



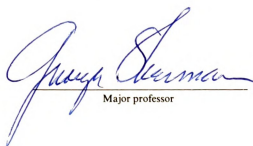
This is to certify that the
 dissertation entitled
 A STUDY OF AVERAGE THIRD GRADE READERS'
 ORAL READING PERFORMANCE IN MATERIAL
 OF VARYING FRY DETERMINED READABILITIES

presented by

Janet Sue Dixon

has been accepted towards fulfillment
 of the requirements for

Ph.D. degree in Education


 Major professor

Date May 13, 1987



RETURNING MATERIALS:

Place in book drop to
remove this checkout from
your record. FINES will
be charged if book is
returned after the date
stamped below.

--	--	--

A STUDY OF AVERAGE THIRD GRADE READERS'
ORAL READING PERFORMANCE IN MATERIAL
OF VARYING FRY DETERMINED READABILITIES

By
Janet Sue Dixon

A DISSERTATION
Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
Department of Teacher Education

1987

Copyright by
JANET SUE DIXON
1987

ABSTRACT

A STUDY OF AVERAGE THIRD GRADE READERS' ORAL READING PERFORMANCE IN MATERIAL OF VARYING FRY DETERMINED READABILITIES

By

Janet Sue Dixon

The Problem

In order to match readers of given ability with materials of suitable difficulty, practitioners will frequently ask the reader to read aloud from the material in question. The time involved with such a procedure, however, makes it impractical when large numbers of readers or materials are under consideration. The question raised by this study is whether or not standardized test scores and Fry Readability Graph data can be used to effectively accomplish the same purpose.

Method

The subjects in this study were 50 third grade students with grade equivalency scores on the Reading Test of the California Achievement Tests within three months above or below their grade placement at the time of testing. Each subject read aloud from a set of five selections, one each having Fry determined readabilities of first, second, third, fourth and fifth grade. Thus the subjects reading achievement was held relatively constant while the readability of the selections varied. Traditional oral

reading assessment procedures were used to evaluate the readings. It was expected that the readers would make more miscues and read more slowly as the readability of the selections increased and that the first, third and fifth grade paragraphs would be at the readers' independent, instructional and frustrational reading levels respectively.

Findings

Generally speaking the readability scores did not appear to discriminate well. When only quantity of miscues was considered performance on all paragraphs tended to be virtually the same and at the readers' instructional reading levels. In terms of rate, unacceptable miscues and fluency the second grade paragraph appeared easiest and the fifth grade selection the most difficult. Additional data analysis, however, found miscues were highly predictable and factors triggering them could be identified. These factors were not related to those traditionally associated with readability formulae, but were virtually identical to factors reported in miscue research conducted more than ten years earlier.

DEDICATION

To the memory of my mother

Katherine Boley Dixon

because she would be proud of this.

ACKNOWLEDGEMENTS

My sincere thanks to my committee chairman, Dr. George Sherman, for his patience, understanding and support throughout this long study. My thanks also to the members of my committee, Dr. John Baldwin, Dr. Perry Lanier and Dr. Lonnie McIntyre for their encouragement and patience as well.

I am also deeply indebted to the following people in the Bay City Public School System, without whose cooperation this study could not have been done:

Dr. Charles Link, Assistant Superintendent for Curriculum and Instruction, for his support and internal coordination of the study,

Dr. Douglas MacPherson, Director of Research and Evaluation, for his assistance during data analysis,

Mr. Wesley Garner, Director of Chapter I Services, for cooperation during the testing phase,

The Principals in the participating buildings:
Mr. Leon Katzinger, Mr. Clem Kaye, Mr. Warren
Likien, Mr. Pete Mayo and Mr. Ron Stachowiak, for
their help in communicating with parents and
arranging for data collection,

The third grade teachers in the participating
buildings: Mrs. Beverly Ballor, Mrs. Gloria
Brooks Garcia, Mrs. Janice Harbour, Mrs. Edith
Hinkley, Mrs. Nancy Lusher, Mrs. Nancy Maier, Mrs.
Janet Moll, Mrs. Rita Narlock, Mrs. Sandra
Remensnyder, Mrs. Sandra Stachowiak, Mrs. Susan
Tanner, Mrs. Irene Tobias, Mrs. Shirley Wegener
and Mrs. Joan Wilson, for so graciously adjusting
their schedules in order to allow subjects in
their rooms to participate in the data collection.

