possible categories by one investigator (Monroe, 1928; Daniels and Diack, 1956) and placed in a single category by others, making comparisons between such studies difficult to interpret.

Some investigators (e.g., Bennett, Gates, Leibert, and Weber) have assumed that all word substitutions and mispronunciations are caused by inaccurate perception of the printed word and, therefore, have classified words according to the position(s) of the error(s) in the word, ignoring the possibility of visual-auditory errors. Other investigators (e.g., Monroe, Killgallon, Schale, and Schummers) have used only phonetic or sound-symbol categories, classifying such errors as walked for walking as a "sound" error (Stuever, 1969).

Effect of Differing Definitions

Swanson's (1937) data describing the oral reading performance of university freshmen suggest that the particular definition of omission and addition errors which the investigator chooses to use may have considerable effect upon the relative sizes of the Omissions, Additions, Substitutions and/or Mispronunciations categories. Fifty-three per cent of the total number of omission error made by the ten good silent readers and 43 per cent of those made by the seventy poor readers were letter omissions. Another 13 per cent (of the poor reader group's omissions) were omissions of syllables. Skipping, which Swanson defined as the omitting of words or parts of words followed by a correction of the omission, accounted for 20 per cent of the Omissions category for good

readers and 10 per cent for poor readers. Fewer than one-third of the omission errors, then, were omissions of whole words. In the same study, slightly over half of the insertion errors were word inserts, one-fifth were additions of syllables, and approximately one-fourth were insertions of letters.

Swanson classified omitted and inserted letters and syllables as well as words in the Omissions and Insertions categories. If he had included omitted letters and word parts in the Substitutions category, the apparent size of this error-type would have substantially increased and the percentage of Omissions and Insertions proportionately decreased as Table III indicates. Table III also summarizes the effect of differing term-definitions upon the apparent frequency of specific kinds of oral reading errors made by children of third and fourth grade reading levels. (The figures in Table III were obtained by reclassifying Monroe's 1928 and Herlin's 1963 error categories according to the definitions of omission and insertion errors discussed above.)

Some investigators and oral-reading-test authors consider only the repeating of two or more consecutive words to be a repetition, whereas others count the repeating of word-parts as repetition errors. Goodman (1965) found that repetitions made by primary-grade pupils were almost evenly divided between single-word and two-or-more word regressions.

Both good and poor readers among Swanson's (1937) university freshmen made slightly more single-word repetitions than repetitions of groups of words (47 per cent as compared with 35 per cent; for good readers; 32 and 27 per cents respectively for poor readers). Forty-one per cent of the repetitions made by poor readers and 18 per cent of those made by good readers, however, were repetitions of word-parts.

TABLE III

EFFECT OF DEFINITION OF OMISSIONS

AND INSERTIONS

	Grade or Reading Level	Kind of Reader	Substitutions and/or Mispronunciations	Omissions (words only)	Insertions (words only)	Substitutions and/or Mispronunciations	Omissions (letters, syllables or words)	Insertions
Swanson (1937)	University Freshmen	Good	48%	5%	13%	24%	18%	24%
	University Freshmen	Poor	57%	6%	7%	40%	17%	13%
Monroe (1928)	3	Normal	74%	4%	4%	49%	19%	15%
	3	Retarded	71%	2%	3%	45%	18%	13%
	4	Normal	71%	5%	8%	45%	22%	17%
	4	Retarded	74%	4%	4%	43%	24%	15%
Herlin (1963)	3	Normal	56%	5%	3%	32%	15%	17%
	4	Normal	62%	4%	2%	37%	19%	13%
	4	Normal	64%	2%	2%	37%	20%	12%

Thus, the frequency of repetition errors would seem to be very dependent upon the definition of repetition that is used by the investigator.

Most Prevalent Error-types.

Despite the difficulties in comparing cross-study results which were noted above, certain conclusions regarding error patterns appear to be warranted. Substitutions and/or Mispronunciations tended to be the most prevalent error-type at all levels and with all kinds of readers. As may be seen from Table IV, half or more of the errors reported by any one investigator tended to fall in this category.

In what is perhaps the major study of developmental changes in oral reading errors among randomly-selected readers, Schale (1964) compared types of errors made on grade-level paragraphs from the Gray Oral Reading Tests by fifteen boys and fifteen girls in even-numbered grades, two through twelve inclusive. Each subject also read two passages below grade placement and two passages above grade level. Schale found that substitutions were one of the two most frequent kinds of errors at each grade level studied. Furthermore, the number of substitutions was not affected by the relative level of difficulty (insofar as this was controlled in Schale's study) nor the grade level of the reader. Partial mispronunciations, on the other hand, were important only at the sixth-grade level and above, with wrong sound and accent becoming more frequent as the material became more difficult.

In an earlier study of error-type importance at different grade levels, Gilmore (1950) administered Form A of the Gilmore Oral Reading Test to 446 pupils in grades one through eight. He, too, concluded that substitutions were the most prevalent kind of error at each grade level.

Among Weber's (1970) first grade pupils, substitutions were eight

TABLE IV

SUMMARY OF ERROR-TYPE FREQUENCIES IN STUDIES REPORTING PERCENTAGES OF OCCURRENCES

	Grade Level	Kind of Reader	N	Test Material	Mispron. and Subst.	OMISSIONS			INSERTIONS			Words Aided or Refused	REPETITIONS		REVERSALS	
						Word	Word, Affix, Inflect. End.	Letter, Syl. or Word(s)	Word	Word, Affix, Inflect. End.	Letter, Syl. or Word(s)		Word or Word Part	Two Words	Full	Partial And Full / Letter Rot. Full Partial
Malmquist	1st	Poor	31	Test (Malm.)	43.9%		8.2%		0.7%			43.1%[a]	2.4%			1.7%
	1st	Aver.	172	Test (Malm.)	53.7%		14.3%		1.1%			15.0%[a]	12.1%			3.9%
	1st	Good	40	Test (Malm.)	51.6%		22.7%		4.0%			2.2%[a]	17.2%			2.2%
Weber	1st	Poor (Low)	9	Basal Mat.	81.6%			7.2%			8.8%	--	--			2.5%
	1st	Good	12	Basal Mat.	78.7%			9.4%			9.5%	--	--			2.4%
McCracken	2nd	Poor	9	Sheldon	20 %		5 %			11 %		38 %	24 %[b]			
	2nd	Aver.	17	Sheldon	25 %		8 %			11 %		28 %	26 %[b]			
	2nd	Good	9	Sheldon	25 %		8 %			19 %		10 %	38 %[b]			
Schale	2nd	Random S.	15	Gray Oral	17.3%	4.0%			6.6%			1.3%	70.7%[b]		0	
	4th	Random S.	15	Gray Oral	33.9%	5.7%			4.7%			6.6%	49 %[b]		0	
Schummers	3rd	Random S.		Lyons IRI Carnahan	43 %[d]	3.4%			2.0%			42 %	6.7%[c]		3.2%	
Murss	3rd	Unsuccessful		Lippincott Basal	55 %	29 %			3 %				(12% Self-Corrections)			
	3rd	Successful		Scott, Foresman Basal	53 %	12 %			10 %				(15% Self-Corrections)			
Monroe	2nd Level Rdg.	Retarded		Gray Oral Iota Wd List Word Discrim.	69.0%[d]	1.5%			1.5%			4.5%	11.8%			11.6%
	2nd	Normal	1	Gray Oral Iota Wd List Word Discrim.	73.2%[d]	3.4%			2.0%			4.4%	8.0%			8.9%
	3rd Level Rdg.	Retarded		Gray Oral Iota Wd List Word Discrim.	71.0%[d](48.5%)	2.4%		(18.0%)	2.9%		(13.1%)	0.9%	13.6%			9.1%
	3rd	Normal		Gray Oral Iota Wd List Word Discrim.	74.1%[d](45.2%)	3.7%		(18.7%)	4.4%		(14.9%)	4.4%	6.5%			7.0%
	4th Level Rdg.	Retarded			73.9%[d](44.8%)	3.7%		(24.2%)	4.3%		(14.8%)	1.0%	8.9%			8.2%
	4th	Normal			70.8%[d](43.0%)	5.4%		(22.1%)	7.8%		(17.1%)	4.9%	5.2%			5.8%
Harlin	3rd	Unselected		Monroe Battery	56.2%[d](32.3%)	5.2%		(15.3%)	3.4%		(17.2%)	13.2%	10.2%			11.8%
	3rd	Unselected		Durrell												
	4th	Unselected		Monroe Battery	62.1%[d](36.6%)	4.0%		(18.0%)	2.4%		(13.4%)	6.0%	14.0%			11.5%
	4th	Unselected		Durrell	64.2%[d](36.7%)	2.4%		(19.6%)	1.8%		(12.1%)	15.2%	9.5%			7.0%
Gates Norms	3.5	Unspecified		Gates Oral Par. 1-4	78 %	10 % (and refused).			2 %			Included in "Omissions"		10 %	0	
Madden and Pratt	3rd		136	Soc. Stu.	59 %	5 %			5 %			21 %	9 %		0.6%	
	3rd		136	Science	56 %	8.4%			6 %			17.1%	11 %		1 %	
	4th		145	Soc. Stu.	53 %	17 %			9 %			6 %	14 %		1.0%	
	4th		145	Science	51 %	15 %			12 %			9 %	13 %		0.4%	
Christenson (1966)	4th	One Classroom Mostly Above Average	22	Instruct. Level IRI	39.3%		18.6%[e]			13.6%[e]		0.8%		27.5%	0	
	4th	One Classroom Mostly Above Average	22	Frust. Level IRI	42.9%		17.2%[e]			16.4%[e]		2.8%		20.7%	0	
	4 - 6	Three Classrooms Mostly Above Aver.	68	Instruct. Level IRI	39.7%		20.9%[e]			13.2%[e]		0.3%		25.9%	0	
	4 - 6	Three Classrooms Mostly Above Aver.	68	Frust. Level IRI	42.9%		16.9%[e]			15.3%[e]		1.9%		23.0%	0	
Goodman-Burke (1968)	4, 5	Proficient	12	6th Grade Story	56 %			29 %[f]			12 %				3 %	
Hardin and Ames	CAs 7-8 - 15-10	Remedial	26	Gray Oral	51.6%	7.9%			6.0%			10.9%	23.5%[b]			
	CAs 7-8 - 15-10	Remedial	26	IRI-Sheldon Series Manual	44.4%	10.4%			11.1%			5.3%	28.8%[b]			
Robeck	2nd - 7th	Rem.-Lacking Word Attack	20		28 %	Few or None			Few or None			18 %	16-20%		Few or None	
Swanson	Univ. Fresh.	Poor	70	IRI	40 %(57.4%)	(5.6%)		17 %	(7 %)		13 %		22 %			
	Univ. Fresh.	Good	10	IRI	24 %(47.8%)	(5.0%)		18 %	(13.2%)							

[a] After 15 seconds
[b] One or more words
[c] One or more words except when word is repeated for adequate correction
[d] Combination of Monroe's Faulty Vowels, Faulty Consonants, Added Sounds, Omitted Sounds and Substitutions
[e] Included punctuation marks
[f] Inflectional ending not included

times as frequent as the next largest error type. (Word refusals and regressions were not tabulated.)

In general, the percentage of substitutions made by poorer readers tended to be similar to that made by the more successful readers in the same study. Only Malmquist's (1958) poor first grade readers, McCracken's (1961) poor and average readers at the second grade level, and Schummers' (1956) third grade boys had as many words refused as words substituted or mispronounced. Schummers' findings are somewhat difficult to interpret since more than half of the words-refused or hesitation errors (4,901 out of 8,831) were made by only 22 of the 237 third-grade children in the sample.

Table V summarizes the relevant findings of investigators who used check lists and reported their results in terms of the number or percentage of subjects exhibiting a behavior. Again, substitutions and/or mispronunciations appeared to be the most prevalent error-type, reported under such terms as _low sight vocabulary_, _inadequate word mastery skills_, and _inaccurate guessing_ or _guesses from context_.

Investigators using a remedial or disabled population consistently reported errors on easy words (Swanson, 1937; Davis, 1931; Mulroy, 1932; Wells, 1935; Barbe, Williams, and Ganaway, 1958; Long, 1959; Packman, 1970). Baker (1945, p. 575) noted that the reading difficulty most frequently encountered among her remedial pupils in grades four through eight was confusion of word forms and the "confusions are most frequent in words which are regarded as basic or common to all reading material." Wells (1935) concluded that if some of his subjects had only misread the words they could not read they would hardly have been considered reading problems.

Table V

RESULTS OF INVESTIGATORS USING CHECKLISTS

	Killgallon 4th	Duffy and Killgallon 3rd	Daw 4th	Daw 5th	Long 2nd N=51	Long 4th	Long 6th	P	Barbe I	Barbe JH	Davis Low 4th	Davis High 4th
Low Sight Vocabulary	50%*	38%	26%	33%	82%	56%	41%	36.4%	48.7%	58.3%	61%	64%
Poor Word Mastery Skills	40%	32%	41%	38%				22.7%	41.0%	41.7%	49%	38%
(Cons.)					50-61%	33%	27%	(9.1%	48.7%	41.7%)		
(V)					62%	68%	47%	(Ignores Wd Endings)				
(Com. Syl.)					100%	54%	20%					
Guesses from Context (Inaccurate Guessing)	45%	41%	7%	6%	74%	43%	25%	40.9%	20.5%	41.7%		
Additions and Omissions	47.5%**	45%	14%	5%					7.7%	33.3%	5%	2%
Repeats Frequently	80%	25%	20%	12%					2.6%	25.0%	7%	6%
Refusals	22.5%											

* Substitution

** Insertions 32.5%
Omissions 15.0%

Inaccurate recognition of common words and the confusion of words which were somewhat similar in appearance were also noted by Mulroy (1932). Although more than half of the errors on Gray's Oral Reading Check Tests made by Mulroy's older students were mispronunciations and failure to recognize words, substitutions were the most common errors among the fifth and low-sixth grade groups. A tendency to make errors on easy words was also noted by investigators using unselected samples (Sheldon and Hatch, 1950; Duffy and Durrell, 1935; Daw, 1938).

The inability to use word attack skills appeared to be characteristic of many pupils in the studies cited. It is not clear whether the inadequate word attack skills resulted in mispronunciations, refusals, or both.

Addition and omission of whole words tended to be relatively infrequent as compared with word substitutions and mispronunciations. The importance of repetition errors, from the standpoint of frequency of occurrence, varied widely from study to study. Schale (1964) found repetitions to be one of the two most prevalent "errors" at every grade level. In Killgallon's (1942) study, they were the most frequent kind of "error" at the instructional level. Cooper (1952), on the other hand, concluded that repetitions failed to discriminate between levels of difficulty because of their low frequency of occurrence. Table VI summarizes the findings of various investigators regarding the frequency of repetition "errors." The percentages refer to the percentage of subjects making repetitions or the ratio of the repetition "errors" to the total errors made.

Influences on Error Patterns

Not all of the variations in error-frequencies which have been

Table VI

FREQUENCY OF REPETITIONS

25 Per Cent or More	10 - 20 Per Cent	6 or 7 Per Cent	Not Counted
McCracken (1961)	Monroe (Retarded Readers 2nd and 3rd Gr., 1928)	Schummers (1956)	Goodman (1965)
Schale (1964)	Madden and Pratt (3rd Gr. Science and 4th Gr. (1941)	Monroe (Normal Rdrs)(1928)	Goodman and Burke (1968)
Hardin and Ames (1969)	Daw (1938)	Davis (1931)	Y. Goodman (1967)
Swanson (1937)	Malmquist (aver. and good rdrs.) (1958)	Cooper (1952)	Burke (1969)
Christenson (1966)	Herlin (Monroe battery, (1963)	Mulroy (1932)	Allen (1969)
Duffy and Durrell (1935)	Robeck (1963)	Malmquist (Poor rdrs., (1958)	Weber (1970)
Barbe, Williams, and Ganaway (1958)		Herlin (Durrell battery, 1963)	Nurss (1970)
Killgallon (1942)		Madden and Pratt (3rd Gr. Soc. Stu., 1941)	Clay (1968)
Gilmore (1947)			

noted can be attributed to differences in error definitions and classification schemes. It is probable that the kinds of error patterns identified by researchers are dependent in part upon the grade level or levels investigated, the kinds of readers studied, the test(s) or materials used, and the relative difficulty of the test material for the individual subject.

Effect of grade level. Powell (1971) has proposed that the errors which significantly affect reading success at the second grade level may be different from those affecting the success of sixth grade readers. Ilg and Ames (1950, p. 252) believed that certain kinds of errors may be indicative of certain levels of skills development and "might well be relatively benign and characteristic responses of certain age levels."

The studies of such investigators as Schale (1964), Madden and Pratt (1941), and Gilmore (1950) provide evidence that some types of errors are more prevalent at certain grade levels than at others. Schale (1964) concluded from her intensive study of the oral reading error patterns of students in even-numbered grades two through twelve that no response and repetition errors decreased as grade levels increased; partial and gross mispronunciations increased as grade levels increased; and the relative frequency of substitutions, word omissions, and word insertions did not change from grade to grade. Reversals occurred too infrequently to be evaluated.

Results of other investigators agree with Schale's findings that refusal or no response error occurrence is inversely related to grade level, although the level at which this error-type ceases to be of importance varies from study to study. Gilmore (1950) and Goodman (1965) found a sharp decline in no response errors after first grade and

a steady decrease at each successively higher grade level. Madden and Pratt (1941) reported a sharp decrease in frequency after grade three. For Schale's (1964) group, the decline came after grade four.

Watkins' (1953) and Monroe's (1928) results suggest that age rather than reading achievement level may be the important variable. Watkins compared the reading performance of third grade pupils who had made normal progress in reading with the performance of "retarded readers" in grades four, five, and six of comparable reading ability, IQ, and sex. The normal-progress readers made more word refusals on the Iota word list from the Monroe Diagnostic Reading Examination than did the older "retarded" group with third-grade reading ability. Word refusal errors decreased after second grade among Monroe's (1928) "retarded" reader groups which, like Watkins' "retarded" sample, were made up of older children; no such decline occurred among the "normal" readers in second, third, and fourth grades.

There is less support from other studies for Schale's conclusion that repetitions decrease as grade levels increase. Madden and Pratt (1941) found an increase in repetitions from grade three to grade nine (although there was little difference in the percentage of repetitions made in grades four through eight). Barbe, Williams, and Ganaway (1958) found an increase from primary to intermediate levels, but reported no repetition incidences among the junior high group. Gilmore (1950) found repetitions to be one of the most common errors at seventh and eighth grades but too infrequent to be studied below these levels.

Gilmore (1950), like Schale, found a tendency for mispronunciations to increase as the grade level increased, starting with grade five (grade six in the Schale study). Stafford's (1967) high-achieving

intermediate-grade group made significantly more gross mispronunciations and mispronunciations of syllables or accent than did her poor-reader group. The greater proportion of unknown multisyllabic words in more-advanced materials, as well as a greater tendency on the part of the reader to attempt a pronunciation, probably accounts for the increase in mispronunciations found at higher grade levels.

The relative frequency of substitution errors does not appear to change with advancing maturity (Schale, 1964; Gilmore, 1950; Madden and Pratt, 1941). Schale (1964) found a slight increase in word omissions and a slight decrease in word additions from grade two to grade four, although she included both in her "no change" category. Monroe (1928) and Madden and Pratt (1941) found an increase in both types of errors from grade three to grade four; and Ilg and Ames (1950), an increase from grade two to grade four. Insertions were so infrequent at all levels in Gilmore's (1950) study that he considered them of little importance; omissions were not studied.

Effect of differing pupil-samples. In general, neither the sex of the pupil nor the level of intelligence appears to have a significant effect on the kinds of oral errors made. Schale (1964) found sex differences to be negligible in regard to frequency of oral reading error-types throughout the elementary and high school years.

Christenson (1966) found no significant differences in the frequencies of oral reading error-types made by boys in grade four, five, and six, and those made by girls at either the independent or the instructional levels. At the frustration level, boys made significantly more substitutions and girls made significantly more refusals.

Although the boys in Schummers' (1956) study made a significantly

greater proportion of secondary errors than did the girls, within the primary or the secondary error category there was little difference in the distribution of error types for boys and girls. (Primary errors were errors which changed the form of the word, whereas secondary errors did not change the word form.) The larger number of hesitation (refusal) errors made by boys (48 per cent of the total errors as compared with 30 per cent of the total errors for the girls) accounted for most of the sex difference. Since 55 per cent of the hesitation errors were made by a relatively small number of children who could not read some of the stories, the observed difference between girls' and boys' error-distributions is difficult to interpret.

In general, no significant sex differences were found in the ability of second and fourth grade pupils to comprehend materials written with frequent and infrequent oral language patterns (Tatham, 1970).

Schummers (1956) found that the percentage distribution of errors did not differ significantly among three intelligence groupings of third grade children: Low (IQs of 71-100), Medium (IQs of 101-118), and High (IQs of 119-152).

The instructional background of the pupils may influence the kinds of errors that are made. Daniels and Diack (1956, p. 39) have criticized studies of children's reading errors for failing to give "necessary information" about what particular training or experiences the children had before the investigation was carried out, contending that "few valid statements can be made about children's errors unless the question of teaching method is considered."

Herlin (1963) did not find significant differences in the error

patterns of two groups of third and fourth grade pupils who had been instructed in different basal reader series and in different school settings.

Daniels and Diack (1956) found differences in both numbers and kinds of errors made by two groups of first year junior children taught by differing methods of reading instruction. One group was instructed by a phonic word-method (the Royal Road Readers, written by the investigators) and the other group by a "mixed methods" approach. After a year of instruction, the phonic word-method group made fewer errors on all measures, including tests of "irregular" words. Significantly more no response errors were made by the mixed-methods group. The phonic word-method group made a larger percentage of errors categorized as Phonic Rendering of Irregular Words (e.g., once read as onk). The investigators concluded that a very high percentage of the errors made by the children in both groups, however, was due to their ignoring some of the letters in the words presented to them.

Some differences were observed in the types of errors made by Scottish children who had earlier and more intensive training in word analysis and American children of the same age or a year older. Scottish children tended to make fewer errors than American children of the same age but made a proportionately greater number of miscues which significantly changed the meaning (Elder, 1966).

Whether pupils are good or poor readers appears to have more effect on their error patterns than does the instructional background of the subjects. Although there is no kind of reading error that is associated solely with poor readers, the relative frequency of certain error-types may discriminate between good and poor readers (Stafford, 1967; Monroe,

1928; Malmquist, 1958; Sheldon and Hatch, 1950, 1951; Hatch and Sheldon, 1950).

Monroe (1932) analyzed the errors made by 415 reading disability cases and 101 control subjects on Gray's Oral Reading Paragraphs and two tests of words in isolation. Error norms were developed from the control group's performance. Monroe found that the "reading-defect" cases greatly exceeded the controls in the number of total errors, vowel errors, consonant errors, reversals, omission of sounds, repetitions, and addition of sounds.

There is considerable evidence that poor readers in the primary grades tend to make more word refusal errors than good readers of the same age (Malmquist, 1958; McCracken, 1961; Schummers, 1956; Stafford, 1967). Poor readers in Malmquist's first-grade sample made nearly eight times as many errors of this type as the average readers; in McCracken's second-grade group, poor readers made four times as many refusal errors as the good readers; and in Schummers' third-grade sample, the poor group had seven times as many refusal errors.

Bennett's (1942) and Monroe's (1928) findings suggest that older disabled readers make more reversal errors than do good readers of the same age.

Poor readers tend to read more slowly and to make more word recognition errors (Schummers, 1956; McCracken, 1961; Monroe, 1932; Packman, 1970). Packman found that poor fourth-grade readers obtained significantly lower word recognition scores and oral and silent reading rates on the Standard Reading Inventory than did better readers even when the reading comprehension levels were similar. She suggested that if similar levels of reading comprehension are used to define placement levels

for good, average, and poor readers, then the criteria for word recognition accuracy and rate of oral and silent reading will have to be different for poor readers than for better readers.

Bennett (1942, p. 36) concluded from an analysis of 34,274 errors made by 595 retarded readers that

> ...a pronounced characteristic of pupils retarded in word recognition seems to be the tendency not to inhibit associated responses until a word is clearly seen in all its parts--beginning, middle, and ending.

A similar conclusion was reached by Daniels and Diack (1956, p. 43) who expressed it as the "part seeing of one word and whole saying of another." Kagan (1965) found a relationship between impulsivity on design matching tests in grade one and errors in word recognition a year later. He surmised that reading errors may arise from an impulsive disposition.

Good readers tend to make more meaningful substitutions than do poor readers (Swanson, 1937; Fairbanks, 1937; Smith, 1954). Substitutions made by the good readers among Swanson's and Fairbank's university freshmen usually did not change the meaning of the passage; changes made by the poor readers tended to change the meaning significantly. The poor readers substituted for easy words as frequently as they did for more difficult ones.

Miscues made by good readers at the beginning levels of reading more closely approximate the stimulus words in terms of letters than do substitutions made by poorer readers (Bennett, 1942; Weber, 1970). At the intermediate-grade levels, however, disabled readers made a greater number of substitutions of words with similar form than good readers of the same age level (Stafford, 1967; Swanson, 1937).

Good readers tend to make proportionately greater numbers of word

omission and addition errors although the mean number of omissions and additions per subject may be higher for poor readers (Monroe, 1932; Smith, 1954; Schummers, 1954; Malmquist, 1958; McCracken, 1961). Nurss' good third-grade readers made a greater percentage of addition errors than did the nineteen poor readers. The poor readers had more omission errors, however, because they tended to skip the words they did not know.

Use of the norms devised by Monroe (1932), Bond, Balow, Hoyt (1970), and Gates (1962) requires the assumption

> ....that a retarded sixth grader, for example, reading at the third grade level should be compared for deviations from others reading at the third grade level, not for deviations from others placed in the sixth grade. (Herlin, 1963, p. 5).

In Bryant's (1968, p. 346) opinion, however,

> ....an older child who is a poor reader may perform differently on certain subtests of the Gates-McKillop than would be expected of a younger child reading at the same reading level.

Watkins (1953) and Monroe (1928) observed some differences between older disabled readers and younger pupils of the same reading levels. The disabled readers made significantly fewer word refusal errors than the normal-progress readers. The disabled readers read more slowly (Watkins, 1953). The disabled readers showed more variability in their reading errors (Monroe, 1928). The scores of the normal-progress readers tended to cluster around the mean, whereas the retarded readers' scores showed high and low deviations from the mean (Watkins, 1953).

Influence of materials. The particular words used in a selection may tend to restrict or to weight the error-types made on it. Bennett's (1942) test sentences contained only 594 of the 600 most frequently used

words from the Gates Primary Word List; of these, 74 per cent were monosyllables and 24 per cent were words of two syllables. She suggested the reason no nonsense words were recorded among the 34,274 erroneous responses may have been the "absence of longer, more complex words."

In discussing the interpretation of errors made on the Gates Diagnostic Tests, Gates (1947, pp. 316-7) stated:

> In the first four paragraphs in the Gates Oral Reading Test are included a relatively large number of words which when observed in reverse order or in partially reverse order give a series of letters which make up a real and usually a familiar word....The frequency of such errors would be less in typical reading matter.

Monroe, likewise, included an unusually large proportion of easily reversed words and letters such as p, q, d, b, g, n, and u which are frequently rotated (or confused) in the Iota Word Test which made up a part of the Monroe Diagnostic Battery. Although Gates does not consider the substitution of b for d to be a reversal error, Monroe included such letter rotations as a sub-division of her reversals category. Both Monroe (1928, 1932) and Herlin (1963), who used the Monroe battery and classification system, reported a high incidence of reversal errors.

Payne (1930, p. 145), on the other hand, found that reversals represented only 0.1 per cent of the total errors above the second grade level. She postulated that

> the reason for the relatively low percentages of reversals may be that, with two exceptions, the words presented were not of a reversible nature; that is, they were not wholly or in part the reverse of another word familiar to the children.

Studies of Goodman (1965), Goodman and Burke (1968), Burke (1969), Weber (1970), and others provide evidence that the structure of the written material influences the kind and percentage of reading miscues

occurring. Weber (1970), and others provide evidence that the structure of the written material influences the kind and percentage of reading miscues occurring. Weber (1970) found that 91 per cent of the errors made by twenty-one first-grade readers were grammatically appropriate to the preceding context. She concluded that both strong and weak readers used the

> ...constraints of preceding grammatical context to reduce the range of responses....When he has not yet learned to exploit all the information provided in the graphic display of a word, the preceding structure of a sentence may be his principal source of information for identifying a word. (p. 443)

Other investigators reached similar conclusions with the subjects ranging from slow first graders to highly-proficient readers in grade six. Two-thirds or more of the miscues made by both good and poor readers had complete syntactic acceptability (Burke, 1969; Allen, 1969; Weber, 1970; Clay, 1968; Goodman and Burke, 1968). Most of the miscues were semantically acceptable with the preceding text, and half or more had complete semantic acceptability (Burke, 1969; Allen, 1969; Weber, 1970; Goodman and Burke, 1968).

Bennett (1942, p. 36) observed, nearly thirty years ago, that

> the structure of the context in which the word is incorporated seems...to operate in some intangible way to govern the response, so that a verb stimulus, for example, calls forth in most cases a verb in response, even though it is erroneous.

Recent studies agree that the grammatical function of the miscue response tends to be the same as the grammatical function of the stimulus word (Burke, 1969; Clay, 1968; Goodman and Burke, 1968; Weber, 1970; Y. Goodman, 1967; Allen, 1969; Christenson, 1966).

Nurss (1969) investigated the relationship between sentence complexity and oral reading errors. Two second-grade groups read aloud six

one-sentence stories representing three levels of structural complexity. For the group of 108 second-graders who had been given a basic vocabulary screening test, significant effects due to levels of syntactic complexity were found for hesitations, self-corrections, and total errors (self-corrections, substitutions, additions, omissions, repetitions). For the 36 second-graders who had not been pre-tested, significant effects were found for total oral errors, other errors (additions, omissions, substitutions, and repetitions combined), and vocabulary in context (omissions, substitutions, and words supplied combined). The errors more frequently made sense semantically and grammatically in the sentences of less complex structure than in those of high structural depth for both groups. The vocabulary-screened group made fewer hesitations at non-grammatical junctures in the less complex sentences, which was interpreted to indicate a better grasp of the grammatical structure.

Daniels and Diack (1956) presented four tests of isolated words and two short sentence tests to two groups of seven-and eight-year-olds who had been considered "nonreaders" nine months earlier. Their results indicated there was less likelihood of a word's being read correctly if it occurred in a sentence than if it was presented in isolation. The sentences provided few semantic clues, however.

Goodman (1965), on the other hand, found that most first, second, and third graders in his randomly-selected sample read correctly in story-context one-half to four-fifths of the words they had missed on word lists. No one read a word correctly on the list and missed it consistently in the story although second and third grade children made numerous "one-time substitutions" of words read correctly in isolation.

Schale (1964) suggested that the types of errors made on word lists

may differ from those made on meaningful material; thus, Monroe's (1928, 1932) findings disagreed with Schale's findings because Monroe combined the frequencies of oral reading errors made on the Gray Oral Reading Tests in which the "meaning of words might be inferred from the text" with errors made on words in isolation "thus blurring the results of both testing situations" (p. 103).

No study was found in which errors from word lists were analyzed separately from errors in contextual material except that of Payne (1930) who presented words from the Gray Oral Reading Paragraphs, along with more unfamiliar words and phrases, to four hundred children in grades two through five by means of a tachistoscopic device. The word lists were followed immediately by the reading of the Gray Oral Reading Paragraphs. Payne found that many of the same errors were made in the contextual reading as on the words in isolation; there were slightly fewer errors on the contextual material. Some transfer effect may have taken place since the word lists always immediately preceded the paragraphs.

Hardin and Ames (1969) compared error patterns on two kinds of contextual materials. They found no significant differences in the frequency and rank order of the oral reading errors obtained by twenty-seven disabled readers on the Gray Oral Reading Tests and an informal reading inventory when the subjects were treated as a group. The error percentages of individual subjects showed much variation from one instrument to the other. There tended to be more aid errors on the Gray and more insertion errors on the informal reading inventory but the observed differences were not statistically significant.

Herlin (1963) compared the error patterns of 243 third and fourth grade pupils on the Monroe battery (Gray Oral Reading Tests, Iota Word

Test, and a Word Discrimination Test) with error patterns obtained on comparable subtests of the Durrell Analysis of Reading Difficulty (Oral Paragraphs, Word Recognition Tests, and Word Discrimination Test). Median grade scores were obtained from grade scores on the three tests within a battery and the California Reading Test. Like Monroe, Herlin combined errors from contextual reading and word lists. Errors were categorized according to Monroe's classification system (faulty vowels, faulty consonants, reversals, additions of sounds, omissions of sounds, word substitutions, repetitions, additions of words, omissions of words, and refusals); reading-grade and age-grade norms were then developed for each test battery. Herlin concluded that the reading grade norms obtained on the Monroe tests were similar to Monroe's 1932 norms on the same test battery although there was a tendency for more sound addition errors, repetitions, and refusals to be made in the later study.

Herlin found significant differences in the relative frequency of certain error-types on the Durrell and Monroe tests. The Durrell battery had significantly more errors on consonant sounds, omissions of sounds, substitution of words, and word refusals; while there were significantly more reversals, additions of sounds, and omissions of words on the Monroe tests. No differences were found in the numbers of vowel errors, repetitions, words added, and total errors. Correlations were lowest for omissions and additions of words, omissions of sounds, and reversals. A detailed analysis of the errors of six disabled readers indicated that, although the mean gross error counts differed, the rank order of the error types were similar on the two tests. Diagnosis from gross errors correlated highly with diagnosis from age-grade norms but were lower with reading grade norms. The grade level norms on

the Durrell may have been distorted by the inclusion of the Word Recognition Test scores which Herlin found to be a year higher than the paragraph mean grade-equivalent.

Della-Piano and Herlin (1964) computed error ranks based on gross errors per five-hundred-words-read and on standard (z) scores for forty-three subjects reading one or more years below expectancy (from the Herlin sample) to determine if directions for remediation would differ, using the two types of scores. For most pupils the norms gave no different information or reading-error rank than did the raw error scores. The use of norms was recommended, however, because of discrepancies among vowel errors, substitutions, and additions and omissions of words for a few students. The Durrell and Monroe raw-score error ranks correlated more highly with each other than the raw-score error ranks from either tests correlated with the corresponding normative error ranks. The Durrell-Monroe normative error rank correlations were lowest of all which would seem to raise some questions about the investigators' conclusions.

For the most part, error classifications have been too imprecise or too dissimilar to permit valid and detailed cross-study comparisons of error patterns in different materials. Studies which have used similar error analysis schemes have differed in other important respects. Schummers (1956) used a classification system similar to that of Monroe (1928, 1932) which was also used by Herlin (1963). He found such a large proportion of refusal errors, however, that other error percentages were necessarily deflated. It is probable that the studies differed in the relative difficulty, as well as in the types, of materials. An inspection of Table VII indicates similar error percentages for word

TABLE VII

RESULTS OF STUDIES USING MONROE'S CLASSIFICATION SYSTEM

Error-Types	Monroe Grade 3	Herlin (Mon.) Grade 3	Schummers Grade 3	Monroe Grade 4	Herlin (Mon.) Grade 4	Herlin (Dur.) Grade 4
Vowel-Consonant Errors	42.6%	30.5%	24.8%	37.9%	33.6%	33.3%
Reversals	7.0%	11.8%	3.2%	5.8%	11.5%	7.0%
Addition of Sounds	10.5%	13.8%	4.5%	9.3%	11.0%	10.3%
Omission of Sounds	15.0%	10.1%	6.4%	16.7%	14.5%	17.2%
Substitution of Words	6.0%	1.8%	7.2%	6.9%	3.0%	3.4%
Addition of Words	4.4%	3.4%	2.0%	7.8%	2.4%	1.8%
Omission of Words	3.7%	5.2%	3.4%	5.4%	4.0%	2.4%
Repetitions	6.5%	10.2%	6.7%	5.2%	14.0%	9.5%
Refused Words and Words Aided (Hesitations)	4.4%	13.2%	41.7%	4.9%	6.0%	15.2%

substitutions, word omissions, and repetitions in the Schummers and Monroe (third grade norms) studies.

Leibert (1965) used the classification system devised by Bennett (1942) to compare types of errors made on words from the Word Recognition subtest of the Gates Advanced Primary Reading Test under three test situations: Recall in which the subjects pronounced the words presented in isolation; Recognition 1 in which subjects marked the words pronounced by the examiner; and Recognition 2 in which subjects marked the words indicated by pictures. The age and type of subjects as well as the materials and test conditions differed in the Leibert and Bennett studies. Leibert's sample consisted of second-grade pupils, whereas Bennett's subjects were older disabled readers. Leibert's tests were isolated words, while Bennett's materials consisted of groups of sentences. As may be seen from Table VIII, beginning correct-ending incorrect errors were most prevalent under all test conditions; Bennett's disabled readers made more reversal and medial vowel errors.

Effect of difficulty level. In Spache's (1950, p. 442) opinion, "...errors probably change in nature according to the difficulty of the material being read." Schale (1964, p. 108) concluded from her study that the "...mean number of each type of error on passages at and above grade level for each grade differs somewhat from mean errors made only on grade level passages."

McCracken and Mullen (1970) classified the oral reading errors made by 170 pupils in grades one through six on the Standard Reading Inventory into seven error types: repetitions, words pronounced by the examiner, mispronounced words, omissions, additions, substitutions, and misread punctuation. A tentative analysis indicated a shift in error pattern

TABLE VIII

RESULTS OF STUDIES USING BENNETT'S CLASSIFICATION SYSTEM

	A	B	AB(A)	AB(B)	AB	Medial Vowel Only Incorrect	Final "s" Omission	Reversals	Substitutions
Bennett (1942)	31.0%	16.0%	6.0%	4.0%	8.0%	15.0%	5.0%	12.0%	3.0%
Leibert Recall (1965)	37.1%	0	10.0%	21.4%	11.4%	5.7%	12.9%*	---	1.4%
Leibert Recognition[1] (1965)	55.6%	0	25.9%	0	7.4%	7.4%	0	---	3.7%
Leibert Recognition[2] (1965)	60.7%	8.9%	16.1%	3.6%	8.9%	1.8%	0	---	0

A One, two, or more letters alike at beginning but unlike at the end
B One or more final letters alike but unlike at the beginning
AB(A) Beginning and ending alike, but more alike at the beginning
AB(B) Beginning and ending alike, but more alike at the ending
AB Beginning and ending alike (also alike in form and number)
Medial Vowel Type - Differences occur only in the medial vowel
Substitutions - No common beginning or ending
Reversals - Orientation of words or letters incorrect.
Omission of final "s"

*Omission of ending

between the maximum instructional level and the first level of frustration. The investigators concluded:

> No significant shift in error pattern seems to exist between successive levels if both are in frustration or both in instructional.... This implies that instructional level errors should be used in determining instructional needs and that using errors made at frustration level to determine instructional needs may lead to incorrect instructional programs. (McCracken and Mullen, 1970, p. 110)

The influence of material-difficulty on types of oral reading errors was investigated by Laurence McLeod more than fifty years ago. He (1918, p. 532) observed that

> ...increasing the difficulty of reading material causes an increase in the percentage of gross and minor mispronunciations, a decrease in the percentage of omissions, insertions, and repetitions, and little change in the percentage of substitutions.

The subjects in McLeod's study were in the advanced sections of their grade level; thus, his findings may not be applicable to other kinds of readers.

The effect of the difficulty level of the material on the prevailing error patterns has been investigated in three doctoral dissertations, all with nonselective populations. Schale (1964) compared the types of reading errors made on grade-level passages from the Gray Oral Reading Tests by fifteen boys and fifteen girls at each even-numbered grade level, two through twelve inclusive, with errors made on two passages below the grade placement level and two passages above. Although all pupils were able to read the grade-level paragraph successfully, a few subjects at each grade level were unable to read one or both of the above-grade-level passages; while others were able to read considerably above their grade placement level.

Schummers (1956) studied the effect of increasing difficulty levels

on the error patterns of 237 third-grade pupils. All pupils, regardless of reading ability, attempted five stories from Lyons and Carnahan basal reader series ranging in readability from 1.7 for Story 1 to 5.6 for Story 5. Errors were combined for different groups of subjects but in no combination was the reading ability of the pupils taken into consideration, except indirectly as it was reflected by IQ or rather broad bands of reading accuracy.

Christenson (1966) compared the oral reading errors on independent, instructional, and frustration-level passages of an informal reading inventory for 68 fourth, fifth, and sixth grade subjects.

Unfortunately, neither Schale nor Schummers controlled the relative difficulty of the materials for individual students so that the passages above grade level may not have been as difficult for some subjects as the passages at or below grade placement were for others. The unusually large number of hesitation (refusal) errors reported by Schummers (an average of one word refusal in every eleven running words for the forty-five Low IQ boys on the second grade story), and the low word-accuracy levels for the Medium and Low IQ boys suggest that even the "easy" and/or grade-level stories may have been "difficult" for many of the subjects.

Christenson (1966) combined errors on materials which were at, below, and above his subjects' individual reading levels as identified by his informal reading inventory. He used the word pronunciation and comprehension accuracy levels recommended by Betts (1946) to identify independent, instructional, and frustration levels; however, Christenson's designation of independent and instructional levels as the highest levels at which either word recognition or comprehension met the accuracy requirements changed the criteria to an unknown extent. Literal

Table IX

EFFECT OF DIFFICULTY OF MATERIAL ON ERROR PATTERNS

Errors Which Increase with Increased Difficulty of Material			Errors Which Decrease with Increased Difficulty of Material		
Type of Miscues	Grade	Investigator	Type of Miscues	Grade	Investigator
No response or refusals	2nd	Schale	Omissions	2-12	Schale
	4th	Killgallon		4-6	Christenson***
	4th > 6th	Christenson	Repetitions	2-12	Schale
				4-6	Christenson****
(Hesitations)*	3rd	Schummers	Spontaneous Corrections	3rd	Schummers
Mispronunciations	4th > 5th	Christenson			
(Partial and Gross Mispronunciations)*	2-12	Schale			
(Omission of Sounds) (Vowel Errors)**	3rd	Schummers			
(Wild, Inappropriate Guessing)*	4th	Killgallon			
Reversals	3rd	Schummers			
	4th	Killgallon			

* Word or phrase within parenthesis refers to the term used by the Investigator for the miscue.
** Kinds of mispronunciation errors found to increase with story - difficulty.
*** Non-significant
**** Significant - difference between Independent - Instructional and Independent - Frustration levels only.

interpretation of "highest" and "lowest" levels resulted in six subjects' being assigned higher Independent than Instructional levels; seven subjects', higher Instructional than Frustration levels; and one subject, an Independent level which was four book-levels higher than the Instructional level, and three book-levels higher than the Frustration level.