But most of all I am indebted to the children who
participated in the study, to their parent who so trustingly
gave their permission and in particular, to the fifty
children who read the research passages with such eagerness
and enthusiasm. They will always be remembered with great
delight.

TABLE OF CONTENTS

LIST OF TABLES.....	xi
---------------------	----

CHAPTER

I. THE PROBLEM.....	1
Overview.....	1
Background of the Problem.....	1
Introduction to the Problem.....	5
Measuring Reading Ability.....	6
Measuring Readability.....	10
Concerns in Test-Formula Matching.....	12
Statement of the Problem.....	14
Purpose of the Study.....	15
Questions Directing the Study.....	16
Need for the Research.....	17
Definition of Terms.....	18
II. REVIEW OF THE LITERATURE.....	22
Overview.....	22
Part I: Determining Readability.....	22
Historical Background.....	22
Development of Readability Formulae.....	25
Historic Trends.....	27
Methodology.....	28
Limitations of Readability Formulae.....	29
Limitations in Factors Studied.....	30
Limitations in Criterion.....	33
Validity of Readability Formulae.....	36
Original Presentation of the	
Readability Method.....	37
Original Criterion Prediction.....	37
Correlation with Other Readability	
Formulae.....	38
Experimental Validation Studies.....	40
Validation Against Outside Criteria.....	42
Oral Reading Criteria in Readability	
Research.....	43
Development of the Fry Graph.....	44
Recent Trends in Readability Prediction.....	48
Ease and Speed of Use.....	49
Criterion Developments.....	53
Readability in the Early Elementary	
Grades.....	56

Oral Reading in Readability Measurement and Prediction.....	60
Part II: Determining Reading Ability.....	62
Introduction.....	62
Standardized Tests.....	62
The California Achievement Tests.....	67
Oral Reading Assessment.....	69
Development of Traditional Practices..	70
Traditional Versus Psycholinguistic Diagnosis.....	78
Summary of the Literature Review.....	85
Formula Limitations.....	86
The Fry Graph.....	87
The Assessment of Reading Ability....	87
III. DESIGN OF THE STUDY.....	90
Overview.....	90
Questions Guiding the Study.....	91
Hypotheses.....	92
Population.....	93
Sample Selection.....	93
Measurement of Student Reading Ability....	94
Instrument Selection and Construction.....	94
Passage Selection.....	94
Determination of Readability.....	96
Data Collection.....	96
Data Recording.....	97
Data Analysis.....	98
IV. PRESENTATION AND ANALYSIS OF RESULTS.....	100
Introduction.....	100
Presentation of Results.....	101
Additional Data Analysis.....	110
Presentation of Additional Data Analysis..	113
Descriptive Miscue Analysis.....	122
Summary of Results.....	129
Summary of Results from Four Measures of Difficulty.....	129
Summary of Results from Functional Reading Levels.....	132
Summary of Results from Miscue Frequency Data.....	134
V. SUMMARY AND CONCLUSIONS.....	136
Introduction.....	136
Summary.....	136
Conclusions.....	138
Discussion.....	141
Implications.....	144
Recommendations.....	146