Thus, it appears that errors from materials of varying relative-difficulty levels may have been combined in all of the studies, possibly obscuring differences. Moreover, obtaining all errors on below-grade-level materials from the same "easy" stories and errors on above-grade-placement passages from the same "difficult" stories, as in the Schale and Schummers studies, may magnify any tendency of the stories' sentence structure or choice of words to produce certain types of errors, thus confounding the effect of difficulty-level with that of materials.

As the difficulty of the material increased, in the studies cited above, the relative frequency of certain errors tended to decrease; other errors tended to increase; and still other errors apparently were not affected by the difficulty of the material (Schale, 1964; Christenson, 1966; Schummers, 1956). As may be seen from Table IX, the relative frequency of repetitions and omissions, as compared with total error occurrence, tended to decrease as the material became more difficult. Schale (1964) found the decrease to be significant for both omissions and repetitions. Christenson (1966) found a significant decline in repetitions between the independent-instructional and independent-frustration levels; a tendency for omissions to decrease as material became more difficult, although not significant, was consistent for the total group, for boys and girls, and for the fourth, fifth, and sixth grades analyzed separately. A steady decline in word recognition

accuracy as the relative difficulty of the material increased was reported by Schummers (1956) and by two later investigators (McCracken, 1961; McCracken and Mullen, 1970). The number of spontaneous corrections decreased from the third-grade level story to the fifth-grade level one for Schummers' (1956) third-grade sample.

The relative number of mispronunciations, no response errors (called refusals by Christenson and hesitations by Schummers), and reversals tended to increase as the story-difficulty levels increased. Schale (1964) reported an increase in both partial and gross mispronunciations. Schummers (1956) found a significant increase in frequency of sound omissions and vowel errors in above-grade level stories. Killgallon, in an earlier study, (1942, p. 106) observed that over 40 per cent more pupils made errors by guessing in frustration-level material on an informal reading inventory, with the guesses tending to become "extremely wild and inappropriate." (Killgallon's fourth-grade sample contained both good and poor readers.)

Schale (1964) found a greater number of no response errors in above-grade level passages for the second-grade subjects; at fourth grade, however, there were fewer no response errors in the more difficult paragraphs, above fourth grade, no response errors were very infrequent. Schummers (1956) observed a tendency for the number of hesitations (no response errors) to increase as the story levels increased, especially among the total Low IQ group and the Low and Medium IQ boys. Killgallon (1942) found that the number of pupils making refusal errors at frustration level (as compared with the number at instructional level) increased over 50 per cent; the refusals were less frequently preceded by attempts to pronounce the words.

Schummers (1956) reported that reversals tended to increase as story-levels increased. Letter, word, and partial reversal errors appeared only at frustration level among Killgallon's (1942) fourth grade sample. No reversal errors were observed in the Christenson (1966) and Schale (1964) studies (possibly because of their more limited definitions of reversals). Christenson (1966) found a significant increase in medial errors at the frustration level.

Schale (1964) concluded that the difficulty of the material did not affect the per cent of total errors of substitutions, insertions, and no response errors (above the second grade). Christenson (1966) found a significantly greater percentage of substitutions at the instructional and frustration levels than at the independent level, but little difference in the relative number of substitutions at the instructional and frustration levels.

Effect of passage length. Stuever (1969), using the same pupil sample as the present investigator, found that more errors were made on the first twenty-five words read than on the second twenty-five words by all groups except the subjects reading the 1.5 level story, with a tendency for a disproportionate number of behavioral and structure errors to occur. She surmised that the use of too-short passages (fewer than 125-150 words) would result in a spuriously-high error ratio (resulting in a too-low instructional level) and a distorted error pattern.

Shedd (1968) reported more errors on the first paragraph of the Gates-McKillop Oral Reading Test than on the second paragraph for 52 per cent of the students enrolled in a university summer reading program. Dunkeld (1970) found that 37 per cent of the subjects made lower word-recognition scores and 59 per cent made lower comprehension

scores on the first passage that they read than they made on a later passage. The effect on error patterns was not determined in the Shedd and Dunkeld studies, however.

Significance of Oral Reading Errors

Authors of standardized oral reading tests, like those of informal measures, disagree as to the significance of various reading errors. Ray (1969), in a preface to his qualitative analysis of oral reading behavior, stated:

> A cursory examination will show disturbing differences between clinical instruments at the quantitative level of analysis relative to error importance, error types, and error frequency. In the clinical instruments examined, equal weight is given to those error types selected as criteria for oral reading performance... The clinician may well question giving the same consideration to a hesitation error as to a mispronunciation error; yet, an evaluation of reading performance using the Gilmore or Durrell would require equal consideration. The current professional interest in re-evaluation of the concept of instructional level may be related to lack of commonality between instruments of measure.

To determine the types of reading errors most related to reading growth, Monroe (1928) computed coefficients of correlation between the frequency of various reading error-types and the median reading-grade score of normal and "retarded" readers. She concluded that

> ...a child may advance in reading grade even though he may omit words or parts of words, add superfluous words, substitute words, and repeat words frequently. On the other hand, confusing the orientation and sequence of letters, mistaking vowel sounds, mistaking consonant sounds, and adding sounds to words are more important errors from the standpoint of progress in reading than are the former.

Gilmore (1947) investigated the relationship of eight types of

reading errors to the oral reading score on the Gilmore Oral Reading Test, reading rate, oral comprehension, and the silent comprehension score on the Stanford Reading Test. He concluded that word substitutions were the most important errors studied and tended to be related to poor oral and silent reading comprehension, poor oral reading scores, and slow oral reading rate. Words-pronounced (by the examiner), which were second in importance, were related to poor silent reading comprehension, a slow rate of reading, and to some extent, to oral reading comprehension. Corrected substitutions were related only to poor oral reading scores. Mispronunciations were made more frequently by good readers than by poor readers in the intermediate grades. Repetitions, hesitations, insertions, and punctuation errors were not related to comprehension, rate, or oral reading score. Omission errors had been made too infrequently in the pilot study to be included in the major investigation.

An inspection of Tables II and X indicates the least agreement among investigators and test writers in relation to the significance of repetitions and corrected errors. Goodman (1967, 1968), who views the reading process as a "psycholinguistic guessing game" in which the reader responds to minimal visual cues, hypothesizes that regressions result from the use of minimal cues which do not produce a satisfactory guess. Allen (1969), Burke (1969), Y. Goodman (1967), and others who accept Goodman's theory of the reading process consider regressions to be correction attempts on the part of the reader which may be responses to a miscue but are not miscues in and of themselves (Goodman and Burke, 1968).

McCracken (1968), who includes corrected errors in the total error

TABLE X

TYPES OF ERRORS INCLUDED ON STANDARDIZED TESTS

Error-Type	Durrell	Gates-McKillop	Gilmore	Gray	Silvaroli	Diagnostic Reading Scales	SRI
Substitutions	X	X	X	X	X	X	X
Mispronunciations	X	X	X	X	X	X	X
Words Pronounced by Examiner	X	X (Hes. 5 sec.)	X (5 sec.)	X* (5 sec.)	X	X (5 sec.)	X (5 sec.)
Insertions	X	X	X	X		X	X
Omissions	X	X	X	X	X	X	X
Repetitions	X (1 word)	X[3] (2 words)	X[1] (1 word or part)	(1 word)	(1 word)	(2 or more)	Word**
Corrected Errors			X (Mispronunciation or substitutions)				X**
Hesitations	X	X (5 sec.)	X (2 sec.)				
Omitted Punctuation	X		X				if inserted elsewhere**
Inversion (Word Order)				X		X	

*5 sec. if no response; 10 sec. if trying to pronounce

**total errors

count but not in the word recognition score on the Standard Reading Inventory, justified the counting of such errors because corrections slow down children's reading and may, therefore, affect whether or not they do read. Goodman (1968, p. 19), who is more interested in the reading strategy employed by the reader than in word accuracy levels per se, suggests that "...the most significant factor in analyzing any miscue may be whether or not it was corrected...and under what circumstance."

Goodman and his students have conducted a series of investigations into factors which influence whether or not miscue-corrections will be attempted. Allen (1969, p. 86) concluded that, since most miscues are syntactically acceptable, "the degree to which a miscue is semantically acceptable seems to determine whether or not the miscue will be corrected." The reader tends not to correct errors that are totally acceptable syntactically and semantically, or errors that are totally unacceptable semantically (Allen, 1969; Burke, 1969). Goodman and Burke (1968) and Burke (1969) have reported the observation of many instances of a reading pause following a miscue during which the reader may have reread preceding material before continuing to read. The investigators assumed that a "silent correction" was made.

Dunkeld (1970) compared word recognition scores on an informal reading inventory, using six different combinations of errors, with the comprehension score (used as a criterion of passage difficulty) of a stratified random sample of the subjects in his study. Correlations between word recognition and comprehension scores above passage-level four were generally low and variable.

Evidence of "the diminishing relationship between word recognition

and comprehension" at higher grade levels reported by Dunkeld (1970, p. 62) has been found by other investigators, also. Packman (1970) found a higher relationship between oral comprehension and word recognition scores among the poor readers than among the average or good readers in her fourth-grade sample. Allen (1969) reported a negative correlation of .72 between the number of miscues per hundred words and per cent of comprehension at the second grade level, but found little relationship between the number of miscues and comprehension with his fourth or sixth grade group. Burke (1969) found no direct relationship between the number of miscues and comprehension with proficient sixth grade readers.

TABLE XI

Correlations Between Comprehension Scores and Six Different Word-Recognition Scores

Passage Level	No. of Cases	W-R 1	W-R 2	W-R 3	W-R 4	W-R 5	W-R 6	Symptom Score
P	18	.48	.55	.50	.50	.49	.54	.39
1	24	.56	.74	.74	.75	.73	.72	.67
2	24	.60	.66	.67	.69	.68	.64	.69
3	21	.13	.44	.47	.47	.49	.48	.54
4	24	.15	.35	.18	.32	.37	.37	.02
P-4	111	$.38\frac{2}{5}$	$.54\frac{4}{5}$	$.51\frac{1}{5}$	$.54\frac{3}{5}$	$.55\frac{1}{5}$	.55	$.46\frac{1}{5}$

W-R Score 1 Substitutions only
W-R Score 2 Conventional miscues
W-R Score 3 Conventional miscues plus repetitions of two or more words
W-R Score 4 Conventional miscues plus repetitions
W-R Score 5 Conventional miscues plus repetitions and corrections
W-R Score 6 Conventional miscues plus corrections

Insofar as oral comprehension level is a valid indicator of error-significance, the use of conventional miscues (substitutions, insertions,

omissions, mispronunciations, and examiner-pronounced words) only, conventional miscues plus corrections, or conventional miscues plus corrections and/or repetitions appears to be equally defensible for reading levels of primer through fourth grade. (See Table XI which is taken in part from Dunkeld, 1970, p. 62.) McCracken (1966) reported higher reliability for total errors (conventional miscues plus corrections and repetitions) on the Standard Reading Inventory than for word recognition errors (conventional miscues). Dunkeld (1970) found almost no relationship between degree of comprehension and symptoms of reading difficulty above the third-reader level. Likewise, head movement, finger pointing, and vocalization were not related to comprehension levels of good or poor fourth grade readers in Packman's (1970) study.

Other issues related to the counting and classifying of reading errors include the "weighting" of significant errors, the counting of repeated error-responses on the same words, and the scoring of dialectal variants. No studies could be found which provided evidence relevant to these issues; rather, procedural rationale appeared to stem primarily from theory or opinion.

Within an error-type, some miscues appear to be relatively serious from the standpoint of the effect on comprehension, while others seem unimportant. Dunkeld (1970) attempted to "weight" errors according to their severity but could find no objective way of doing so. He concluded that since small errors increase as material becomes more difficult, the accumulative effect may make them important. On the Standard Reading Inventory (McCracken, 1966) "small word" substitutions such as a for the are counted as total errors but not as word recognition errors above the first-grade reading level.

Some investigators count repeated miscues on the same words as separate errors each time they occur; others count them only once. On the Standard Reading Inventory (McCracken, 1966), proper nouns and "difficult" words at a particular story-level are counted only once (unless the error was corrected by the examiner the first time it occurred, in which case it is counted twice); all other errors are counted each time they occur. Dunkeld (1970, p. 52) counted repeated miscues as separate errors in his investigation because

> ...some children are able to use the context to help them overcome a pronunciation difficulty. Others are not. Counting repeated miscues as miscues, except when they occur on proper nouns, differentiates between the children who succeed and the children who do not.

In her review of studies related to oral reading errors, Weber (1968, p. 104) commented:

> Another problem arises which weakens the value of the oral reading tests as diagnostic tools....Evaluating the oral reading of a child with noticeable deviations from standard English pronunciation is problematic since the tests fail to distinguish clearly between a reading error and an error that is simply not acceptable in standard speech....The child who has acquired the fundamentals of the reading process but has not mastered control over standard pronunciation may well suffer unjustifiably on oral reading tests since the tests provide no substantive instructions to the examiner on how to handle variations from his own speech.

Recent investigators (Stuever, 1969; Dunkeld, 1970; Allen, 1969; Goodman, 1969; Y. Goodman, 1967) have treated certain dialect differences and speech variants involving pronunciation as acceptable responses and not as miscues (e.g., fella for fellow, goin' for going, axed for asked). Syntactic differences such as we was for we were and semantic differences such as bucket for pail were treated as miscues by these investigators.

Kasdon (1970), however, considered the following deviations to be

"use of dialect" on the part of Negro students rather than reading errors: additions or omissions of 's, substitution of was for were, substitution of mens or menzes for men, pronunciation of except as egskept or eskept, and with pronounced wif.

Dunkeld (1970, p. 52) defended the counting of grammatical deviations as miscues "because appearances suggest that they are indicative of reading difficulty in the same way, for example, that transpositions are indicative of another.

Y. Goodman (1967) concluded that dialect miscues did not affect the reading comprehension or proficiency of six beginning readers. Nurss (1970) found a greater number of dialect errors among the high-achieving third grade readers than the low-achieving group. She interpreted this to mean that dialect errors indicated better comprehension.

Error classifications. Although Payne (1930) considered the specific error-types made by a reader to be largely the product of chance, other investigators believe that errors are not haphazard in nature but reflect to a great extent individual reading patterns and difficulties (Bennett, 1942; Goodman and Burke, 1968). By analyzing a sufficient number of oral reading errors, specific instructional needs can be identified (Durrell, 1935; Monroe, 1935; Ray, 1969).

Mulroy's (1932) results support the latter point of view. Mulroy found that two months of remedial instruction based on analyses of errors made on Gray's Oral Reading Check Tests produced a greater gain in silent reading (as measured by the New Stanford Reading Test) than was made by a control group which had only instruction in silent reading.

There is little consistency in the classification of particular errors, however. As Harris (1970, p. 198) has pointed out:

> ...that for there might be classified variously as wrong ending, wrong several parts, vowel error, consonant error, or two or more of these. No uniform rules for tabulating errors have been established.

There are basically two approaches to error classification: (1) the visual-perceptual approach of Gates (1947, 1962) in which the discrepancy between the stimulus word and the oral response is analyzed in terms of letters; and (2) the phonic or sound-symbol-relationship approach of Monroe (1928, p. 373) in which "...the comparison made between the child's mispronunciation and the original word is primarily one of sounds."

"Any classification of errors will depend very much upon one's concept of what reading involves" (Daniels and Diack, 1956, p. 36). Those utilizing a visual perceptual approach consider that the error resulted from "inadequate attention to the visual pattern of the word" (Durrell, 1935) or the ignoring of certain letter or word-parts. Errors are classified according to the position of the error in the word (or the portions of the word that is correct). Gates' (1962) visual perception errors include: (1) wrong beginning; (2) wrong middle; (3) wrong ending; and (4) wrong in several parts. Errors in which the order of letters is incorrect (orientation errors) are categorized, in the Gates' classification system, as full reversals or reversal of parts. The possibility that the error may have been caused by the association of the wrong sound with a symbol is ignored.

Test authors and investigators who have used a visual perception analysis of errors include: Gates and McKillop (1962), Bennett (1942), Weber (1970), Leibert (1965), Bond, Clymer, Hoyt (1955), Bond, Balow, Hoyt (1970), and Daniels and Diack (1956). Schummers (1956) and Christenson (1966) classified errors by position in addition to another

classification system.

There are differences in error classification even among those using a visual perception approach. Some investigators (Gates and McKillop, 1962; Schummers, 1956; Christenson, 1966; Daniels and Diack, 1956) classified by the elements incorrect, while others (Bennett, 1942; Weber, 1968; Leibert, 1965) looked for the correct features of the word. The results are not entirely comparable, as may be seen by Table XII

There is even less agreement about what constitutes a medial (or beginning, or ending) error. Bennett classified grate for gate as beginning and ending correct (medial error). Daniels and Diack, likewise, classified shop for step and bought for brought as medial errors. On the 1955 edition of the Silent Reading Diagnostic Tests, Form D-A, brook for book, slur for star, failing for flying, and bush for brush were all designated as medial errors. On the 1970 revision of the test, failing for flying is still a medial error, but sleps for steps, gapes for grapes, and spool for stool are classified as initial errors.

The Gates-McKillop (1962) norms for middle third grade indicate approximately five times as many ending errors as middle errors

> ...partly because many of the words in the first four paragraphs of the test are of one or two syllables and hence have no clearly pronounced middle part or syllable. (Gates, 1947, p. 214)

(Row for raw is considered a medial error, however. Gates-McKillop, 1962). Weber (1968, p. 111), on the other hand, concluded that

> ...given some training in reading, children make most errors in the middle of words, fewer on letters at the end of words, and fewest on letters at the beginning.

Investigators using a phonic or sound-symbol relationship classification assume that the error is caused by faulty sound-symbol associations and classify according to "sound" elements. The categories used

Table XII

RESULTS OF STUDIES USING POSITIONAL CLASSIFICATION SYSTEMS

		Wrong Beginning	Wrong Ending	Wrong Middle	Wrong in Several Parts	Reversals and Partial Reversals
Gates-McKillop norms	3rd gr.	0-6%	53-67%	11-12%	22-29%	0-6%
Schummers	3rd gr.	15.09%	50.37%	34.55%		
Christensen						
Instructional	4-6	11.3%	18.2%	14.9%	55.6%	
Frustration	4-6	9.9%	14.9%	27.2%	48.0%	
Bennett*		19. %	39%	33%		12%
Weber*	1st	47%	69%			
Leibert*						
Recall	3rd	1.4%	48.6%	48.5%		
Recog. 1	3rd	3.7%	59.3%	40.7%		
Recog. 2	3rd	8.9%	60.7%	30.4%		

* May be wrong in several parts

	Beginning Correct, Ending Incorrect	Ending Correct, Beginning Incorrect	Both Beginning and Ending Correct	Both Beginning and Ending Incorrect	Final "S" Omission	Reversals
Bennett (1942)	31%	16%	33%	3%	5%	12%
Leibert (1965)						
Recall	37.1%	0	48.5%	1.4%	12.9%	
Recog. 1	55.6%	0	40.7%	3.7%	0	

Table XII (Continued)

Recog.	2	60.7%	8.9%	30.4%	0	0

	Beginning Correct	Ending Correct	Beginning-only correct	Ending-only correct	Stem correct
Weber	53%	31%	11%	7%	14.1%

by Monroe (1928, 1932) include: (1) faulty vowels: (2) faulty consonants; (3) additions of sounds; and (4) omissions of sounds. A reversals category was included for errors "caused by" improper sequence of letters or words and/or letter rotations. The possibility of visual perception errors are ignored by this classification system; such errors are classified under sound categories.

Investigators who have used sound-symbol classifications include: Monroe (1928), Gray (1963), Killgallon (1942), Schale (1964), Herlin (1963), and Schummers (1956).

Ray (1969, p. 8) terms both of the above approaches "one-dimensional approaches" and states:

> A disabled reader whose error profile was diagnosed by both /Gates' and Monroe's/ classification systems could easily have been placed in diametrically opposed remedial programs.

The Ray Error Analysis System (Ray, 1969) attempts to utilize the most desirable features of both the visual perception and visual auditory approaches. "Decoding" errors are assigned to the visual perception category if the response to the word is instantaneous and to the visual auditory if there are discernible efforts to "sound out" the word. The B-S-R Error Analysis (used by Stuever, 1969, and in the present investigation) also utilizes this approach.