APPENDICES

APPENDIX A

Letter from Principals to Parents.....	149
Parental Permission Slip.....	151

APPENDIX B

The Research Passages.....	152
----------------------------	-----

APPENDIX C

The Fry Readability Graph.....	158
--------------------------------	-----

APPENDIX D

Formulae and Computational Procedures.....	159
--	-----

APPENDIX E

Summary of Computations.....	161
------------------------------	-----

REFERENCES.....	168
-----------------	-----

LIST OF TABLES

TABLE

IV-1	MEANS OF WORD RECOGNITION ACCURACY SCORES BASED ON TOTAL NUMBER OF MISCUES.....	101
IV-2	MEANS OF READING RATE SCORES.....	102
IV-3	PERCENTAGES OF SUBJECTS READING AT EACH FUNCTIONAL READING LEVEL ON PARAGRAPH #1.....	105
IV-4	PERCENTAGES OF SUBJECTS READING AT EACH FUNCTIONAL READING LEVEL ON PARAGRAPH #3.....	107
IV-5	PERCENTAGES OF SUBJECTS READING AT EACH FUNCTIONAL READING LEVEL ON PARAGRAPH #5.....	109
IV-6	MEANS OF WORD RECOGNITION ACCURACY SCORES WHEN ONLY UNACCEPTABLE MISCUES WERE COUNTED.....	114
IV-7	MEANS OF GENERAL IMPRESSION OF FLUENCY SCORES.....	115
IV-8	FREQUENCIES OF MISCUES OCCURRING ON EACH WORD IN PARAGRAPH #1.....	117
IV-9	FREQUENCIES OF MISCUES OCCURRING ON EACH WORD IN PARAGRAPH #2.....	118
IV-10	FREQUENCIES OF MISCUES OCCURRING ON EACH WORD IN PARAGRAPH #3.....	119
IV-11	FREQUENCIES OF MISCUES OCCURRING ON EACH WORD IN PARAGRAPH #4.....	120
IV-12	FREQUENCIES OF MISCUES OCCURRING ON EACH WORD IN PARAGRAPH #5.....	121

IV-13	SUMMARY OF DIFFERENCES FOUND BETWEEN PARAGRAPHS ON FOUR MEASURES OF DIFFICULTY.....	131
IV-14	PERCENTAGES OF SUBJECTS READING AT EACH FUNCTIONING READING LEVEL ON EACH PARAGRAPH.....	133

CHAPTER I

THE PROBLEM

Overview

In this chapter the problem will be introduced, background information will be presented and the importance of the problem will be established. The questions directing the research will be given and terminology pertinent to the study will be defined.

Background of the Problem

About 1840 the McGuffey Readers introduced the concept of graded difficulty to American Schools. Forerunners of the modern basal series, their author, William Holmes McGuffey had based the texts on two important premises which still govern the way most reading is taught today: (a) The difficulty of reading material (readability) can be controlled and (b) controlling readability facilitates learning to read.

While McGuffey's method for controlling readability would be debatable, the relationship between task difficulty and learning which he recognized, has since been well supported in the research and by successful instructional practices. It would eventually affect not only the teaching of reading but the development of instructional theory and the structure of curriculums. Taxonomies (Bloom, 1956) hierarchies (Gagne, 1968; Gagne, 1969), task analysis

(Anderson and Faust, 1973, Chapter 3; DeCecco, 1968, Chapter 2), programmed instruction (Glaser, 1965; DeCecco, 1968, Chapter 12; Lumsdaine, 1960, 1964), and mastery learning (Carroll, 1963; Block, 1971; Block and Anderson, 1975; Bloom, 1976; Smith, 1977) would be among the terms and methods made familiar by educators and educational psychologists as they described the process of breaking complex learnings into simpler underlying tasks, usually arranged in some hierarchical form. Ideally the learner begins at a point in this sequence where he can succeed and master tasks of gradually increasing difficulty until the complex learning has been accomplished.

Maximizing success is central to this process of controlling task difficulty, for the facilitating effect of success on learning has long been recognized by virtually every learning theorist. On the other hand, while failure experiences may contribute positively to the learning process under some conditions, (Gage and Berliner, 1984, p. 396-197; Weiner, 1972) such experiences can also be devastating, and the undesirable consequences to the learner who repeatedly fails have been frequently documented. Such learners typically have shown increased anxiety, less persistence, lowered aspirations, increased tendency to repeat inappropriate responses or to use fantasy or superstitious behaviors rather than realistic problem solving strategies (Sears, 1940; Baker, 1941; Baker, Demo and Lewin, 1941; Lantz, 1945).