A highly detailed taxonomy "of cues and miscues in reading" has been devised by Goodman (1968) to examine the reading strategies employed by the reader. This system analyzes various aspects of the miscue (graphic proximity, phonemic proximity, semantic proximity, syntactic proximity, and several others) on a number of linguistic levels. The system is too intricate to be used by classroom teachers and clinicians, but has been used in research studies by Y. Goodman (1967),

Goodman (1965), Goodman and Burke (1968), and Allen (1969).

Summary

A survey of the literature revealed numerous studies which have compared grade-equivalent scores on standardized and informal reading tests with randomly-selected or total-classroom groups, but only a few which have compared scores on oral standardized reading tests, and fewer still which have used disabled-reader groups.

Several studies have found that disabled readers tend to be less accurate in oral reading than normal readers of the same reading level (Watkins, 1953; Monroe, 1928; Packman, 1970). From a comparison of the results of other investigators, it appears that disabled readers may be less penalized by a scoring system based primarily on rate of oral reading than one based on oral reading accuracy (Attea, 1966; Long, 1959; Hardin and Ames, 1970; Patty, 1965). No study was found which investigated this variable. Since most oral diagnostic reading tests are used with disabled readers, an investigation of the comparability of grade scores on well-known reading tests for this reader group appears to be warranted.

Two investigators reported differences of three-fourths year or more between grade-scores on the oral paragraphs and the word lists on the Durrell Analysis of Reading Difficulty for randomly-selected third and fourth grade children. Since differences between words in isolation and contextual reading levels are often used diagnostically with disabled readers, it is important to determine if a similar difference would be found with disabled reader groups.

Studies comparing error patterns between oral reading tests have

tended to use broad categories (e.g., substitutions) which may have obscured possible differences between tests (Hardin and Ames, 1969) or to use one-dimensional kinds of classification systems which ignored certain kinds of errors (Herlin, 1963). If oral reading errors are to give direction to remediation, then the commonality of error patterns among oral reading instruments, for disabled readers, should be investigated using a precise error analysis system.

Several investigators have reported that error patterns vary according to the level of difficulty of the material (Schale, 1964; Christenson, 1966; Schummers, 1956); however, these investigators did not adequately control the relative difficulty level of the test selections for individual subjects, possibly distorting or obscuring the results. Further study of this question is warranted.

From a review of the literature, it appears there is justification for an investigation of the commonality of error patterns and reading rates for disabled readers between tests and between levels of reading performance, and for a comparison of grade-level equivalents on different types of reading instruments.

CHAPTER III

DESIGN AND METHODOLOGY

This chapter contains a description of the population of the study, the testing procedures, the test instruments used in collecting the data, and the statistical treatment of the data.

Description of the Population

The population for this study[1] consisted of all the fourth-grade pupils in the public and parochial schools of a county in north central Oklahoma who were average or above in intelligence and reading below grade level. The schools ranged from rural, two-teacher schools to those in which two or more fourth-grade classrooms were contained in the same building. The reading programs utilized basal readers; a phonics-oriented basal series was also used by some teachers. A cross section of socio-economic levels was represented. The fourth-grade population was primarily Caucasian, but also included pupils of American Indian, Negro, and Spanish-American extraction.

Pupils meeting the criteria for the study sample were identified through a three-step screening process:

1. A survey was made of fourth-grade classrooms in the city and and town schools to obtain the names of all pupils whose reading

---

[1]This study was one of three independent studies utilizing the same pupil sample.

ability fell in the lowest one-third of their class.

2. The Stanford Achievement Test, Primary II, Form W: Reading Section was administered to the poor readers identified by the classroom survey and to all of the fourth-grade pupils in the smaller schools. (The Primary II test was considered more appropriate than the Intermediate I level since these were poor readers.) A total of 415 tests were administered by the investigating team.[2] Those pupils who scored at or below the 4.0 grade level on the Word Meaning and Paragraph Meaning sections of the Stanford were given the Peabody Picture Vocabulary Test, Form A. All of this group who obtained intelligence quotients of 80 or above were administered the final screening tests, consisting of the Standard Reading Inventory and the Wechsler Intelligence Scale for Children.

3. Pupils whose maximum instructional levels, as determined by the Standard Reading Inventory, Form B, fell at or between 2^1 and 3^2 and whose full-scale IQ on the Wechsler Intelligence Scale for Children was 90 or above were administered the test battery used in the three studies. Pupils with known physical handicaps which would interfere with their reading the test materials (or with the subsequent analysis of oral reading errors) were not included in the study.

The sample, as identified by the three steps above, consisted of ninety-two children. This number was subsequently lowered to seventy-seven for the following reasons: (1) failure to meet the original criteria; (2) incomplete test data; or (3) inaudible or incomplete

---

[2]The investigating team consisted of Rita Stuever, Bettie Vanice, and the writer, all of whom collected data for separate dissertation studies.

tape recordings. Pupils from twenty-eight of the thirty-four elementary schools in the county were included in the study. Four of the twelve rural schools and two of the four schools in towns of 350 persons or less were not represented. The distribution of the pupils among the rural, small-town, and city schools was not changed by the reduction in sample size. Screening test data for the total sample and for boys and girls separately are contained in Table XIII.

Testing Procedures

The Stanford reading test was administered by the investigating team in classroom settings according to directions in the test manual. All other tests were administered individually by members of the investigating team or other trained examiners from the Oklahoma State University Reading Center. Subjects were taken from the classroom to a suitable area where the testing could proceed undisturbed, with only the subject and the examiner present. The children were told the purpose of the testing and asked if they would be willing to assist the examiners. All children agreed to participate in the study; most children enjoyed the testing sessions.

The twelve tests used in the three studies were counter-rotated and placed in twelve different testing sequences so that, within the twelve test-orders, each test was preceded once and followed once by every other test; and each test was administered first once, second once, third once, and so on. Subjects were randomly assigned to testing-sequences in such a way that each test order, I-XII, was followed approximately the same number of times. This was done so that no one test would be consistently affected by such intervening variables as

TABLE XIII

SCREENING TEST DATA FOR STUDY SAMPLE

	n	CA Mean Range	PPVT Mean Range	WISC Mean Range	Stanford Reading Test WD MNG Mean Range	Stanford Reading Test PAR MNG Mean Range	SRI Mean
Boys	55	9-11 (9-4 to 10-11)	101 (81-138)	106 (90-117)	3.2 (2.3 - 3.8)	3.0 1.9 - 3.9	2.9
Girls	22	9-11 (9-6 to 10-11)	95 (82-118)	101 (90-114)	3.1 (2.3 - 4.0)	3.1 1.9 - 3.9	2.8
Total	77	9-11 (9-4 to 10-11)	99 (81-138)	104 (90-117)	3.2 (2.3 - 4.0)	3.0 1.9 - 3.9	2.9

test fatigue, anxiety at the beginning of a testing session, or the learning of words in one test which appeared on a subsequent test.

The tests were administered according to the detailed Manual of Directions given each examiner, and in the order indicated by the test-sequence number assigned to the subject. During the test administration, the examiners recorded the miscues made by the subjects on copies of the test selections. All tests used as variables in this study were tape-recorded; the errors and times required for reading were carefully rechecked by this investigator.

Although the administration of the test batteries to the ninety-two subjects extended over approximately six weeks, all of the tests for any one subject were completed within a week and were administered by the same examiner.

Instruments Used

Stanford Achievement Test, Primary II, Form W: Reading (1964)

This test was used to identify pupils who were below grade level in reading ability. It measures two aspects of reading achievement: word meaning and comprehension. The Word Meaning Test consists of 36 multiple-choice items, each of which requires a selection from four alternatives of the correct word to complete a sentence. The sentence may define the word or ask for a synonym. The Paragraph Meaning Test contains 31 paragraphs (60 multiple-choice items) of increasing difficulty from which one or more words have been omitted. The correct word for each omission must be selected from among four choices. (Two paragraphs are followed by questions rather than utilizing the modified close technique.) Paragraphs range in length from 10 to 99 words.

This test measures, primarily, the ability to recognize facts or details and to make inferences.

Content or curricular validity is based on the content of the typical elementary school curriculum. Typical courses of study and textbooks were examined by the authors; experimental tests were tried out and the items reviewed by a number of classroom teachers and reading specialists. Split-half reliability coefficients of .83 for the Word Meaning Test and .93 for the Paragraph Meaning Test were reported for grade three.

The tests were standarized on a minimum of 10,000 pupils per grade level. Farr (1969, pp. 44,45) concluded that "the Stanford Achievement Tests-Reading are carefully constructed tests for measuring general reading ability" with norms ". . .as representative of actual national student performance as those of any other published test available." Robinson (1968) considered an earlier edition to be among the best survey tests of reading achievement for the elementary grades.

Peabody Picture Vocabulary Test, Form A

This test was used as a gross screening measure of intellectual ability. It indicates auding ability or listening vocabulary levels. The test consists of 150 plates arranged in order of difficulty and 150 stimulus words, each of which is defined or illustrated by one of the four line-drawings on the plate with a corresponding number. The student points to, or otherwise indicates, the picture on the page which best portrays the meaning of the stimulus word pronounced by the examiner. Norms are provided for ages ranging from eighteen months to eighteen years. Any one student is given only the portion of the test which is within his ability range.

Standardization was based on 4,012 white children and youth in and around Nashville, Tennessee. Lyman (1968) considered the PPVT to be a highly usable test of moderate reliability and largely unpublished validity. Neville (1965) found no significant difference between the mean PPVT results and the mean full-scale IQ on the Wechsler Intelligence Scale for Children for 54 fifth-grade children. He tentatively concluded that the PPVT could be substituted for the WISC with poor readers. However, Callaway (1970) found that although the mean scores were similar, IQs on the PPVT averaged nearly ten points lower than IQs on the Stanford-Binet Intelligence Scale for 79 of 177 disabled readers at the University of Georgia Clinic. In the present investigation children with PPVT IQs of 80 and above were retained in the sample for later evaluations by the WISC.

Wechsler Intelligence Scale for Children (WISC)

This test was used as the criterion instrument in defining average-or-above intelligence. The Wechsler Intelligence Scale for Children consists of twelve subtests which are combined into two subscales to yield three measures of intelligence: verbal, performance, and full-scale. The Verbal Scale subtests include those entitled Information, Comprehension, Arithmetic, Similarities, Vocabulary, and Digit Span. The Performance Scale consists of Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding, and Mazes. Ordinarily, only five Verbal and five Performance subtests are administered; however, all twelve tests were given to every subject in the standardization of the WISC and in clinical situations the use of twelve tests is recommended because of the qualitative and quantitative data they add.

Split-half reliabilities for the 10½ age-group reported by

Wechsler are .96 for the Verbal Scale, .89 for the Performance, and .95 for the Full Scale Score. Correlations of .75 to .90 have been found between full-scale IQs on the WISC and IQs on the Stanford-Binet Intelligence Scale (Freeman, 1962).

The WISC was standardized on a sample of 100 white boys and 100 white girls at each age from five through fifteen years. Burnstein (1965) concluded that the WISC is a well-standardized, stable instrument, correlating well with other tests of intelligence.

Standard Reading Inventory (SRI)

This test has two equivalent forms. Form B was used as a screening measure to identify pupils with maximum instructional reading levels of 3^2 or below. The Standard Reading Inventory is an individually-administered reading test for measuring reading achievement at pre-primer through seventh reader levels. Each form contains eleven word lists for testing word recognition in isolation, eleven stories for oral reading, and eight stories for silent reading. Comprehension of the oral and silent reading passages is tested by inference and detail questions. The following levels are identified by the use of a scoring sheet: independent, instructional, and frustration. Separate ratings are made for vocabulary (both in context and in isolation), errors (total and word recognition), comprehension, and reading speed. The standards used in scoring the Standard Reading Inventory are based upon the criteria recommended by Betts (1946).

Content validity is assumed from the manner in which the test was constructed. Words were used in the stories and word lists at the levels in which they were introduced in three basal reader series. Sentence length, content, and general style were also based on the

reader series. The Spache (1968) and Dale-Chall (1948) Readability Formulas were used in analyzing the stories. The difficulty levels of the stories were also evaluated subjectively by twenty-five reading experts. Two studies corroborating the content validity indicated above and two studies of concurrent validity were reported in the test manual (McCracken, 1966). Additional studies of concurrent validity were reported by Botel, Bradley, and Kashuba (1970) and McCracken and Mullen (1970).

Evidence of reliability was obtained in two studies of elementary children who took both forms of the SRI. The correlation between the instructional levels on the two forms was .91 in one study and .95 in the other.

In the present investigation, the Form A word lists were used as a measure of reading level. To save testing time, the first five words were omitted from each word list above the pre-primer level. The level of the most difficult list on which 55 per cent or more of the words were pronounced correctly (providing there were no more than seven consecutive failures) was assumed to be the maximum instructional level (McCracken, 1966, p. 45). The Form A oral stories were used as measures of reading level and rate, and as sources of oral reading errors. Two measures of instructional level were obtained from the oral subtests: one measure was secured from the regular scoring of the SRI, using both total errors and word recognition errors, and a minimum word-accuracy level of 91 per cent (one error in eleven running words); the other measure was based on word recognition errors (in which the total error score was disregarded) and a minimum word-accuracy level of 95 per cent. The regular scoring of the SRI was used for all purposes

except the second or word-recognition criteria of instructional level.

Durrell Analysis of Reading Difficulty, New Edition

The Oral Reading Test and the Word Recognition and Word Analysis Test were used from this well-known diagnostic battery to obtain measures of reading levels (for words in isolation and for oral reading); oral reading rates and errors were obtained from the Oral Reading Test, also. The Word Recognition and Word Analysis Test measures quick word recognition and delayed word analysis techniques. The word lists for grades 2-6 contain a total of fifty words. By means of a hand tachistoscope, the words are exposed one at a time for one-half second. If a word is not pronounced correctly on the flash exposure, the shutter is opened and the child given time to work out the pronunciation. His response to the word on the untimed exposure is recorded under Analysis. The Word Recognition score is the total of the words pronounced correctly on the flash presentation. The Word Analysis score is the total of the words credited on the Word Recognition Test plus the words identified on the untimed exposure. After seven successive errors, the test is discontinued. In this study words were not pronounced for the subjects on the Word Recognition and Word Analysis Test.

The Oral Reading Test consists of eight reading passages, each of which is followed by four to seven comprehension questions measuring recall of details. Testing extends from a basal paragraph in which there are no errors to a paragraph in which seven or more errors are made, or until the time required for reading any paragraph is more than two minutes. The number of errors is used only to determine the appropriate paragraphs to be read. Grade norms are based upon the

time required for reading each paragraph. The median of the grade levels assigned each paragraph is the grade score for oral reading.

In the present study, the grade score on the highest paragraph read successfully was used as a second measure of reading level. The first paragraph on which seven or more errors was made was assumed to be the beginning of FRUSTRATION level for purposes of error and rate comparisons.

No data are reported in the manual concerning the reliability of validity of the tests. Norms were based on "no fewer than a thousand children for each test" but the standardization population is not described. Maxwell (1968, p. 248), in a review of the Durrell for the Mental Measurements Yearbook, stated that the reading paragraphs were well chosen and graded, and the word lists were adequately comprehensive, but pointed out the limitations of the oral and silent reading norms. Nevertheless, the tests have been widely used for many years and continue to be one of the two or three most widely used diagnostic tests (Harris, 1970; Kaluger and Kolson, 1969).

Gates-McKillop Reading Diagnostic Tests, Form I (Oral Reading)

The Oral Reading Test was used as a measure of reading level and rate, and as a source of oral reading errors. It consists of seven paragraphs of increasing difficulty from grade one to approximately eighth grade level which tell a continuous story. Each paragraph receives a raw score dependent upon the number of errors made in reading it. All pupils begin with paragraph one and continue reading until they have made eleven or more errors in two consecutive paragraphs. Comprehension is not checked. The test is scored only for errors; the length of time used by a pupil in reading a paragraph is

not taken into consideration in reaching a grade equivalent. (The revised edition is not timed according to the manual directions.) A detailed analysis is made of the errors made on the first four paragraphs. Norms provide rough comparisons of the errors of a given child with those made by the average child making the same number of total errors.

Spache (1968, p. 251) called the last three paragraphs "highly artificial and stilted" and expressed the opinion that the oral reading test probably

> functions as a measure of ability to read (?) this esoteric material rather than as a test of general oral reading ability. It is also doubtful that the intensive analysis of reading errors suggested by the author is realistically related to the child's true reading performance with ordinary materials.

Bryant (1968) on the other hand, considered the analysis of errors made on the first four paragraphs to be "of particular value."

In the present study, the first paragraph in which eleven or more errors were made was assumed to be the beginning of FRUSTRATION level for purposes of rate and error-pattern comparisons. Since the manual specifies that only errors from the first four paragraphs should be analyzed, two sets of errors were obtained from the oral paragraphs for comparison: those made on the first four paragraphs only; and those made on the first paragraph with two or more errors through the paragraph immediately preceding the first paragraph with eleven or more errors. The paragraphs were timed. For purposes of error-comparisons only, single word repetitions were counted. (Only repetitions of two or more words were included in the reading level computations, however.)

No published data on validity or reliability of the Gates-McKillop are available. Torgerson (1968) referred to an unpublished

study by Gates in which a reliability coefficient of .86 was reported for the oral reading test, based upon correlations obtained between Form I and Form II using 90 cases in the third grade. Gates (1947) stated that the norms for an earlier (and very similar) edition of the test were based on a large number of children in grades 1-5. Nowhere is the norm group clearly identified.

Wide Range Achievement Test: Reading (1946 edition)

This test was used as a measure of reading grade level on words in isolation. It consists of 128 words, printed horizontally rather than in a list, with norms ranging from kindergarten to college. A basal of ten successive correct responses is required; the test is discontinued after ten consecutive errors.

Among the uses claimed for the test by the authors is the determination of instructional levels. Harris (1970) cautioned that the test is not a satisfactory measure of general reading ability. However, it appears to be a valid measure of word-pronouncing ability. The manual reported correlations ranging from .78 to .88 between the WRAT and teacher's estimate of grade level, mid-term reading grades, and grade scores on the New Stanford Reading Test. Reliability coefficients ranging from .90 to .95 are reported by the authors.

B-S-R Error Analysis

The B-S-R Error Analysis was devised by Berends, Stuever, and Ray at the Oklahoma State University Reading Center as a means of combining the visual-perceptual approach to error analysis of Gates with the sound-symbol emphasis of Monroe.

By means of the B-S-R Error Analysis system, errors from the oral reading paragraphs and stories were classified into six major

categories: visual perception--word parts, directional confusion, visual-auditory, syllabic division, structure, and behavorial characteristics.

Visual Perception--word parts. Errors were classified in this category if the response to the stimulus word was made instantaneously with no attempt at sounding-out the word. The assumption was that the child looked at one or more parts of a word and said another word which the part(s) suggested.

1. -++ middle and end correct, beginning incorrect:
   there - where, hungry - angry

2. +-+ beginning and end correct, middle incorrect:
   smelling - smiling, serve - slave

3. ++- beginning and middle correct, end incorrect
   (s, ed, ing were classified under structure):
   you - your, not - nor

4. --+ end correct, beginning and middle incorrect:
   pillow - window, thought - forgot

5. +-- beginning correct, middle and end incorrect:
   nothing - neither, well - with

6. -+- middle correct, beginning and end incorrect:
   hampster - champion, danger - tangle

7. --- word completely wrong; also, error on one-or-two letter stimulus word:
   was - and, away - up

Directional Confusion. Errors were classified in this category if the order of letters or words was incorrect and/or letters were rotated.

1. Rotations: bounding - pounding, dog - boy

2. Reversals: whole and partial reversals: was - saw, left - felt
   word sequence errors: you will

Visual-Auditory. Errors were classified in this category if the response was incorrect after a discernible attempt to "sound it out." Visual-Auditory errors reflect "faulty perception of sound-symbol relationships, faulty application of phonic principles, or lack of application of alternative word recognition techniques to sound-symbol relationships. (Ray, 1970)

1. C Error on a single consonant: rejāned - regained

2. CC Error on a consonant blend or digraph:
fē/fē dŭm' - freedom
orpən - orphan

3. V Error on a single vowel: strōōk - struck,
pĕr'ish - perish

4. VV Error on vowel digraphs or diphthongs:
thrēt - threat,
côr/ cŭr' ĭj - courage

5. CCVV Error on both vowels and consonants:
cŏns/ cŏnkĭt - conceit
grĭ/ grĭd/ groud - gnawed

Syllabic Division. Error caused by wrong syllabic division and/or accent: hū' mant/ hū' məty - humanity
grōw ləd - growled

Structure. This category included errors on contractions, compound words, inflectional endings, prefixes, and suffixes.

Behavorial Characteristics. This category included omissions of whole words, additions of whole words, words aided, repetitions and corrections.

Errors were classified in one category only. Repetitions, additions, and omissions of one or more consecutive words were counted as one error. Repetitions made in conjunction with corrections were not counted as errors. Consonant digraphs were considered single consonants (ship for sip would be a -++ error) for purposes of classifying visual perception errors. Speech variants such as runnin' for running, excape for escape, set for sat, winda for window, pertended for pretended, bre' fəst for breakfast, and sumpthən for something were not counted as errors. We was for we were was counted.