While failure situations can become self-defeating and are generally to be avoided, tasks that are too easy are undesirable also, for they will not produce the desired growth. As David Ausubel (1968) has observed:

If the material is too difficult, the learner accomplishes disproportionately little for the degree of effort he expends; if it is too easy, his accomplishments are disappointingly meager in terms of what he could have achieved were greater effort demanded of him. (p. 325)

In addition Ausubel notes:

Inappropriately easy material...fails to stimulate and challenge the learner adequately, fostering boredom and disinterest. (p. 326)

Ideally then the teacher seeks to find that place in the learning sequence where the tasks are of appropriate difficulty for the learner. Sometimes called the student's "instructional level", it is that point where the material offers some challenge but where the student is capable of handling that challenge without undue anxiety or frustration.

For some kinds of learnings the hierarchy involved can be arranged in a relatively linear progression, each subskill more or less prerequisite to the next. Finding the instructional level is largely a matter of testing for mastery of the underlying skills. Learning to read, however, tends to be a developmental process, characterized by stages of increasing complexity, involving many skills, abilities and understandings which the reader must combine more or less simultaneously and appropriately in order to

read a given selection. Finding the "instructional level" then, is not simply a matter of testing for specific skills, but depends on an evaluation of the reader's entire general level of functioning in relationship to the difficulty of the material being read.

In reading, this is commonly done by assessing the learner's oral reading performance directly in the material under consideration. This performance is typically evaluated using some variation of procedures and criteria popularized by Emmett Betts (1946) about 40 years ago. Betts distinguished at least three different reading levels: The instructional level, the independent level (material that is easily read) and the frustration level (material which is too difficult). From observation of the student's oral reading "errors", the teacher makes a determination of the level of difficulty of this material for this student. This information might then be combined with other knowledge, such as the student's interest in the subject, the length of the selection and consequently the persistence needed to finish it, or the format of the book, in deciding if the student will be able to successfully read the selection.

While finding a reader's instructional level may not be a simple, precise or expedient matter, it is important in the teaching of reading in order to eliminate the task avoidance responses commonly associated with frustration and failure. While task avoidance is certainly a hindrance in

any kind of learning it is particularly detrimental to reading progress since reading, like many other complex performances such as playing the piano, driving a car, or learning a sport, seems highly affected by practice.

Not only does practice affect reading by reinforcing and automating previously learned skills, but it is also necessary to integrate those skills into meaningful and fluent reading behaviors. In addition, we know that many people learn to read with little if any apparent formal instruction. Evidently what one needs to know to become a better reader can often be learned intuitively while reading, with little assistance from the teacher, if the teacher can only find material motivating enough so the learner will read it and easy enough so the learner can read it. Failure experiences, on the other hand, can lead to avoidance of further reading cutting off perhaps the most important means by which the failing learner could improve. Finding materials of suitable difficulty for the learner then becomes an integral part of developing reading proficiency.

Introduction to the Problem

If a practitioner wanted to know quickly whether a student could read a particular book, the most logical procedure would be to have the student read a few sample passages aloud. Based on this observation, the practitioner could then make a judgment concerning the student's ability



to handle the material. This procedure, in fact, is frequently used when the question concerns one reader and one book, but when the problem involves many children and many books, the time required for listening to each child read makes such a procedure impractical.

Obviously if there were some effective means of measuring student reading ability and some corresponding method for measuring passage difficulty (readability), the process of matching readers with materials would be greatly facilitated. Measures of both reading ability and readability to exist, and they are frequently used together to make decisions of this nature in research studies, textbook selection and development of new materials. However an examination of these measures poses important questions and suggests serious limitations concerning their use in this way.

Measuring Reading Ability Determination of a student's reading achievement is most frequently made based on results of either an Informal Reading Inventory (IRI) or a Standardized Reading Achievement test. With both methods the results are popularly reported using some form of grade level norm.

The Informal Reading Inventory is more apt to be used by teachers in special reading programs since it is a time consuming, individually administered test which does not lend itself to the structure of most classroom settings. It

can be teacher constructed from materials being used by the student or the teacher may choose to use one of the commercially published tests such as the Durrell Analysis of Reading Difficulty (Durrell, 1937, 1955), the Diagnostic Reading Scales (Spache, 1963, 1972) or the Classroom Reading Inventory (Silvaroli, 1965).

When giving an IRI, deviations from the text made by the reader while reading aloud are recorded. Then, usually using some variation of criteria first popularized by Emmett Betts (1946), the instructor decides if this material is at the student's independent, instructional or frustrational reading level. As previously noted, practitioners often use this procedure by itself to determine directly if a particular book is of suitable difficulty for a given student. In the IRI, however, the selections are presumably graded in difficulty corresponding to the grade levels of basal texts. In this respect it then becomes a prediction device. The assumption is that once the student's reading levels are established in terms of the grade level difficulties of the IRI passages, this information can automatically be transferred to other material, and when the student is reading another selection intended for that grade they can be expected to perform in a similar fashion.

Using an IRI to establish a student's instructional reading level assumes that all reading materials intended for a given grade are of the same level of difficulty. The publishers of basal texts, however, use individual standards

and standardization procedures in developing their books and they do not conform to any universal standard applicable for all basal publishers. Moreover, the standardization methods they use are not typically well described in an easily accessible manner, as they are in the documentation prepared by standardized test publishers. Thus it is difficult to compare norming procedures from series to series or even to know what procedures were used. Teachers, however, will frequently refer to one series as being "more difficult" than another, and formula determined readabilities of basal selections can differ greatly from series to series for materials intended for use by the same grade, and may even differ from selection to selection in the same book (Bradley and Ames, 1976, 1977; Eberwein, 1979). This would suggest that the difficulty of materials for a given grade may differ considerably. It would seem then that it cannot be assumed that a reader's performance based on one basal series will automatically indicate performance in another, nor can the passages from one IRI, and their grade level indicators, necessarily be used as a meaningful standard for judging the difficulty a reader may encounter in other materials.

Standardized tests are most frequently used in classrooms to assess reading achievement since they are fast, convenient and highly reliable. They are excellent for comparing the performances of readers, but they pose serious problems when used to determine reading levels.

They are primarily tests of comprehension and offer no opportunity to observe reading behaviors directly, or to compare the reader's performance to a set of criterion tasks. The scores are based on comparisons of students with a standardization group and are not necessarily related to the difficulty level of reading selections. A grade equivalency score of 2.0 on a standardized test, therefore, does not mean the test taker was able to comprehend material with a beginning second grade readability, but that s/he was able to answer as many questions correctly over the entire test as did the average beginning second grader in the standardization group.

There is empirical evidence that the grade equivalency scores from standardized tests cannot be used to place children at their instructional level, for when they are compared to IRI results they usually yield significantly higher grade placement scores. Using them for this purpose will probably result in students being placed at their frustration level (Sipay, 1964; Glaser, 1964). Also, most norming procedures use one administration of the test during the school year and the between grade norms for each month of the year are interpolated from these results. This practice assumes that reading growth proceeds at an even rate, an assumption that is not supported by research studies (Bernard, 1966; Lennon, 1951).

Finally grade equivalency scores imply that students in different grades with the same scores have the same reading

achievement. However, students scoring above their grade placement and students scoring below their grade placement may perform quite differently on an IRI even though their scores from a standardized test indicated the same grade level in reading achievement (Glaser, 1964; Farr and Carey, 1986, p. 153-154).

Measuring Readability The development of objective methods for measuring the difficulty of reading passages has also presented serious problems. While it is relatively easy to observe that some materials are more difficult to read than others, it is not so easy to identify or measure the factors which account for that difference.

Early in this century, using improved statistical procedures in factor analysis, researchers began to systematically investigate aspects of writing which appear to influence the ease with which material is read and understood. Interest generated by these early studies, along with increasing demands in society to understand and control reading difficulty, eventually led to the development of numerous formulae for calculating what is now termed "readability".