Statistical Techniques Used in the Treatment of the Data

Analysis of Grade-Score Differences

To test hypothesis one, an analysis of variance was performed on the grade-equivalent scores of the nine measures of reading levels, using the IBM 360/Model 50 computer at Oklahoma State University with the Analysis of Variance for Factorial Design program, BMD02V (with one replication), revised May 29, 1968. This program was prepared by the Health Sciences Computing Facility of the University of California --Los Angeles.

The analysis of variance determined whether there were significant differences in the mean grade-equivalent scores on the nine oral reading tests.

The same computer was used with the Duncan Multiple-Range Test program to make multiple comparisons of the means after the F had been found to be significant.

t-Test Comparisons of Mean Differences

Multiple *t*-tests were computed to determine the significance of differences in reading rate and in error proportions among the three oral tests and the various levels of reading difficulty. Since the subjects in this study were paired with themselves, it was necessary to use a *t*-test for paired samples or correlated means. Although there are risks associated with computing many individual *t*-tests, since by chance alone a few *t*-test results may appear to be significant, comparisons of the intermediate values in the levels of reading performance with those above and below, and the comparisons between specific pairs of tests were of interest in this investigation; therefore, it

was decided to use this statistical technique despite its limitations.

The Oklahoma State University computer program of June, 1968, for a Student's t-statistic for single, paired, and unpaired samples was used to test hypothesis two and the intertest comparisons of hypothesis three.

The levels of performance (intratest comparisons) of hypothesis three were tested by the computing by hand of t-tests for correlated samples by means of the formula

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\Sigma d^2}{N(N-1)}}}$$

as explained in Wert, Neidt, and Ahmann (1954, p. 141). The $\Sigma d^2 = \Sigma D^2 - \frac{(\Sigma D)^2}{N}$. The N is the number of pairs, and D is the difference between the scores for each pair. The expression $\bar{X}_1 - \bar{X}_2$ is equivalent to $\frac{\Sigma D}{N}$.

The Kendall Coefficient of Concordance: W

To determine whether or not the error patterns were similar on the three oral reading tests and at the three levels of performance--INSTRUCTIONAL, FRUSTRATION 1, and FRUSTRATION 2, the error-types which were made at each of the above performance levels of each test were ranked according to their relative frequency of occurrence. By use of the Kendall coefficient of concordance, the overall agreement, or degree of association, among the various rankings was ascertained.

Hypotheses four and five were tested by the computing by hand of twelve coefficients of concordance according to the formula in Siegel (1956, p. 231): $W = \frac{S}{\frac{1}{12}k^2(N^3-N)}$

where Rj = the sums of the ranks assigned to each object,

S = the sum of squares of the deviations from the mean of Rj,

k = the number of sets of rankings, and

N = the number of objects ranked.

Correction for tied ranks was made when necessary (Siegel, 1956, p. 234).

The significance of W was determined by the finding of χ^2 by the formula:

$$\chi^2 = \frac{S}{\frac{1}{12}kN(N+1)}$$

$$df = N-1$$

and the subsequent reference to a table of critical values of Chi Square.

Summary

This chapter has described the population used in the study and the test instruments employed in gathering the data necessary for testing the hypotheses. In addition, the statistical techniques used in the treatment of the data have been explained.

CHAPTER IV

TREATMENT OF DATA AND ANALYSIS OF RESULTS

This chapter contains a detailed account of the statistical treatment of the data and the analysis of the results.

This study was concerned with the effect of the testing instrument and the difficulty of material upon the reading performance of disabled fourth-grade children. It included an analysis of the oral reading errors made by the pupils at the INSTRUCTIONAL, FRUSTRATION 1, and FRUSTRATION 2 levels on each of three standardized tests. (See Appendix A for a key to the symbols used to identify levels and test instruments.) Comparisons of the resulting error patterns were made between tests and between levels of performance. Similar comparisons were made of oral rates of reading. Grade-equivalent scores were compared for five measures of contextual reading and four measures of words in isolation.

A hypothesis related to the differences among grade-equivalent scores will be examined first; next the hypothesis concerning reading rate will be discussed; and last , the three hypotheses related to the comparisons of error patterns will be tested.

Tests of the Hypotheses

Hypothesis 1: There are no significant differences among the mean grade-equivalent scores of the nine reading measures.

To test this hypothesis, an analysis of variance was performed on the grade-equivalent scores of the nine measures. The critical F ratio is significant at the .001 level of confidence (as can be seen from Table XIV); thus, the null hypothesis of no significant differences among mean scores can be rejected.

TABLE XIV

ANALYSIS OF VARIANCE OF GRADE-EQUIVALENT SCORES ON NINE ORAL READING TESTS

Source	df	SS	ms	F
Subjects	77	187.25	2.43	
Oral Tests	8	207.01	25.88	114.86*
Residual	616	138.77	0.225	
Total	701	533.03		

*F ratio is significant at the .001 level with values greater than 3.27

The means were further compared by the Duncan Multiple-Range Test to determine which specific pairs of means differed significantly; these data are reported in Table XV. An inspection of this table indicated that the mean grade-equivalent scores of all measures were significantly different except for the following pairs of means: the two scorings of the Durrell Oral Reading Test--the median grade scores and the grade scores from the highest paragraph on which the reader was successful; the Gates-McKillop and both scorings of the Durrell; the SRI Word Lists and the WRAT; and the Durrell Word Recognition and

TABLE XV

COMPARISON OF MEAN GRADE-EQUIVALENT SCORES ON NINE ORAL READING TESTS BY THE DUNCAN MULTIPLE-RANGE TEST

Least Significant Range	Identification of Test	Test Mean	Difference Between Means
S.E. = 0.0537	M1	2.82	M1 vs. M2 = .22*
R_2 = 0.149	M2	3.04	M2 vs. M3 = .33*
R_3 = 0.157	M3	3.37	M3 vs. M4 = .07 M3 vs. M5 = .14 M3 vs. M6 = .39*
R_4 = 0.162	M4	3.44	M4 vs. M5 = .07 M4 vs. M6 = .32*
R_5 = 0.166	M5	3.51	M5 vs. M6 = .25*
R_6 = 0.169	M6	3.76	M6 vs. M7 = .06 M6 vs. M8 = .66*
R_7 = 0.172	M7	3.82	M7 vs. M8 = .60*
R_8 = 0.174	M8	4.42	M8 vs. M9 = .13
R_9 = 0.175	M9	4.55	

*Denotes significance at .05 level.

M1 Standard Reading Inventory (scored on total and word recognition errors)
M2 Standard Reading Inventory (scored on word recognition errors)
M3 Gates-McKillop Reading Diagnostic Tests (Oral Reading)
M4 Durrell Analysis of Reading Difficulty (Oral Reading Test--median grade score)
M5 Durrell Analysis of Reading Difficulty (Oral Reading Test--grade score of last paragraph read successfully)
M6 Standard Reading Inventory Word Lists
M7 Wide Range Achievement Test: Reading
M8 Durrell Analysis of Reading Difficulty Word Analysis Test
M9 Durrell Analysis of Reading Difficulty Word Recognition Test

Word Analysis Tests.

Hypothesis 2: The mean oral reading rates do not differ significantly between tests or between levels of oral reading performance on a single test. (This hypothesis will be examined separately for intertest and intratest comparisons.)

Multiple t-tests were computed to test hypothesis 2. Results of the between-test comparisons are presented in Table XVI. These data indicate the null hypothesis cannot be rejected for the following pairs of means: the GI-4 - SRII, the GI - DF[1], and the DF[2] - SRIF[2]. Hypothesis 2 can be rejected for all other intertest comparisons.

The t values for the intratest comparisons are reported in Table XVII. The t values are significant for all between-levels mean differences; therefore, hypothesis 2 can be rejected for intratest comparisons.

Hypothesis 3: The mean proportion of each error-type or category does not vary significantly between tests or between levels or oral reading performance on a single test. (This hypothesis will be examined separately for each of the following categories or sub-categories on both intertest and intratest comparisons: visual perception errors, visual-auditory errors, directional confusion errors, syllabic division errors, structural errors, corrections, repetitions, and omissions-additions-words aided.)

To test hypothesis 3, multiple t-tests were computed. After the hypothesis has been examined for an error category or sub-category at a specific performance level, significant differences for individual error-types within the category will be presented. The reader should be cautioned that nonsignificant differences in error-type comparisons are

TABLE XVI

BETWEEN-TEST COMPARISONS OF MEAN RATES OF READING

Test	N	S.D.	Mean	t
DI	77	24.35	115.44	
GI-4	77	24.20	105.03	5.23**
DI	77	24.35	115.44	
GI	77	22.85	91.61	12.24**
DI	77	24.35	115.44	
SRII	77	20.27	102.26	8.25**
GI-4	77	24.20	105.03	
SRII	77	20.27	102.26	1.49
GI-4	77	24.20	105.03	
GI	77	22.85	91.61	8.43**
SRII	77	20.27	102.26	
GI	77	22.85	91.61	5.56**
DI	77	24.35	115.44	
SRIF[1]	77	20.71	81.18	14.68**
DI	77	24.35	115.44	
GF[1]	77	18.14	52.68	23.59**
GI	77	22.85	91.61	
DF[1]	77	24.39	87.06	1.98
GI	77	22.85	91.61	
SRIF[1]	77	20.71	81.18	4.69**
SRII	77	20.27	102.26	
DF[1]	77	24.39	87.06	7.13**
SRII	77	20.27	102.26	
GF[1]	77	18.14	52.68	20.39**
DF[1]	77	24.39	87.06	
GF[1]	77	18.14	52.68	12.90**
DF[1]	77	24.39	87.06	
SRIF[1]	77	20.71	81.18	2.41*
SRIF[1]	77	20.71	81.18	
GF[1]	77	18.14	52.68	12.68**
DF[2]	46	23.95	69.78	
GF[2]	46	16.80	41.04	9.41**
DF[2]	42	24.35	72.21	
SRIF[2]	42	20.63	69.38	1.14
SRIF[2]	54	22.03	68.56	
GF[2]	54	17.24	40.76	10.63**

*Significant beyond the 0.05 level.

**Significant beyond the 0.01 level.

TABLE XVII

COMPARISONS OF READING RATE BETWEEN LEVELS OF PERFORMANCE

Instructional-Frustration I Comparisons

Test	N	Par. 1-4	S.D.	I Mean	S.D.	F^1 Mean	S.D.	t
Durrell	77			115.44	24.35	87.06	24.39	15.08*
Gates	77	105.03	24.20	91.61	22.85			8.42*
Gates	77			91.61	22.85	52.68	18.14	17.45*
SRI	77			102.26	20.27	81.18	20.71	13.08*

Frustration II Comparisons

Test	N	I Mean	S.D.	F^1 Mean	S.D.	F^2 Mean	S.D.	t
Durrell	53	118.87	23.65			69.87	23.09	16.61*
Durrell	53			90.60	26.06	69.87	23.09	8.93*
Gates	68	91.35	23.55			40.28	16.44	20.74*
Gates	68			53.54	18.54	40.28	16.44	9.34*
SRI	62	102.19	20.94			67.61	21.58	20.52*
SRI	62			84.00	21.26	67.61	21.58	9.90*

not being reported specifically but have been included in the tabled information.

The t values for the Visual Perception--Intertest comparisons are reported in Table XVIII. This table indicates that the null hypothesis can be rejected for the visual perception category at the INSTRUCTIONAL level only for the G1-4 - SRII and the GI - SRII comparisons. Table XVIII also contains the t values for the individual error-types. The G1-4 and the GI had proportionately more -++ and +-+ errors than the SRII. The DI* also had proportionately more +-+ errors than the SRII. The SRII had proportionately more --- errors than the DI. All other visual perceptual error-type differences were nonsignificant.

The null hypothesis can be rejected at the INSTRUCTIONAL - FRUSTRATIONAL 1 level except for the DI - $SRIF^1$ and the GI - DF^1 comparisons. The GI had proportionately more -++ errors than the $SRIF^1$ and the DF had proportionately more --+ errors than the SRIF. The DI and GI had proportionately more +-+ errors than the $SRIF^1$. The DF^1 and GF^1 had proportionately more +-+ errors than the SRII. The GF^1 had proportionately more ++- errors than the DI or the SRII. There were proportionately more ++- errors on the GI than on the DF^1 or the $SRIF^1$. The GF^1 had proportionately more +-- errors than the DI or the SRII. The DF^1 had proportionately more +-- errors than the SRII. The SRII had proportionately more --- errors than the GF^1 or the DF^1. (Table XVIII).

Table XVIII also contain the t values for the visual perception

---

*If the substitution of begin for began, which occurred frequently on the Durrell, had been classified as a structural error (irregular inflection) instead of a +-+ error, the difference between the Durrell and SRI would no longer be significant.

TABLE XVIII

SUMMARY OF $\bar{t}$ VALUES FOR BETWEEN-TEST COMPARISONS OF VISUAL PERCEPTION ERRORS

Test	Vis. Per. Category	Error-Types						
		-++	+-+	++-	--+	+--	-+-	---
DI - GI-4	1.02	0.24	1.25	1.58	.0005	0.24	0.37	1.27
DI - GI	1.17	0.19	0.67	1.42	0.04	1.08	0.16	0.94
DI - SRII	1.43	1.25	4.97**	0.75	0.30	0.37	0.39	2.24*
GI-4 - SRII	3.44**	2.68**	4.60**	1.48	0.48	1.46	0.08	1.14
GI - SRII	3.78**	2.96**	5.93**	1.08	0.51	1.54	0.21	1.56
DI - GF[1]	2.38**	0.49	1.11	2.19*	0.005	5.93**	1.51	1.73
DI - SRIF[1]	1.71	0.80	2.72**	0.14	0.38	0.36	1.50	0.68
GI - DF[1]	0.65	0.67	1.17	3.32**	0.19	0.98	0.85	0.79
GI - SRIF[1]	4.05**	2.07*	2.39*	3.07**	0.56	1.73	1.68	0.23
SRII - DF[1]	2.72**	2.14*	7.15**	1.98	0.60	2.31*	1.17	2.28*
SRII - GF[1]	5.68**	1.89	7.61**	2.42*	0.56	6.11**	1.88	4.79**
DF[1] - GF[1]	2.25*	0.62	1.11	5.02**	0.14	3.80**	2.14*	2.62*
DF[1] - SRIF[1]	3.16**	1.23	4.18**	0.55	0.77	2.52*	0.72	0.64
GF[1] - SRIF[1]	6.50**	0.77	5.33**	4.32**	0.65	6.63**	3.54**	3.46**
DF[2] - GF[2]	4.48**	0.74	3.43**	2.92**	0.62	5.81**	1.36	1.11
DF[2] - SRIF[2]	0.63	0.84	0.29	0.49	1.99	0.50	1.33	0.52
GF[2] - SRIF[2]	4.35**	1.39	6.22**	2.88**	0.06	7.23**	2.76**	4.40**

*Significant beyond the .05 level.

**Significant beyond the .01 level.

comparisons at the FRUSTRATION 1 level. These data indicate that the null hypothesis can be rejected at the FRUSTRATION 1 level. Among the error-types, the DF^1 and GF^1 had proportionately more +-+ errors than the $SRIF^1$. The GF^1 had proportionately more ++- and +-- errors than the DF^1 or the $SRIF^1$. There were proportionately more +-- errors on the DF^1 than on the $SRIF^1$. The DF^1 and the $SRIF^1$ had proportionately more -+- errors than the GF^1. The DF^1 and the $SRIF^1$ had proportionately more --- errors than the GF^1.

The null hypothesis can be rejected for the Visual Perception category at the FRUSTRATION 2 level (see Table XVIII) except for the DF^2 - $SRIF^2$ comparison. Among the error-types, the GF^2 had proportionately more +-+, ++-, and +-- errors than the DF^2 or the $SRIF^2$. The $SRIF^2$ had proportionately more -+- and --- errors than the GF^2

The t values for the Visual Perception--Intratest comparisons are reported in Table XIX. Only the t value for the $SRIF^1$ - $SRIF^2$ comparison is significant; therefore, the null hypothesis cannot be rejected. Examination of error-type differences indicates there were proportionately more -++ errors at the INSTRUCTIONAL level of the Gates than at the FRUSTRATION 2 level. There were proportionately more +-+ errors at the FRUSTRATION 1 and FRUSTRATION 2 levels of the Gates and the SRI than at the INSTRUCTIONAL level. The FRUSTRATION 1 level of the Gates had proportionately more +-- errors than the INSTRUCTIONAL level; the FRUSTRATION 2 levels of the Gates and the SRI had proportionately more +-- errors than the FRUSTRATION 1 levels. The FRUSTRATION 1 level of the SRI had proportionately more -+- errors than the INSTRUCTIONAL level. The INSTRUCTIONAL level of the Gates had proportionately more --- errors than the FRUSTRATION 1 or FRUSTRATION 2 level. The

TABLE XIX

SUMMARY OF t VALUES FOR LEVELS OF PERFORMANCE COMPARISONS OF VISUAL PERCEPTION ERRORS

Test	Vis. Per. Category	Error Types						
		-++	+-+	++-	--+	+--	-+-	---
DI - DF^1	0.61	0.39	0.24	0.41	0.09	1.98	0.60	0.22
DI - DF^2	0.54	0.90	0.84	0.48	0.18	3.52**	0.80	1.02
DF^1 - DF^2	1.32	1.32	1.61	0.70	0.85	0.95	0.60	1.01
GI - GF^1	1.87	1.89	2.18*	1.57	0.08	5.30**	1.41	3.30**
GI - GF^2	1.95	3.01**	3.02**	0.20	0.45	7.08**	1.06	3.44**
GF^1 - GF^2	0.37	1.96	0.91	1.49	0.04	2.80**	0.23	0.51
SRII - $SRIF^1$	0.19	1.37	2.18*	1.48	1.62	0.07	2.34*	1.77
SRII - $SRIF^2$	1.35	1.29	4.27**	1.87	0.12	2.25*	1.67	2.08*
$SRIF^1$ - $SRIF^2$	2.05*	0.19	1.60	0.08	1.87	3.31**	0.54	0.47

*Significant beyond the .05 level.

**Significant beyond the .01 level.

FRUSTRATION 2 level of the SRI had proportionately more --- errors than the INSTRUCTIONAL level.

The t values for the Visual-Auditory--Intertest comparisons are reported in Table XX. The null hypothesis can be rejected for visual-auditory errors at the INSTRUCTIONAL level except for the GI-4 - SRII comparison. Among the error-type comparisons (Table XX), the GI-4 and GI had proportionately more Consonant Blend/Digraph and Consonant-Vowel errors than the DI. The GI had proportionately more Consonant-Vowel errors than the SRII.

As can be seen from Table XX, the null hypothesis can be rejected for visual-auditory errors at the INSTRUCTIONAL - FRUSTRATION 1 level except for the GI - DF^1 comparison. Among the error-type comparisons (Table XX), the GF^1 had proportionately more Consonant and Consonant Blend/Digraph errors than the D^1. The $SRIF^1$ had proportionately more Consonant Blend/Digraph and Consonant-Vowel errors than the DI. The DF^1 had proportionately more Vowel errors than the SRII. The GF^1 had proportionately more Vowel and Consonant-Vowel errors than the DI or the SRII. The GF^1 had more Vowel Digraph/Diphthong errors than the SRII. The $SRIF^1$ had proportionately more Vowel Digraph/Diphthong errors than the GI.

The t values reported in Table XX indicate that the null hypothesis can be rejected for the visual-auditory comparisons at the FRUSTRATION 1 level except for the DF^1 - $SRIF^1$ mean-difference. The GF^1 had proportionately more Consonant and Consonant-Vowel errors than the DF^1 or the $SRIF^1$.

Table XX indicates that the null hypothesis can be rejected for visual-auditory errors at the FRUSTRATION 2 level except for the DF^2 -

TABLE XX

SUMMARY OF t VALUES FOR BETWEEN-TEST COMPARISONS OF VISUAL-AUDITORY ERRORS

Test	Vis. Aud. Category	Error-Types				
		C	CC	V	VV	CCVV
DI - GI-4	3.92**	1.25	2.45*	0.00	0.45	2.84**
DI - GI	5.57**	1.95	2.42*	1.28	0.45	4.34**
DI - SRII	2.30*	1.00	1.52	0.00	0.51	1.67
GI-4 - SRII	1.54	0.54	0.32	0.00	0.15	1.80
GI - SRII	2.86**	0.86	0.21	1.28	0.15	3.40**
DI - GF[1]	11.36**	4.33**	2.21*	2.76**	1.27	9.34**
DI - SRIF[1]	6.85**	1.28	3.32**	1.56	1.59	4.94**
GI - DF[1]	0.28	0.04	0.38	1.83	1.81	1.49
GI - SRIF[1]	2.30*	0.48	1.11	0.26	2.65*	0.77
SRII - DF[1]	3.55**	0.77	0.21	2.78**	1.86	1.86
SRII - GF[1]	9.81**	3.41	0.03	2.76**	2.42*	9.20**
DF[1] - GF[1]	7.18**	2.87**	0.10	0.20	0.73	7.92**
DF[1] - SRIF[1]	1.90	0.35	1.54	1.63	1.12	1.78
GF[1] - SRIF[1]	6.31**	2.93**	1.38	1.73	0.47	6.88**
DF[2] - GF[2]	4.83**	0.08	1.26	0.18	0.21	5.66**
DF[2] - SRIF[2]	0.52	1.53	0.06	2.37*	0.11	1.17
GF[2] - SRIF[2]	5.46**	1.59	1.74	3.28**	0.15	5.67**

*Significant beyond the .05 level.