Readability formulae attempt to give an objective measure of factors within text which may affect the reading accomplishment needed to handle the passage or the ease with which the material can be read and understood. By their nature, these devices must be based on only a limited

set of factors which can influence the difficulty level of a passage, for while many factors have been studied, invariably only a few emerge as significant enough (or measurable enough) to be included in the final formula. These factors usually include some measure of word difficulty and some measure of sentence complexity. Formulae cannot measure conceptual complexity, reader interest, reader motivation, topic organization, figurative language or such physical factors as format, illustrations, or size of print, all of which may also contribute to the difficulty one encounters when reading a given passage. Moreover, results of studies concerning the validity of readability prediction methods have been conflicting and those studies concerning the ability of the devices to go beyond prediction of relative difficulty to prediction of difficulty for students in given grade levels, have been generally negative.

Because of these limitations, readability formulae have met severe criticism from many leaders in the field of reading. At best, these authorities, and even the authors of the formulae themselves, caution that these devices should be used with great care and only as rough estimates of relative difficulty. In spite of such warnings, however, the grade level indexes yielded by these formulae are still frequently combined with the grade equivalency scores yielded from standardized test data to make decisions concerning the appropriateness of difficulty of certain

materials for given readers.

Concerns in Test-Formula Matching Even if we were assured of the validity of the tests and formulae involved to measure reading ability and passage difficulty respectively, the test-formula matching procedure assumes that the two measures are congruent. The evidence would suggest that probably they are not.

Standardized tests and readability formulae were not developed using the same criterion measures nor were they designed to be used together specifically for matching readers with materials of appropriate difficulty. The test-makers' prime concern has not been readability but rather comparison of performances. Therefore, when readability formulae are used to assess standardized test passages they do not reveal an orderly progression of gradually increasing difficulty as one might expect, and it is possible for a student to receive a grade equivalency score of 2.0 on a standardized test, without any passage on the test having a readability of 2.0.

Moreover, standardized tests are typically measures of comprehension. Readability formulae, on the other hand, do not measure comprehension directly, but rather deal with factors in the text which may affect comprehension. It is also evident that some authors never meant their formula to be an indicator of the level of accomplishment needed as associated with developmental reading achievement, but

rather as measures of "clear writing" style which increases the ease of comprehension for adult readers (McElroy, 1953; Flesch, 1948, 1949, 1954, 1958).

In the literature review for her study concerning "Easy to Read" books for children, Margaret Paolo (1977) found the use of oral reading in readability research has received little attention. Validity studies which have attempted to compare formula predictions with reader's performance have typically used silent reading comprehension, rather than word recognition, as the measure of that performance, even though oral reading would seem a more logical choice since it, like the formulae involved, does not assess comprehension directly but rather deals with word and sentence factors in the text which may affect comprehension.

Because comprehension has been used so exclusively in such validity studies it has left practitioners with little information concerning the usefulness of the various readability formulae. If a reader's achievement test scores and the formula's data suggest a given reader should be able to read a given selection, but we find his comprehension in the material to be low, the results do not tell us if the reader was unable to handle the text at all, or if he could read the text but found the situations or concepts presented to be too complex or unfamiliar for his understanding. If his comprehension of the material is good, it still does not assure us that this material is at or below the student's instructional reading level, for it is possible for a reader

to maintain an acceptable level of comprehension even though experiencing frustration due to word recognition difficulties. This might especially be true if the topic involved is familiar or the selection is short.

Statement of the Problem

Observation of oral reading performance directly in the material under consideration is frequently used to assess a single reader's ability to read a given selection. The time involved in such a procedure, however, makes it impractical when large numbers of students are involved. This has led to the practice of combining standardized test scores as a measure of student reading ability, and readability formula data as a measure of passage difficulty to determine if certain readers will be able to read certain materials.

It would appear that direct observation of the reader's performance in the material provides a more acceptable means for matching readers with materials. The question raised by this study is whether or not standardized test scores and readability formula data can be used together to effectively accomplish the same purpose. If they can, then we would expect a great deal of consistency between and among oral reading, standardized test scores, and readability measures. However, as the preceding text has noted, this is often not the case. Do oral reading assessment procedures, then, which are primarily measures of word recognition, and standardized tests, which are primarily measures of

comprehension, and readability formulae, which attempt to measure characteristics in the text which may affect both comprehension and word recognition all sample enough of the same reading factors to allow a reader's performance on a standardized test to predict that reader's oral reading performance in material measured by a readability formula? In greater detail, to what extent are a reader's grade equivalency scores as measured by a standardized test predictive of his functional reading levels as established by his oral word recognition abilities when reading material of a formula determined readability? And is this test-formula relationship strong enough to make it a useful tool for practitioners and justify its use as a basis for making judgments and decisions in research studies, text selection, and the development of new instructional materials?

Purpose of the Study

The purpose of this study is to investigate the relationship between grade equivalency scores from the California Achievement Tests and Fry Readability Graph (1968) data. Specifically, it examines how effectively grade equivalency scores from the Reading Subtest of the California Achievement Tests, when used to identify a group of "average" readers, and Fry Readability Graph estimates of material difficulty, will predict the degree of difficulty a reader will encounter when reading orally from material of varying Fry determined readabilities. Subsequently, the

study will also investigate the relationship between the Readability Graph scores of these selections and (a) the number of oral reading errors (miscues) made by the readers, (b) the readers' reading rate and (c) the readers' functional reading levels.

Questions Directing the Study

If the grade equivalency scores from the California Achievement Tests and Fry readability data provide an effective means for matching readers with materials of appropriate difficulty, then we would expect the readers to make more word recognition errors (miscues) and to read more slowly as the readability of the passages increases. We would also expect the readers to read the passage with first grade readability at their independent reading level, the passage with third grade readability at their instructional reading level and the passage with fifth grade readability at their frustrational reading level. Based on these expectations, the following questions were posed to be answered by this study.

When average third grade readers, as determined by the Reading Test of the California Achievement Tests, are reading selections with varying Fry determined readabilities:

1. Will the readers' word recognition accuracy, based on their oral reading errors (word miscues), decrease as the grade level readability scores of the selections increase?

2. Will the readers' reading rate (number of words read per minute) decrease as the grade level readability scores of the selections increase?

3. Will the readers read material with a first grade readability at their independent reading level?

4. Will the readers read material with a third grade readability at their instructional reading level?

5. Will the readers read material with a fifth grade readability at their frustrational reading level?

Need for the Research

In spite of continual criticism, the use of readability estimates appears to be rising. Publishers increasingly list estimates of difficulty of their materials with the names of the formula (or frequently formulae) used to make those determinations. Increased demand for "High Interest, Low Vocabulary" and "Easy to Read" books places continual pressure on authors to control readability in their writing. It is probably only the time involved in using the formulae that has kept their use from becoming more prevalent. As microcomputers become commonplace, however, the development of more complex but faster and easier to use computerized formulae promises to remove this restriction and further increase their use.

The widespread acceptance of the readability concept and the demand for readability information and control underscores the serious need teachers and others have for

some indication of the suitability of a given material for a given reader, even if that information may be questionable and unproven. It is important, therefore, that studies be conducted that either help practitioners define readability scores operationally, discredit their use, or provide estimates of how much confidence can be placed in them. Such studies might also indicate how more predictive reader ability - readability indexes could be developed.

Definition of Terms

Readability: Refers in general to any factor that affects the ease with which a selection can be read and understood. More specifically it has come to be associated with the factors measured by readability formulae. In this study it will refer to the scores from the Fry Readability Graph as computed by the text analysis computer program School Utilities Volume 2, available from the Minnesota Educational Computer Consortium.

Fry Readability Graph: A nomograph developed by Edward Fry, Rutgers University. It estimates readability using sentences per 100 words and syllables per 100 words. For books and longer selections, the final estimate is based on an average of three samples. Because the selections in this study are short, the Fry estimate will be based on the actual text involved.

Functional Reading Level: A term used to refer collectively to a reader's independent, instructional and frustrational reading levels.