**Significant beyond the .01 level.

$SRIF^2$ comparison. Among the significant error-type differences (Table XX), the DF^2 and GF^2 had proportionately more Vowel errors than the $SRIF^2$. The GF^2 had proportionately more Consonant-Vowel errors than the DF^2 or the $SRIF^2$.

The t values for the Visual-Auditory--Intratest comparisons are reported in Table XXI. The null hypothesis can be rejected for the levels of performance comparisons except for $GF^1 - GF^2$. Data for the error-types are contained in Table XXI, also. There were proportionately more Consonant errors on the FRUSTRATION 1 level of the Gates than on the INSTRUCTIONAL. The SRI had proportionately more Consonant errors on the FRUSTRATION 2 level than on INSTRUCTIONAL or the FRUSTRATION 1 level. The Durrell had proportionately more Consonant and Consonant Blend/Digraph errors at the FRUSTRATION 2 level than at the INSTRUCTIONAL. The FRUSTRATION 1 level of the Durrell had proportionately more Vowel errors than the INSTRUCTIONAL. There were more Vowel errors on the FRUSTRATION 2 level of the Durrell and the Gates than on the INSTRUCTIONAL. The FRUSTRATION 1 level of the Durrell, the SRI, and the Gates had proportionately more Vowel Digraph/Diphthong errors than the INSTRUCTIONAL. The FRUSTRATION 2 levels of the Durrell and the Gates had proportionately more Vowel Digraph/Diphthong errors than the INSTRUCTIONAL. The FRUSTRATION 1 and FRUSTRATION 2 levels of the Durrell, the Gates, and the SRI had proportionately more Consonant-Vowel errors than the INSTRUCTIONAL level. The FRUSTRATION 2 level of the Durrell and the SRI had proportionately more Consonant-Vowel errors than the INSTRUCTIONAL level.

The t values for the Directional Confusion--Intertest comparisons are reported in Table XXII. All of the t's are significant; thus, the

TABLE XXI

SUMMARY OF *t* VALUES FOR LEVELS OF PERFORMANCE COMPARISONS OF VISUAL-AUDITORY ERRORS

Test	Vis. Aud. Category	Error Types C	CC	V	VV	CCVV
$DI - DF_2^1$	5.00**	1.91	1.67	2.76**	2.84**	2.84**
$DI_1 - DF^2$	7.32**	2.61*	2.53*	3.52**	2.45*	5.97**
$DF^1 - DF^2$	3.92**	1.37	1.43	1.30	0.58	3.30**
$GI - GF_2^1$	8.20**	2.65**	0.30	1.90	2.44*	6.99**
$GI_1 - GF^2$	8.76**	1.05	0.09	3.64**	2.77**	7.73**
$GF^1 - GF^2$	1.34	1.12	0.10	1.33	1.22	1.71
$SRII - SRIF_2^1$	5.36**	0.39	1.66	1.09	3.02**	4.34**
$SRII_1 - SRIF^2$	5.64**	3.11**	1.34	0.18	1.78	5.64**
$SRIF^1 - SRIF^2$	2.14*	3.52**	0.06	0.79	0.74	2.65*

*Significant beyond the .05 level.

**Significant beyond the .01 level.

TABLE XXII

SUMMARY OF *t* VALUES FOR BETWEEN-TEST COMPARISONS OF ERROR-TYPES

Test	DIR. CONF.	SYL. DIV.	STRUCT.	OMISS.	ADD.	WORDS AIDED	REPET.	CORR.
DI - GI-4	4.02**	0.00	2.28*	3.33**	4.01**	1.11	0.11	0.77
DI - GI	4.64**	3.36**	2.53*	3.89**	4.24**	2.08*	0.52	1.70
DI - SRII	2.07*	1.00	5.42**	1.62	2.61*	0.60	0.78	0.79
GI-4 - SRII	2.36*	1.00	3.91**	2.77**	2.55*	0.61	0.34	0.004
GI - SRII	2.94**	3.20**	4.15**	3.46**	2.80**	1.74	1.35	1.15
DI - GF^1	8.29**	7.45**	3.22**	4.83**	4.17**	4.24**	4.44**	8.93**
DI - $SRIF^1$	5.69**	3.61**	5.64**	1.41	1.63	3.02**	0.13	4.03**
GI - DF^1	0.57	0.51	2.28*	4.27**	5.82**	0.76	0.81	3.61**
GI - $SRIF^1$	0.77	1.56	4.06**	3.07**	3.72**	0.87	0.53	3.18**
SRII - DF^1	2.85**	3.55**	1.88	1.06	3.25**	1.36	2.35*	5.00**
SRII - GF^1	6.79**	7.27**	4.14**	3.77**	2.24*	3.95**	6.00**	10.97**
DF^1 - GF^1	3.02**	4.46**	1.64	4.62**	5.03**	3.70**	4.42**	6.32**
DF^1 - $SRIF^1$	0.14	2.01*	1.42	1.03	1.89	1.31	1.58	0.87
GF^1 - $SRIF^1$	3.56**	5.85**	3.28**	3.72**	3.64**	3.42**	5.72**	6.50**
DF^2 - GF^2	0.65	3.89**	6.63**	1.55	3.53**	0.58	4.31**	5.38**
DF^2 - $SRIF^2$	0.29	1.72	3.17**	0.78	1.91	1.38	1.47	2.01
GF^2 - $SRIF^2$	3.33**	4.88**	4.55**	1.39	4.88**	3.15**	5.48**	9.80**

*Significant beyond the .05 level.

**Significant beyond the .01 level.

null hypothesis can be rejected for the INSTRUCTIONAL level. The GI-4, GI, and SRII had proportionately more directional confusion errors than the DI. The GI-4 and GI had proportionately more directional confusion errors than the SRII.

An inspection of Table XXII indicates that the null hypothesis cannot be rejected at the INSTRUCTIONAL - FRUSTRATION 1 level for the GI - DF^1 and the GI - $SRIF^1$ comparisons but can be rejected for the other pairs of means. The GF^1 had proportionately more directional confusion errors than the DI or the SRII. The $SRIF^1$ had proportionately more directional confusion errors than the DI. The DF^1 had proportionately more directional confusion errors than the SRII.

The null hypothesis cannot be rejected at the FRUSTRATION 1 level for the DF^1 - $SRIF^1$ (Table XXII) but can be rejected for the other two comparisons. The GF^1 had proportionately more directional confusion errors than the DF^1 or $SRIF^1$.

Table XXII indicates that the null -hypothesis cannot be rejected at the FRUSTRATION 2 level for the DF^2 - GF^2 and the DF^2 - $SRIF^2$ comparisons but can be rejected for the GF^2 - $SRIF^2$ comparisons. The GF^2 had proportionately more directional confusion errors than the $SRIF^2$.

The t values for the Directional Confusion--Intratest comparisons are reported in Table XXIII. As can be seen, the null hypothesis can be rejected for all levels of all three tests. The FRUSTRATION 1 level had proportionately more directional confusion errors than the INSTRUCTIONAL level, and the FRUSTRATION 2 level had proportionately more directional confusion errors than the INSTRUCTIONAL and the FRUSTRATION 1 levels for each of the three tests.

The t values for the Syllabic Division--Intertest comparisons are

TABLE XXIII

SUMMARY OF t VALUES FOR LEVELS OF PERFORMANCE COMPARISONS OF ERROR-TYPES

Test	DIR. CONF.	SYL. DIV.	STRUCT.	OMISS.	ADD.	WORDS AIDED	REPET.	CORR.
$DI - DF_2^1$	4.04**	3.72**	4.68**	0.72	0.28	1.73	1.17	4.49**
$DI_1 - DF^2$	7.14**	3.38**	6.93**	1.36	3.28*	3.38**	2.03*	5.14**
$DI^1 - DF^2$	4.71**	0.03	3.27**	1.60	3.00**	3.40**	2.00	3.43**
$GI - GF_2^1$	3.78**	5.71**	0.91	1.03	0.71	3.43**	5.54**	10.29**
$GI_1 - GF^2$	6.60**	5.38**	0.73	0.73	0.78	3.52**	7.70**	12.32**
$GF^1 - GF^2$	2.87**	0.40	1.72	1.68	1.49	2.91**	2.93**	4.61**
$SRII - SRIF_2^1$	3.96**	3.19**	0.69	0.45	1.59	2.64*	1.03	4.78**
$SRII_1 - SRIF^2$	7.38**	3.66**	0.67	1.20	2.17*	3.36**	2.86**	6.19**
$SRIF^1 - SRIF^2$	3.97**	1.16	0.37	1.83	0.16	1.48	1.95	2.72**

*Significant beyond the .05 level.

**Significant beyond the .01 level.

reported in Table XXII. The null hypothesis can be rejected at the INSTRUCTIONAL level for the DI - GI and the GI- SRII comparisons, but cannot be rejected for the other pairs of means. Among the error-types (Table XXII), the GI had proportionately more syllabic division errors than the DI or the SRII.

Table XXII indicates that the null hypothesis can be rejected at the INSTRUCTIONAL - FRUSTRATION 1 level except for the GI - DF^1 and the GI - $SRIF^1$ comparisons. The GF^1 and the $SRIF^1$ had proportionately more syllabic division errors than the DI. The DF^1 and the GF^1 had proportionately more syllabic division errors than the SRII.

As Table XXII indicates, all *t* values are significant for the FRUSTRATION 1 level comparisons; therefore, the null hypothesis can be rejected. The GF^1 had proportionately more syllabic division errors than the DF^1 or $SRIF^1$. The DF^1 had proportionately more syllabic division errors than the $SRIF^1$.

The null hypothesis for Syllabic Division comparisons at the FRUSTRATION 2 level can be rejected except for the DF^2 - $SRIF^2$ mean differences (Table XXII). The GF^2 had proportionately more syllabic division errors than the DF^2 or $SRIF^2$, among the error-types.

An inspection of Table XXIII indicates that the null hypothesis can be rejected for Syllabic Division intratest comparisons at the INSTRUCTIONAL-- FRUSTRATION 1 and INSTRUCTIONAL - FRUSTRATION 2 levels, but cannot be rejected for the FRUSTRATION 1 - FRUSTRATION 2 level comparisons of the three tests. The FRUSTRATION 1 and FRUSTRATION 2 levels had proportionately more syllabic division errors than the INSTRUCTIONAL level for each of the tests.

The *t* values for the Structural errors--Intertest comparisons are

reported in Table XXII. All t values for the INSTRUCTIONAL level comparisons are significant; the null hypothesis can be rejected. Among the error-types, the G1-4 and the GI had proportionately more structural errors than the DI. The SRII had proportionately more structural errors than the DI, the G1-4, or the GI.

As can be seen from Table XXII, the null hypothesis can be rejected at the INSTRUCTIONAL - FRUSTRATION 1 level except for the SRII - DF^1 comparison. Among the error-type differences, the $SRIF^1$ had proportionately more structural errors than the GI or the DI. The GF^1 had proportionately more structural errors than the DI. The DF^1 had more structural errors than the GI. The SRII had proportionately more structural errors thanthe GF^1.

Table XXII indicates that the null - hypothesis can be rejected at the FRUSTRATION 1-level only for the GF^1 - $SRIF^1$ comparison. The $SRIF^1$ had proportionately more structural errors than the GF^1.

The null hypothesis concerning structural errors can be rejected at the FRUSTRATION 2 level. (See Table XXII)

The t values for Structural errors--Intratest comparisons are reported in Table XXIII. The null hypothesis can be rejected for the Durrell levels of performance but cannot be rejected for the SRI and Gates tests. There were proportionately more structural errors at the FRUSTRATION 1 and FRUSTRATION 2 levels of the Durrell than the INSTRUCTIONAL; the FRUSTRATION 2 level had proportionately more structural errors than the FRUSTRATION 1 level.

The t values for the Corrections--Intertest comparisons are reported in Table XXII. All of the t's for the comparisons at INSTRUCTIONAL level are nonsignificant; therefore, the null hypothesis

cannot be rejected.

The null hypothesis can be rejected at the INSTRUCTIONAL - FRUSTRATION 1 level, as can be seen from Table XXII. The DI and SRII had proportionately more corrections than the GF^1; the DI and GI had proportionately more corrections than the $SRIF^1$; and the GI and SRII had proportionately more corrections than the DF^1.

As Table XXII indicates, the null hypothesis can be rejected at the FRUSTRATION 1 level except for the DF^1 - $SRIF^1$ comparison. The DF^1 and $SRIF^1$ had proportionately more corrections than the GF^1.

The null hypothesis can be rejected at the FRUSTRATION 2 level for the corrections comparisons except for DF^2 - $SRIF^2$. There were proportionately more corrections on the DF^2 and the $SRIF^2$ than on the GF^2.

The *t* values for differences in mean proportion of corrections between levels of performance are all significant as may be seen from Table XXIII; therefore, the null hypothesis can be rejected. The mean proportion of corrections at the INSTRUCTIONAL level of each test exceeded that at the FRUSTRATION 1 or FRUSTRATION 2 level. The FRUSTRATION 1 level had proportionately more corrections than the FRUSTRATION 2 level.

The *t* values for the Repetitions--Intertest comparisons are reported in Table XXII. All *t* values are nonsignificant for the INSTRUCTIONAL-level comparisons; thus, the null hypothesis cannot be rejected.

The null hypothesis can be rejected for the DI - GF^1, the SRII - DF^1, and SRII - GF^1 pairs of means, but cannot be rejected for the other INSTRUCTIONAL - FRUSTRATION 1 level comparisons. The DI and SRII had proportionately more repetitions than the GF^1; the SRII had

proportionately more repetitions than the DF^1.

The null hypothesis can be rejected at the FRUSTRATION 1 level except for the DF^1 - $SRIF^1$ comparison. The DF^1 and $SRIF^1$ had proportionately more repetitions than the GF^1.

The null hypothesis can be rejected at the FRUSTRATION 2 level except for the DF^2 - $SRIF^2$ comparison. The DF^2 and the $SRIF^2$ had proportionately more repetitions than the GF^2.

The t values for the Repetitions--Intratest comparisons are reported in Table XXIII. The null hypothesis can be rejected for the Gates levels of performance, the SRII - $SRIF^2$, and the DI - DF^2. There were proportionately more repetitions on the INSTRUCTIONAL level of the Durrell and SRI than on the FRUSTRATION 2 level. For the Gates test, there were proportionately more repetitions at the INSTRUCTIONAL level than at the FRUSTRATION 1 or FRUSTRATION 2 level. The FRUSTRATION 1 level had proportionately more repetitions than the FRUSTRATION 2 level of the Gates.

The t values for Omissions-Additions-Words Aided--Intertest comparisons are reported in Table XXIV. The t values are all significant for the comparisons at the INSTRUCTIONAL level; therefore, the null hypothesis can be rejected. The DI and the SRII had proportionately more omissions and additions than the GI-4 or GI. The DI had proportionately more additions than the SRII. The GI had proportionately more words aided than the DI.

As can be seen from Table XXIV, the t values are not significant for the DI - $SRIF^1$ and the SRII - GF^1 pairs of means. The null hypothesis can be rejected for the other comparisons at the INSTRUCTIONAL -

FRUSTRATION 1 level. The DI and SRII had proportionately more omissions and additions than the GF^1. The DF^1 and the $SRIF^1$ had proportionately more omissions and additions than the GI. The DF^1 had proportionately more additions than the SRII. The GF^1 and $SRIF^1$ had proportionately more words aided than the DI. The GF^1 had proportionately more words aided than the SRII.

The null hypothesis at the FRUSTRATION 1 level for Omissions, Additions, and Words Aided can be rejected except for the DF^1 - $SRIF^1$ comparison (Table XXIV). The DF^1 and the $SRIF^1$ had proportionately more omissions and additions than the GF^1. The GF^1 had proportionately more words aided than the DF^1 and the $SRIF^1$.

The null hypothesis at the FRUSTRATION 2 level for Omissions, Additions, and Words Aided cannot be rejected. All t values are non-significant as can be seen from Table XXIV.

The t values for Omissions, Additions, Words Aided--Intratest comparisons are reported in Table XXV. The t's are nonsignificant for the Durrell levels of performance and the $SRIF^1$ - $SRIF^2$ comparison. The null hypothesis can be rejected for the other comparisons. The Durrell had proportionately more additions at the INSTRUCTIONAL and FRUSTRATION 1 levels than at the FRUSTRATION 2 level. The SRI had proportionately more additions at the FRUSTRATION 2 level than at the INSTRUCTIONAL. The Durrell had proportionately more words aided at the FRUSTRATION 2 level than at the INSTRUCTIONAL or FRUSTRATION 1. The SRI and Gates had proportionately more words aided at the FRUSTRATION 1 and FRUSTRATION 2 levels than at the INSTRUCTIONAL level. The Gates also had proportionately more words aided at the FRUSTRATION 2 level than at the FRUSTRATION 1 (Table XXIII).

TABLE XXIV

SUMMARY OF t VALUES FOR BETWEEN-TEST COMPARISONS OF ERROR CATEGORIES

Test	VIS. PER.	VIS. AUD.	OMISS. ADD. WDS. AIDED	REPET.	CORR.	SYL. DIV.	STRUCT.	DIR. CONF.
DI - GI-4	1.02	3.92**	5.57**	0.11	0.77	0.00	2.28*	4.02**
DI - GI	1.17	5.57**	5.47**	0.52	1.70	3.36**	2.53*	4.64**
DI - SRII	1.43	2.30*	3.07**	0.78	0.79	1.00	5.42**	2.07*
GI-4 - SRII	3.44**	1.54	4.03**	0.34	0.004	1.00	3.91**	2.36*
GI - SRII	3.78**	2.86**	3.28**	1.35	1.15	3.20**	4.15**	2.94**
DI - GF^1	2.38**	11.36**	2.69**	4.44**	8.93**	7.45**	3.22**	8.29**
DI - $SRIF^1$	1.71	6.85**	1.05	0.13	4.03**	3.61**	5.64**	5.69**
GI - DF^1	0.65	0.28	6.14**	0.81	3.61**	0.51	2.28*	0.57
GI - $SRIF^1$	4.05**	2.30*	5.51**	0.53	3.18**	1.56	4.06**	0.77
SRII - DF^1	2.72**	3.55**	3.55**	2.35*	5.00**	3.55**	1.88	2.85**
SRII - GF^1	5.68**	9.81**	0.08	6.00**	10.97**	7.27**	4.14**	6.79**
DF^1 - GF^1	2.25*	7.18**	2.82**	4.42**	6.32**	4.46**	1.64	3.02**
DF^1 - $SRIF^1$	3.16**	1.90	1.47	1.58	0.87	2.01*	1.42	0.14
GF^1 - $SRIF^1$	6.50**	6.31**	2.08*	5.72**	6.50**	5.85**	3.28**	3.56**
DF^2 - GF^2	4.48**	4.83**	1.26	4.31**	5.38**	3.89**	6.63**	0.65
DF^2 - $SRIF^2$	0.63	0.52	0.05	1.47	2.01	1.72	3.17**	0.29
GF^2 - $SRIF^2$	4.35**	5.46**	0.41	5.48**	9.80**	4.88**	4.55**	3.33**

*Significant beyond the .05 level.

**Significant beyond the .01 level.

TABLE XXV

SUMMARY OF t VALUES FOR LEVELS OF PERFORMANCE COMPARISONS OF ERROR CATEGORIES

Test	VIS. PER.	VIS. AUD.	OMISS. ADD. WDS. AIDED	REPET.	CORR.	SYL. DIV.	STRUC.	DIR. CONF.
$DI - DF_2^1$	0.61	5.00**	0.25	1.17	4.49**	3.72**	4.68**	4.04**
$DI_1 - DF^2$	0.54	7.32**	0.91	2.03*	5.14**	3.38**	6.93**	7.14**
$DF^1 - DF^2$	1.32	3.92**	1.31	2.00	3.43**	0.03	3.27**	4.71**
$GI - GF_2^1$	1.87	8.20**	2.73**	5.54**	10.29**	5.71**	0.91	3.78**
$GI_1 - GF^2$	1.95	8.76**	3.68**	7.70**	12.32**	5.38**	0.73	6.60**
$GF^1 - GF^2$	0.37	1.34	2.43*	2.93**	4.61**	0.40	1.72	2.87**
$SRII - SRIF_2^1$	0.19	5.36**	2.58*	1.03	4.78**	3.19**	0.69	3.96**
$SRII_1 - SRIF^2$	1.35	5.64**	2.66**	2.86**	6.19**	3.66**	0.67	7.38**
$SRIF^1 - SRIF^2$	2.05*	2.14*	0.67	1.95	2.72**	1.16	0.37	3.97**

*Significant beyond the .05 level.

**Significant beyond the .01 level.

Tables XXIV and XXV contain summaries of the error category comparisons which have been discussed individually.

Hypothesis 4: There is no significant relation among the rankings of error-types or categories by the three oral tests at INSTRUCTIONAL level. To test this hypothesis, two Kendall Coefficients of Concordance: W were computed. The results are reported in Table XXVI. As can be seen, there is a significant relation among the rankings of the three tests; and the null hypothesis can be rejected at the .001 level for the error-type rankings and .01 level for the error categories.

TABLE XXVI

SUMMARY OF THE COEFFICIENTS OF CONCORDANCE FOR THE RANKED ORDER OF TWENTY ERROR TYPES AND EIGHT ERROR CATEGORIES

		Test for Significance of W		
Tests	W	χ^2	df	Sign.
DI - GI - SRII	.919	52.4	19	.001
DI - GI - SRII	.978	20.5	7	.01

Hypothesis 5: There is no significant relation among the rankings of error-types or categories by oral reading passages at the INSTRUCTIONAL, FRUSTRATION 1, and FRUSTRATION 2 levels of performance.

To test this hypothesis, ten Kendall Coefficients of Concordance were computed. The results are reported in Table XXVII. The null hypothesis can be rejected for both error-type and error category rankings.

Summary

This chapter has presented the statistical results from the treatment of the data. Analysis of variance was used to determine if there were significant differences in the mean grade-equivalent scores on nine oral reading measures. After the F ratio had been found to be significant, multiple comparisons of the means were made by use of the Duncan Multiple-Range Test. Five pairs of grade-equivalent scores did not differ significantly.

Multiple *t*-tests were computed to test the null hypothesis of no significant differences in oral reading rates between tests or between levels of performance on a single test. The hypothesis was rejected for levels of performance comparisons and for fifteen of the eighteen comparisons between reading tests.

Multiple *t*-tests were used, also, to determine the significance of differences in error-proportions between tests and between levels of performance on a single test. The hypothesis was examined separately for eight error categories and sub-categories. Significant differences in the relative proportion of specific error-types were also identified. Coefficients of Concordance were computed to determine the agreement among the oral reading tests in the rankings of error categories and error-types. W, the coefficient of concordance, was significant at the .01 or .001 level of confidence for all rankings of error-types and categories except the ranking of the error categories by three levels of the Gates test which was significant at the .05 level of confidence.

TABLE XXVII

SUMMARY OF THE COEFFICIENTS OF CONCORDANCE FOR THE RANKED ORDER OF TWENTY ERROR TYPES AND EIGHT ERROR CATEGORIES

		Test for Significance of W		
Tests	W	χ^2	df	Sign.
DI - DF^1 - DF^2	.902	51.4	19	.001
GI - GF^1 - GF^2	.863	49.2	19	.001
SRII - $SRIF^1$ - $SRIF^2$	.938	53.5	19	.001
DI - DF^1 - DF^2 SRII - $SRIF^1$ - $SRIF^2$	.904	103.1	19	.001
DI - DF^1 - DF^2 GI - GF^1 - GF^2 SRII - $SRIF^1$ - $SRIF^2$	.806	137.8	19	.001
DI - DF^1 - DF^2	.947	19.9	7	.01
GI - GF^1 - GF^2	.677	14.2	7	.05
SRII - $SRIF^1$ - $SRIF^2$	.968	20.3	7	.01
DI - DF^1 - DF^2 SRII - $SRIF^1$ - $SRIF^2$	.943	39.6	7	.001
DI - DF^1 - DF^2 GI - GF^1 - GF^2 SRII - $SRIF^1$ - $SRIF^2$	.735	46.3	7	.001

CHAPTER V

SUMMARY AND CONCLUSIONS

General Summary of the Investigation

This study was concerned with the reading grade equivalents, oral reading rates, and prevailing error patterns of fourth-grade disabled readers on standardized oral reading tests.

The sample consisted of all the fourth-grade pupils in one county in northern Oklahoma who met the criteria set for the study: a full-scale IQ of 90 or above on the Wechsler Intelligence Scale for Children; a maximum instructional level on the Standard Reading Inventory, Form B, at or between 2^1 and 3^2; and no discernible physical handicaps which would interfere with their reading of the test materials or the subsequent analysis of the reading errors. The final sample consisted of seventy-seven children.

The oral reading at sight of the oral paragraphs and/or stories from the Durrell Analysis of Reading Difficulty, the Gates-McKillop Reading Diagnostic Tests, Form 1, and the Standard Reading Inventory, Form A, was timed and tape-recorded. Errors made at three levels of reading performance--INSTRUCTIONAL, FRUSTRATION 1, and FRUSTRATION 2--on each of the oral reading tests were analyzed using the B-S-R Error Analysis. Comparisons were made of the error patterns and the mean oral reading rates between pairs of tests and between levels of performance. Grade scores were obtained from the Durrell Word

Recognition and Word Analysis Test, the word lists from the SRI, the reading section of the Wide Range Achievement Test, and the Gates-McKillop, Durrell, and SRI oral paragraphs/stories. Additional measures of oral reading level were obtained from second scorings of the Durrell paragraphs and the SRI stories.

Analysis of variance was used to determine if there were significant differences in oral reading grade-equivalents among the nine measures. The specific means differing significantly from each other were identified by use of the Duncan Multiple-Range Test. Multiple *t*-tests were computed to determine the significance of differences in error proportions and in oral reading rates between tests and between levels of performance. The Kendall Coefficient of Concordance: W was employed to ascertain the agreement in error ranks among the three tests at the three levels of reading performance.

Conclusions

The results of this study indicate that there are significant differences in grade-equivalent scores among the nine oral reading measures. All words-in-isolation mean grade-scores were significantly higher than the mean grade-levels based upon oral paragraphs. However, the difference in mean scores between the SRI oral stories and the Gates-McKillop or Durrell paragraphs, or between the SRI word lists and the word lists on the Durrell was greater than the difference between the Durrell paragraphs and the WRAT or SRI word lists. The grade equivalents of individual subjects varied considerably from test to test. The SRI total-errors-scoring of the oral stories tended to be lowest and the Durrell Word Recognition Test to be highest.

There was a significant (although probably not an important) difference between the total-errors scoring of the SRI and a word-recognition-errors-only scoring. The two scorings resulted in the same grade level for forty-three pupils. Of the thirty-four differences, the word-recognition scoring produced the higher grade-equivalent level twenty-nine times. For twelve pupils, the results differed by more than a half-grade level.

The SRI was significantly lower than the other reading measures. Approximately three-fourths of the subjects obtained a higher grade level on the Gates-McKillop, the third-lowest mean, than on the word-recognition scoring of the SRI; five pupils obtained the same score on both tests. The two tests differed by more than five-tenths grade for thirty-five pupils.

The two scorings of the Durrell did not differ significantly, perhaps because the possible-grade-score ratings tended to be similar on the paragraphs read successfully by this sample. Fifty-four pupils obtained the same grade-equivalent by either scoring; in twenty of the twenty-three different grade-equivalents, the last-successful-paragraph scoring was the higher. The two scores differed by more than five-tenths grade in only one case.

Unlike Attea's (1966) findings, neither scoring of the Durrell differed significantly from the oral paragraph grade-score on the Gates-McKillop. Attea found the Gates-McKillop to be "consistently higher." In this study, the Durrell was higher forty-three times; the Gates-McKillop twenty-seven. In nineteen of the thirty-one grade scores which differed by more than a half-year, the Durrell was the higher. These results suggest that disabled readers tend to make

relatively higher scores on the Durrell as compared to the Gates-McKillop than do slightly-younger randomly-selected third-grade pupils of similar reading level. Presumably the disabled readers were less penalized by the time-scoring of the Durrell than by the Gates' scoring based on oral reading accuracy.

In the present study, the grade score on the Durrell oral paragraphs was 1.11 grades lower than the score on the Durrell Word Recognition Test, and .98 grade lower than the Word Analysis score. These differences agree with Herlin's (1963) findings with randomly-selected third and fourth graders and were slightly greater than those reported by Attea (1966).

The mean grade-equivalent on the WRAT did not differ significantly from that on the SRI word lists; for twenty-one pupils, however, there was more than a half-year's difference. Likewise, the Durrell Word Recognition and Word Analysis Test mean scores did not differ significantly; the scores of five pupils differed by more than five-tenths grade. The mean grade scores on the WRAT and the SRI did differ significantly from those on the Durrell word tests.

Significant differences in reading rates were found among the tests. The Durrell (DI) was significantly higher than the other measures at INSTRUCTIONAL level. The reading rate on the SRI INSTRUCTIONAL (SRII) was not significantly different from the rate on the first four paragraphs of the Gates (GI-4), but was significantly higher than that on the Gates INSTRUCTIONAL (GI). The GI reading rate was similar to that of the Durrell FRUSTRATION^{1} (DF^{1}); the Durrell FRUSTRATION 2 (DF^{2}) and the SRI FRUSTRATION 2 (SRIF^{2}) did not differ significantly.

The difference in reading rates between the tests of INSTRUCTIONAL level may have been caused in part by differences in the difficulty of the material read. The average readability level of the DI, as determined by the Spache (1968) formula was 2.19; for the SRII, it was 2.33; for the paragraphs of the GI-4 on which there were two or more errors, it was 2.87; and on the GI, the readability level was 3.6. Table XXVIII indicates the mean reading rate of the last paragraph falling within the INSTRUCTIONAL level on the Durrell and the SRI and a breakdown of the readability level of the last successful paragraph for individual pupils.

TABLE XXVIII

NUMBER OF PUPILS READING AT EACH READABILITY LEVEL ON LAST SUCCESSFUL PARAGRAPH

	Rate of Reading	Readability Levels						
		1.3-1.5	1.6-1.7	2.1-2.2	2.7	3.3-3.4	3.5	3.7
DUR	112.26	1	6	35	22		13	
SRI	97.04	1	5	20	16	22		13

That the difference in rate is not entirely the result of differences in readability levels, as determined by a formula, is indicated by Table XXIX. Although the readability levels were similar, the reading rate on the SRI second-grade selection was twenty words a minute slower than the rate on the Durrell paragraph. Other variables, such as the format, sentence structure, or length of passage, apparently affected the reading rate, also.

TABLE XXIX

READING RATES ON PARAGRAPHS
OF SIMILAR READABILITY

Test	Paragraph Number	Readability Level	N	Rate of Reading
Gates	2	2.1	73	107.71
Durrell	3	2.1	73	123.99
SRI	2[1]	2.2	77	103.14
Durrell	5	3.5	63	90.46
SRI	3[2]	3.7	63	77.59
Gates	4	3.4	63	90.02

Table XXX reports the error-types for which there were significant changes in error-proportion from the INSTRUCTIONAL level to FRUSTRATION 1 and/or FRUSTRATION 2 level(s). The results agree closely with those reported by previous investigators (Schale, 1964; Killgallon, 1942; Christenson, 1966; Schummers, 1956) which were summarized in Table IX on page 58. An increase in the number of non-word responses as the material became more difficult, which is reported in Table XXXI, also supports the findings of other investigators. An increase in ++- errors and in +-+ errors on the Gates and SRI, and a decrease in -++ and --- errors on the Gates were observed as the material increased in difficulty, although the relative proportion of Visual Perception errors did not change from one level of performance to another. These results suggest that error-types from one or more levels of performance should not be combined in error analyses.

Changes in the ranks of error-types among the levels of performance

TABLE XXX

ERRORS WHICH CHANGE IN RELATIVE FREQUENCY AS DIFFICULTY LEVEL OF MATERIAL INCREASES

Errors Which Increase	Errors Which Decrease	Errors Which Do Not Change
Visual Auditory	Repetitions (Gates, $DI\text{-}DF^2$ and $SRII\text{-}SRIF^2$)	Total Visual Perception
VV	Corrections	Omissions (change in error rank, however)
V ($DF^1 - DF^2$; $GF^1 - GF^2$)	Additions (Gates only)	
C	-++ errors (Gates only)	
CCVV	--- errors (Gates only)	
Syllabic Division		
Directional Confusion		
Structure errors (Durrell only)		
Words Aided		
+-+ errors (Gates and SRI		
++- errors		

may be seen in Table XXXII. This table indicates that although there were proportionately fewer corrections at the FRUSTRATION 1 level than at the INSTRUCTIONAL level and at the FRUSTRATION 2 level than at the FRUSTRATION 1 level for all tests, corrections remained the most frequent kind of error-type on the Durrell and the SRI. Conversely, a change in error rank for the omissions error-type is indicated by Table XXXII, but the ratio of omission-occurrence to total-error occurrence did not change significantly from one level of performance to another. The most dramatic changes in error ranks occurred on the Gates-McKillop which has considerably more unknown words and considerably fewer semantic and syntactic cues at the FRUSTRATION 1 and 2 levels than do the other tests.

TABLE XXXI

PER CENT OF TOTAL SUBSTITUTIONS
WHICH WERE NON-WORDS

Test	Per Cent	Test	Per Cent	Test	Per Cent
DI	1.7	GI-4	3.0	SRII	3.4
DF[1]	13.7	GI	12.0	SRIF[1]	13.5
DF[2]	28.4	GF[1]	44.2	SRIF[2]	15.1
		GF[2]	58.4		

As may be seen by Table XXXIII there were virtually no differences in the rankings of error categories by the three tests at INSTRUCTIONAL level. However, there were differences in the relative proportions

Table XXXII

RANKED ORDER OF TWENTY ERROR TYPES

	-++	+-+	++-	+--	-+-	--+	---	C	CC	V	VV	CCVV	OMISS	ADD.	WORDS AIDED	REPET.	CORR.	DIR. CONF.	SYL. DIV.	STRUCT
SRII	11	8	6	9	13	12	3	18	15	20	17	14	5	7	16	2	1	10	19	4
SRIF[1]	11	6	9	12	14	18	3	20	17	19	15	10	7	5	13	2	1	8	16	4
SRIF[2]	13	6	11	9	14	15	2	17	18	20	19	8	10	7	12	3	1	5	16	4
DI	7	4	8	10	12	11	3	18	18	18	18	14.5	6	5	14.5	2	1	13	18	9
DF[1]	10	4	11	9	15	14	3	20	19	18	17	13	7	5	16	2	1	8	12	6
DF[2]	15	6	12	7	17	14	3	19	18	16	20	10	8	9	11	4	1	5	13	2
GI	7	4	5	9	16	14	3	18	17	20	19	11	10	13	15	2	1	8	12	6
GF[1]	12	1	8	5	20	15	4	16	19	18	17	3	14	13	10	6	2	7	9	11
GF[2]	14	1	9	3	20	15	5	18	19	16	17	2	12	13	6	10	7	4	8	11

and in the error rank of specific error-types. The more important differences among the three tests are reported in Table XXXIV. As may be seen, the SRI tended to have more structural errors and fewer +-+ errors than the other two tests at both INSTRUCTIONAL and FRUSTRATION 1 levels. The Durrell had relatively more additions and omissions; whereas, the Gates had proportionately more Visual-Auditory, Syllabic Division, and Directional Confusion errors. The Gates also tended to have relatively more Visual Perception errors than the other two tests. In general, the error patterns on the Durrell and the SRI were similar; there were fewer differences at the FRUSTRATION 1 level than at the INSTRUCTIONAL level, and fewer still at the FRUSTRATION 2 level. The differences observed on the Gates may reflect differences in difficulty level. The percentages of oral reading accuracy at the various levels of performance on the two tests is reported in Table XXXV.

TABLE XXXIII

RANKED ORDER OF ERROR CATEGORIES
AT INSTRUCTIONAL LEVEL

	Vis. Per.	Vis. Aud.	Omiss. and Wds. Aided	Repet.	Corr.	Syl. Div.	Struct.	Dir. Conf.
DI	2	7	4	3	1	8	5	6
GI	1	7	4	3	2	8	5	6
SRII	2	7	5	3	1	8	4	6

TABLE XXXIV

ERROR-TYPE DIFFERENCES AMONG THE ORAL READING TESTS

DI	GI-4	GI	SRII
More Addition errors	More Directional Confusion errors	More Directional Confusion errors	More Structural errors
More Omission errors than GI-4 and GI	More CCVV than DI	More CCVV errors	More Omission errors than GI-4 and GI
	More V than DI	More Syllabic Division errors	More --- errors than Durrell
		More Words Aided than DI	

More +-+ and -++ errors than SRII

Fewer Omissions and Additions

DF[1]	GF[1]	SRIF[1]
More Syllabic Division Errors than SRIF[1]	More CCVV errors	More Structural errors than GF[1]
More +-- and +-+ errors than SRIF[1]	More ending errors	Fewer +-+ errors
	More Words Aided	
	More Directional Confusion errors	
	More Syllabic Division errors	
	More +-+ errors than SRIF[1]	
	Fewer --- errors	
	Fewer Additions and Omissions	
	Fewer Repetitions	
	Fewer Corrections	

A comparison of reading rates and error-patterns between the GI-4 and GI supports Gates' stipulation that errors should be classified only on the first four paragraphs of the Gates-McKillop. For these subjects it appeared that the reading of paragraph five and above was essentially a word-pronouncing task. As compared with the Gates-McKillop norms (1962), the subjects in the study made a greater number of two-word repetitions and omissions on the first four paragraphs of the test. Among the words mispronounced (Gates' definition), there were fewer wrong endings and more wrong-in-several part errors than the Gates norms indicate.

TABLE XXXV

ORAL READING ACCURACY
PER CENTS

Test	Per Cent	Test	Per Cent	Test	Per Cent
DI	92.3	GI	87.3	SRII	91.0
DF[1]	84.2	GF[1]	70.1	SRIF[1]	84.8
DF[2]	77.8	GF[2]	55.8	SRIF[2]	80.7

The results of this study would not justify the use of a word-in-isolation test as a measure of general reading ability for disabled readers.

Since most errors at the INSTRUCTIONAL level are visual perceptual or behavioral-type errors, to adequately sample the pupil's ability to

apply phonic generalizations and sound-symbol relationships, it may be necessary to use FRUSTRATION level material.

There were more efforts to sound-out words on the Gates, perhaps because more words were not in the pupils' recognition vocabulary and fewer other cues were available than on the SRI and the Durrell oral sections. To evaluate phonic and syllabication skills, it may be desirable to use material with relatively few semantic and syntactic cues, such as the upper paragraphs of the Gates; whereas, the more meaningful material of the Durrell or the SRI may be preferable for evaluating the use of contextual clues.

Although pupils may need to read in somewhat difficult material in order to assess their visual-auditory skills, many more directional confusion and beginnings-only-correct types of errors were observed when the material became very difficult. It appears that the pupils reverted to an earlier level of skill development when faced with a decoding task that was too difficult. To obtain useful information about the pupils' decoding skills, then, the material should be moderately difficult but not so formidable that the readers fail to apply the skills they possess.

Recommendations

1. Since differences in error patterns which were observed among the three tests may reflect differences in the relative difficulty of material, it is suggested that error patterns on the tests be compared with error patterns from instructional materials of various difficulty levels.

2. A study should be made of error patterns between paragraphs of

comparable difficulty within the various reading tests.

3. The relationship between error patterns on words in context and patterns of errors on words in isolation should be explored.

4. Investigation should be made of the relative use of syntactic and semantic cues on the various reading tests.

5. It is suggested that the relation between the Visual-Auditory errors at FRUSTRATION 1 level and the errors made on isolated word lists be investigated.

6. A study should be made of the relation between the decrease in rate of reading and the increase in relative difficulty of material for a specific reader.

7. It is suggested that a study be made of the similarity between the patterns of errors of individual pupils in different materials, using a detailed analysis of errors.

BIBLIOGRAPHY

Allen, Paul D. "A Psycholinguistic Analysis of the Substitution Miscues of Selected Oral Readers in Grades Two, Four, and Six, and the Relationship of These Miscues to the Reading Process: A Descriptive Study." (unpub. Ed.D. dissertation, Wayne State University, 1969.)

Attea, Mary. "A Comparison of Three Diagnostic Reading Tests." (unpub. Ed.D. dissertation, State University of New York at Buffalo, 1966.)

Baker, Norma. "Confusion in Word Recognition." Elementary School Journal, XLIV (June, 1945), 575-577.

Barbe, Walter B., Thelma Williams and Virginia Ganaway. "Types of Difficulties in Reading Encountered by Eighty Children Receiving Instruction at a Reading Clinic." Journal of Educational Research, LI (February, 1958), 437-443.

Beldin, H. O. "Informal Reading Testing: Historical Review and Review of the Research." Reading Difficulties: Diagnosis, Correction, and Remediation. W. K. Durr, ed. Newark, Delaware: International Reading Association, 1970.

Bennett, Annette. "An Analysis of Errors in Word Recognition Made by Retarded Readers." Journal of Educational Psychology, XXXIII (January, 1942), 25-38.

Betts, Emmett A. Foundations of Reading Instruction. New York: American Book Co., 1946.

Bond, Guy L., Bruce Balow, and Cyril Hoyt. Silent Reading Diagnostic Tests. Chicago, Ill.: Lyons and Carnahan, 1970.

Bond, Guy L., Theodore Clymer, and Cyril J. Hoyt. The Developmental Reading Tests - Form D-A. Chicago, Ill.: Lyons and Carnahan, 1955.

Botel, Morton. "A Comparative Study of the Validity of the Botal Reading Inventory and Selected Standardized Tests," in J. Allen Figurel (Ed.), Reading and Realism, 1968 Proceedings, XIII, Part 1. Newark, Delaware: International Reading Association, 1969.

Botel, Morton, John Bradley, and Michael Kashuba. "The Validity of Informal Reading Testing," in William K. Durr (Ed.), Reading Difficulties: Diagnosis, Correction, and Remediation. Newark, Delaware: International Reading Association, 1970.

Brown, Sandra Rose. "A Comparison of Five Widely Used Standardized Reading Tests and an Informal Reading Inventory for a Selected Group of Elementary School Children." (unpub. Ed.D. dissertation, University of Georgia, 1963.) Dissertation Abstracts, 996.

Burke, Carolyn L. "A Psycholinguistic Description of Grammatical Restructuring in the Oral Reading of a Selected Group of Middle School Children." (unpub. Ed.D. dissertation, Wayne State University, 1969.)

Burnstein, A. G. Sixth Mental Measurement Year Book. Highland Park, N. J.: Gryphon Press, 1965.

Byrant, N. Dale. Reading Tests and Reviews. Highland Park, N. J.: Gryphon Press, 1968.

Callaway, Byron. "Factors Related to Reading of Children Referred to a University Reading Clinic," in William K. Durr (Ed.), Reading Difficulties: Diagnosis, Correction, and Remediation. Newark, Delaware: International Reading Association, 1970.

Cavoures, Dorothy G. "Phoneme Identification in Primary Reading and Spelling." (unpub. Ed.D. dissertation, Boston University, 1964.)

Christenson, A. Adolph. "A Diagnostic Study of Oral Reading Errors of Intermediate Grade Children at Their Independent, Instructional, and Frustration Reading Levels." (unpub. Ed.D. dissertation, Colorado State College, 1966.)

Christenson, Adolph. "Oral Reading Errors of Intermediate Grade Children at Their Independent, Instructional, and Frustration Reading Levels," in J. Allen Figurel (Ed.), Reading and Realism, 1968 Proceedings, XIII, Part 1. Newark, Delaware: International Reading Association, 1969.

Clay, Marie M. "A Syntactic Analysis of Reading Errors." Journal of Verbal Learning and Verbal Behavior, VII (1968), 434-438.

Cooper, J. Louis. "The Effect of Adjustment of Basal Reading Materials on Reading Achievement." (unpub. Ed.D. dissertation, Boston University, 1952.)

Courtney, Brother Leonard, F.S.C. "The Relationship Between the Oral and Silent Reading of College Students." (unpub. Ph.D. dissertation, University of Chicago, 1964.)

Dale, Edgar and Jeanne Chall. "A Formula for Predicting Readability." Educational Research Bulletin, XXVII (January 21 and February 18, 1948) 16-20, 28, 37-54.

Daniel, John Emerson. "The Effectiveness of Various Procedures in Reading Level Placement." Elementary English, XXXIX (October, 1962), 590-600.

Daniels, J. C. and H. Diack. Progress in Reading. Nottingham Institute of University of Education, 1956.

Davis, Georgia. "Procedures Effective in Improving Pupils of Poor Reading Ability in Regular Reading Classes." Elementary School Journal, XXXI (January, 1931), 336-348.

Davis, William Quinby. "A Study of Test Score Comparability Among Five Widely Used Reading Survey Tests." (unpub. Ph.D. dissertation, Southern Illinois University, 1968.) Dissertation Abstracts.

Daw, Seward Emerson. "The Persistence of Errors in Oral Reading in Grades Four and Five." Journal of Educational Research, XXXII (October, 1938), 81-90.

Della, Piana, Gabriel M. and Wayne R. Herlin. "Are Normative Oral Reading Error Profiles Necessary?," in J. Allen Figurel (Ed.), Improvement of Reading Through Classroom Practice. Newark, Delaware: International Reading Association, 1964.

Duffy, Gertrude, and Donald Durrell. "Third Grade Difficulties in Oral Reading." Education, CVI (September, 1935), 37-40.

Dunkeld, Colin G. M. "The Validity of the Informal Reading Inventory for the Designation of Instructional Reading Levels: A Study of the Relationships Between Children's Gains in Reading Achievement and the Difficulty of Instructional Materials." (unpub. Ph.D. dissertation, University of Illinois, 1970.)

Dunn, Lloyd M. Peabody Picture Vocabulary Test. Circle Pines, Minnesota: American Guidance Service, 1959.

Durrell, Donald. "Tests and Corrective Procedures for Reading Disabilities." The Elementary English Review, XII (4) (April, 1935), 91-95.

Durrell, D. D. Durrell Analysis of Reading Difficulty. New York: Harcourt, Brace, and World, 1955.

Elder, R. D. "A Comparison of the Oral Reading Groups of Scottish and American Children." (unpub. doctoral dissertation, University of Michigan, 1966.)

Eller, William and Mary Attea. "Three Diagnostic Reading Tests: Some Comparisons," in J. Allen Figurel (Ed.), Vistas in Reading, 1966 Proceedings, XI, Part 1. Newark, Delaware: International Reading Association, 1967.

Fairbanks, G. "The Relation Between Eye-movements and Voice in the Oral Reading of Good and Poor Silent Readers." Psychological Monographs, XLVIII (215) (1937), 78-107.

Farr, Roger and Nicholas Anastasiou. "Tests of Reading Readiness and Achievement: A Review and Evaluation." Reading Aids Series. V. L. Simula, ed. Newark, Delaware: International Reading Association, 1969.

Freeman, Frank S. Theory and Practice of Psychological Testing. New York: Holt, Rinehart and Winston, 1962.

Gallant, Ruth. "Use of Cloze Tests as a Measure of Readability in the Primary Grades," in J. Allen Figurel (Ed.), Reading and Inquiry, 1965 Proceedings, X, Part I. Newark, Delaware: International Reading Association, 1965.

Garlock, Jerry, Robert S. Dollarhide, and Kenneth D. Hopkins. "Comparability of Scores on the Wide Range and the Gilmore Oral Reading Tests." California Journal of Educational Research, XVI (March, 1965), 54-57.

Gates, Arthur I. The Improvement of Reading, Third Edition, New York: The Macmillan Co., 1947.

Gates, Arthur I. and A. S. McKillop. Gates-McKillop Reading Diagnostic Tests. New York: Bureau of Publications, Teachers College, Columbus University, 1962.

Gilmore, John V. "The Relationship Between Certain Oral Reading Habits and Oral and Silent Reading Comprehension." (unpub. Ed.D. dissertation, Harvard University, 1947.)

Goodman, Kenneth S. "A Linguistic Study of Cues and Miscues in Reading." Elementary English, XLII (October, 1965), 639-649.

_____. "Reading a Psycholinguistic Guessing Game." Journal of the Reading Specialist, IV (1967), 126-35.

_____. "Analysis of Oral Reading Miscues: Applied Psycholinguistics." Reading Research Quarterly, V(1) (1969), 9-29.

Goodman, Kenneth S. and Carolyn L. Burke. "Study of Children's Behavior While Reading Orally." (Final Report of U. S. Office of Education Project No. 5425) Detroit: Wayne State University, 1968.

Goodman, Yetta M. "A Psycholinguistic Description of Observed Oral Reading Phenomena in Selected Young Beginning Readers." (unpub. Ed.D. dissertation, Wayne State University, 1967.)

_____. "Using Children's Reading Miscues for New Teaching Strategies." Reading Teacher, XXIII (February, 1970), 455-459.

Gray, William S. Gray Oral Reading Test. Indianapolis, Indiana: Bobbs-Merrill, 1963.

Hardin, Veralee B. and Wilbur S. Ames. "A Comparison of the Results of Two Oral Tests." Reading Teacher, XXII (January, 1969), 329-334.

Harris, Albert J. How to Increase Reading Ability. New York: David McKay, 1970.

Hatch, Shirley and William D. Sheldon. "Strengths and Weaknesses in Reading of a Group of Fourth Grade Children." Elementary English, XXVII (April, 1950), 254-260.

Herlin, Wayne Richard. "A Comparison of Oral Reading Errors on the Monroe Diagnostic Reading Examination and the Durrell Analysis of Reading Difficulty." (unpub. Ph.D. dissertation, University of Utah, 1963), Dissertation Abstracts.

_____. "A Comparison of Oral Reading Errors on the Monroe Diagnostic Reading Examination and the Durrell Analysis of Reading Difficulty." (unpub. Ph.D. dissertation, University of Utah, 1963.)

Hill, Mary B. "Experimental Procedures in the Study of the Process of Word Discrimination in Reading." Journal of Educational Research, XXIX (February, 1936).

Ilg, F. L. and L. B. Ames. "Developmental Trends in Reading Behavior," Journal of General Psychology, LXXVI (1950), 291-312.

Jastak, J. F. and S. R. Jastak. The Wide Range Achievement Test (WRAT). Wilmington, Delaware: Guidance Associates, 1946.

Kagan, J. "Reflection-Impulsivity and Reading Ability in Primary Grade Children." Child Development, XXXVI, 1965, 609-628.

Kaluger, George and Clifford J. Kolson. Reading and Learning Disabilities. Columbus, Ohio: Charles E. Merrill, 1969.

Kasdon, Lawrence M. "Silent Reading Before Oral Reading," Ohio Reading Teacher, II (October, 1967), 6-8.

_____. "Oral Versus Silent-Oral Diagnosis," in Dorothy L. DeBoer (Ed.), Reading Diagnosis and Evaluation, Proceedings of the Thirteenth Annual Convention. Newark, Delaware: International Reading Association, 1970, 86-92.

Kelley, T. L., R. Madden, E. Gardner, and H. Rudman. Stanford Achievement Test, Form W, Primary II. New York: Harcourt, Brace, and World, 1964.

Killgallon, Patsy. "A Study of Relationships Among Certain Pupil Adjustments in Language Situations." (unpub. Ph.D. dissertation, Pennsylvania State College, 1942.)

Leibert, Robert E. "An Investigation of the Differences in Reading Performance on Two Tests of Reading." (unpub. Ed.D. dissertation, Syracuse University, 1965.)

Long, Donna J. "An Analysis of the Reading Difficulties of Retarded Readers in Second, Fourth, and Sixth Grades." (unpub. Ph.D. dissertation, University of Iowa, 1959.)

Lyman, Howard B. Peabody Picture Vocabulary Test. Sixth Mental Measurement Yearbook. Highland Park, New Jersey: Gryphon Press, 1965.

McCracken, Robert A. "The Oral Reading Performance of a Second Grade Class Using an Informal Reading Test." Journal of Educational Research, LV (November, 1961), 113-117.

_____. "Standardized Reading Tests and IRIs." Education, LXXXII (February, 1962), 366-369.

_____. "Using Reading as a Basis for Grouping." Education, (February, 1964), 357-359.

_____. Standard Reading Inventory. Bellingham, Washington: Pioneer Printing, 1966.

_____. "The Informal Reading Inventory as a Means of Improving Instruction." The Evaluation of Children's Reading Achievement. Perspectives in Reading No. 8, T. C. Barrett (Ed.). Newark, Delaware: International Reading Association, 1967.

_____ and Neill D. Mullen. "The Validity of Certain Measures in an I.R.I.," in William K. Durr (Ed.) Reading Difficulties: Diagnosis, Correction, and Remediation. Newark, Delaware: International Reading Association, 1970.

McLeod, Laurence S. "The Influence of Increasing Difficulty of Reading Material Upon Rate, Errors, and Comprehension in Oral Reading." Elementary School Journal, XVIII (March, 1918), 523-532.

Madden, Mable and Marjorie Pratt. "An Oral Reading Survey as a Teaching Aid." Elementary English Review, XVIII (April, 1941), 122-126, 159.

Malmquist, Eve. Factors Related to Reading Disabilities in the First Grade of the Elementary School. Stockholm: Almquist and Wiksell, 1958.

Maxwell, James. Reading Tests and Reviews. Highland Park, New Jersey: Gryphon Press, 1968.

Monroe, Marion. Children Who Cannot Read. Chicago: University of Chicago Press, 1932.

_____. "Diagnosis and Treatment of Reading Disabilities." Educational Diagnosis. NSSE Thirty-fourth Yearbook. Bloomington, Illinois: Public School Publishing Company, 1935.

_____. "Methods for Diagnosis and Treatment of Cases of Reading Disability." Genetic Psychology Monographs, IV (1928), 335-456.

Mulroy, Mary Dorothea. "The Improvement of Reading Through An Analysis of Errors in Oral Reading." (unpub. Master's thesis, University of Chicago, 1932.)

Neville, Donald. "The Relationship Between Reading Skills and Intelligence Test Scores." The Reading Teacher, 1965, XVIII, 257-62.

Nurss, Joanne R. "Oral Reading Errors and Reading Comprehension." Reading Teacher, XXII (March, 1969), 523-527.

_____. "A Diagnostic Comparison of Two Third Grade Reading Classes," in William K. Durr (Ed.), Reading Difficulties: Diagnosis, Correction, and Remediation, 1970.

Packman, Linda Arlene. "Relationships Between Selected Measures of Behavior and Levels of Reading Comprehension for Good, Average, and Poor Readers." (unpub. Ed.D. dissertation, University of Pennsylvania, 1970.)

Patty, Delbert L. "A Comparison of Standardized Oral Reading Test Scores and Informal Reading Inventory Scores." (unpub. Ed.D. dissertation, Ball State University, 1965.)

Payne, Cassie Spencer. "The Classification of Errors in Oral Reading." Elementary School Journal, XXXI (October, 1930), 142-146.

Pflieger, Elmer F. "A Study of Reading Grade Levels." Journal of Educational Research, XLII (March, 1949), 541-546.

Powell, William R. "Reappraising the Criteria for Interpreting the Informal Inventories," in Dorothy DeBoer (Ed.), Reading Evaluation, 1968 Proceedings, Volume 13, Part 4. Newark, Delaware: International Reading Association, 1969.

_____. "The Validity of the Instructional Reading Level," in R. E. Leibert (Ed.), Diagnostic View Points in Reading. Newark, Delaware: International Reading Association, 1971.

Ray, Darrel. "A Qualitative Analysis of Oral Reading Behavior." (unpublished paper, Oklahoma State University, 1969.)

Robeck, Mildred C. "Readers Who Lacked Work Analysis Skills: A Group Diagnosis." Journal of Educational Research, 1963, LVI, 432-34.

Robinson, Helen, ed. Gray Oral Reading Tests. Indianapolis, Indiana: Bobbs-Merrill, 1963.

_____. Reading Tests and Reviews. Highland Park, New Jersey: Gryphon Press, 1968.

Schale, Florence C. "Changes in Oral Reading Errors at Elementary and Secondary Levels." (unpub. Ph.D. dissertation, University of Chicago, 1964.)

Schiffman, Gilbert. "Current Information on Achievement Levels in Reading," in J. Allen Figurel (Ed.), Reading as an Intellectual Activity, 1963 Proceedings, Volume 8. New York: Scholastic, 1963.

Schummers, John L. "Word Pronunciation in the Oral Sight-Reading of Third-Grade Children." (unpub. Ph.D. dissertation, University of Minnesota, 1956.)

Shedd, Charles L. "Ptolemy Rides Again or Dyslexia Doesn't Exist?" The Alabama Journal of Medical Science, V (October, 1968), 481-503.

Sheldon, William D. and Shirley Hatch. "Strengths and Weaknesses in Reading of a Group of Third-Grade Children." Elementary School Journal, L (April, 1950), 445-452.

_____. "Strengths and Weaknesses in Reading of a Group of Second-Grade Children." Educational Administration and Supervision, XXXVII (November, 1951), 405-414.

Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

Sipay, Edward R. "A Comparison of Standardized Reading Achievement Test Scores and Functional Reading Levels." (unpub. Ph.D. dissertation, University of Connecticut, 1961.)

Sister M. Catherine Elizabeth Davis, R.S.M. "The Relative Effectiveness of Certain Evaluative Criteria for Determining Reading Levels." (unpub. Ed.D. dissertation, Temple University, 1964.) Dissertation Abstracts.

Smith, Lulu Davis. "An Analysis of Errors in Oral Reading." (unpub. Master's thesis, Harrisburg, Mississippi, 1954.) Cited by Ralph C. Staiger, "Meaning Clues to Word Recognition," Education, LXXV (May, 1955), 579.

Spache, George. "A Comparison of Certain Oral Reading Tests." Journal of Educational Research, XLIII (February, 1950), 441-452.

_____. Good Reading for Poor Readers. Champaign, Illinois: Garrard Publishing Company, 1968.

_____. Reading Tests and Reviews. Highland Park, New Jersey: Gryphon Press, 1968.

Stafford, Clarice M. Salli. "An Analysis of the Types of Oral Reading Errors in a Sample of Fourth Grade Pupils." (unpub. Ed.D. dissertation, Wayne State University, 1967.) Dissertation Abstracts.

Stuever, Rita Fae. "Analysis of Relationship of Length of Passage to Categories of Oral Reading Errors." (unpub. Ed.D. dissertation, Oklahoma State University, 1969.)

Swanson, Donald E. "Common Elements in Silent and Oral Reading." Psychological Monographs, XLIV, No. 3 (1937), 36-60.

Tatham, Susan Masland. "Reading Comprehension of Materials Written with Select Oral Language Patterns: A Study at Grades Two and Four." Reading Research Quarterly, V (Spring, 1970), 402-426.

Taylor, E. A. and J. H. Crandall. "A Study of the Norm Equivalence of Certain Tests Approved for the California State Testing Program." California Journal of Educational Research, XIII (1962), 186-192.

Torgerson, T. L. Reading Tests and Reviews. Highland Park, New Jersey: Gryphon Press, 1968.

Watkins, Mary. "A Comparison of the Reading Proficiencies of Normal-Progress and Reading Disability Cases of the Same I.Q. and Reading Level." (unpub. Ph.D. dissertation, University of Minnesota, 1953). Dissertation Abstracts XIV, Pt. 1, 1953, 644.

Weber, Rose Marie. "A Linguistic Analysis of First-Grade Reading Errors." Reading Research Quarterly, V (3), 427-451, 1970.

_____. "The Study of Oral Reading Errors: A Survey of the Literature." Reading Research Quarterly, IV (1), 96-119, 1968.

Wechsler, David. Wechsler Intelligence Scale for Children. New York: The Psychological Corp., 1949.

Wells, F. L. "A Glossary of Needless Reading Errors." Journal of Experimental Education, IV (September, 1935), 34-43.

Wert, James, Charles Neidt, and J. Stanley Ahmann. Statistical Methods in Educational and Psychological Research. New York: Appleton-Century-Crofts, Inc., 1954.

Williams, Joan Lee. "A Comparison of Standardized Reading Test Scores and Informal Reading Inventory Scores." (unpub. Ph.D. dissertation, Southern Illinois University, 1963.)

APPENDIX

IDENTIFICATION OF SYMBOLS

APPENDIX

IDENTIFICATION OF SYMBOLS

Dur	Durrell Analysis of Reading Inventory
SRI	Standard Reading Inventory
Gates	Gates-McKillop Reading Diagnostic Tests
DI	Durrell INSTRUCTIONAL
DF^1	Durrell FRUSTRATION 1
DF^2	Durrell FRUSTRATION 2
GI-4	Gates-McKillop first four paragraphs
GI	Gates-McKillop INSTRUCTIONAL
GF^1	Gates-McKillop FRUSTRATION 1
GF^2	Gates-McKillop FRUSTRATION 2
SRII	Standard Reading Inventory INSTRUCTIONAL
$SRIF^1$	Standard Reading Inventory FRUSTRATION 1
$SRIF^2$	Standard Reading Inventory FRUSTRATION 2

VITA

Margery Lois Berends

Candidate for the Degree of

Doctor of Education

Thesis: AN ANALYSIS OF ERROR PATTERNS, RATES, AND GRADE EQUIVALENT SCORES ON SELECTED READING MEASURES AT THREE LEVELS OF PERFORMANCE

Major Field: Higher Education

Biographical:

Personal Data: Born in Chase County, Kansas, the daughter of Mr. and Mrs. Harry H. Berends.

Education: Attended public elementary schools in Chase County, Kansas; graduated from Matfield Rural High School, Matfield Green, Kansas in 1944; received the Bachelor of Arts and Bachelor of Science degrees from Kansas State Teachers College, Emporia, Kansas in 1950; received the Master of Arts degree with a major in elementary education from the University of Colorado, Boulder, Colorado, in 1958; completed requirements for the Doctor of Education degree at Oklahoma State University, Stillwater, Oklahoma in July, 1971.

Professional Experience: Rural school teacher in Chase County, Kansas, 1944-45; teacher of grades four, five, and six at Clements Grade School, Clements, Kansas, 1945-47; fifth grade teacher in Hutchinson Public Schools, Hutchinson, Kansas, 1950-1962; remedial reading teacher-consultant in Hutchinson Public Schools, Hutchinson, Kansas, 1962-1966; assistant professor of education (part-time), NDEA Reading Institute at Oklahoma State University, summers of 1965 and 1967; graduate assistant at Oklahoma State University, 1966-67; assistant professor of education (part-time) at Oklahoma State University, summer of 1966; consultant for reading programs in various Oklahoma schools, 1966-1968; instructor in Education Extension, Oklahoma State University, 1967-68; assistant professor and director of the Reading Center, Northwestern State College, Alva, Oklahoma, 1968-