Independent Reading Level: Refers to material which a reader can read easily. In this study it will refer to material a reader can read with 99% or better word recognition accuracy.

Instructional Reading Level: Refers to material a reader is capable of reading with some help. It is the level of difficulty which, ideally, should be used for instruction. In this study it will refer to material a reader can read with 95% to 99% word recognition accuracy.

Frustrational Reading Level: Refers to material that is too difficult for a reader to read under any conditions. In this study it will refer to material a reader reads with 90% or less word recognition accuracy.

Miscue: A deviation from text which a reader makes when reading orally. The term miscue is generally preferred to the terms "mistake" or "error" because it more accurately suggests what is occurring during the reading process, suggesting that such deviations from text are not random errors but, in fact, are cued by the thought and language of the reader in his encounter with the written material

(Goodman and Burke, 1972).

Oral Reading Errors: Refers to a miscue made by a reader when reading orally. The following types of miscues will be counted as oral reading errors in this study: (a) Omissions, (b) insertions, (c) substitutions, (d) partial or gross mispronunciations (not caused by dialect or speech difficulties) and (e) words aided.

Betts' Criteria: Criteria, developed and popularized by Emmett Betts (1946), and used widely in oral reading assessment procedures to determine a reader's functional reading levels. In this study the Betts' word recognition criteria of 99% word recognition accuracy will be used to designate a selection as being at a reader's independent reading level, 95% to 99% word recognition accuracy will be used to designate a selection as being at a reader's instructional reading level and less than 90% word recognition accuracy will be used to designate a selection as being at a reader's frustrational reading level.

Reading Rate: Refers to the speed with which material is read. Researchers have used reading rate as an index of speed of response which they in turn consider as an indicator of automaticity (Samuels, 1979). In this study reading rate will be given in terms of words read per minute and will be determined by dividing the number of

words in a selection by the number of seconds taken by the reader to read the selection, multiplied by 60.



CHAPTER II

REVIEW OF THE LITERATURE

Overview

In this chapter a review and synthesis of selected literature relevant to the study will be presented. The review will be divided into two parts. Part I, Determining Readability, will concentrate on (a) the development of the readability concept, its measurement, and prediction and (b) the use of oral reading in readability prediction and validation. Part II, Determining Reading Ability, will concentrate on the development and use of (c) standardized tests as a measure of reading ability, and (d) the Informal Reading Inventory and oral reading assessment procedures.

Part I

Determining Readability

Historical Background

The awareness that reading material can differ in difficulty and the search for ways to control that difficulty are probably as old as writing itself. Klare and Buck (1954, p. 42) have noted, for instance, that much of early literary criticism was concerned with comparing "ornate" and "plain" styles among writers, and Klare (1963, p. 29) cites a quotation from I Corinthians 14:9 as a favorite among advocates of clear language: "Except ye utter

by the tongue words easy to be understood, how shall it be known what is spoken? For ye shall speak into the air."

While awareness of style and admonishments to writers may be evident early in the history of writing, the idea that readability can be consciously and systematically controlled seems to be a much more recent historical development. In the early years of American education, for instance, there was evidently no attempt to prepare books specifically to meet the needs of beginning readers. Colonial children learned to read by struggling as best they could with whatever books were available. Usually those books were of a religious nature intended for adults rather than children. Chief among them, for instance, was the New England Primer, which was so named, not because it was the child's first book, or because it contained easy to read material appropriate for beginning readers as the term "primer" implies today, but because it contained religious teachings which were considered "primary" for the child's spiritual existence (Smith, 1986, p. 18-25; Ford, 1952).

It should be noted that in colonial times education was primarily for the few, the wealthy and those with facility for learning, and the primary purpose for reading was religious. Once public school education became established by law, however, and as concern grew for creating an educated electorate, the situation began to change. As Klare and Buck (1954) have noted



Saving Everyman's child from illiteracy was a different job from teaching the sons of merchants to read the Scriptures. It required different tools. (p. 40)

Klare and Buck (1954, p. 41) observe that, when compared with texts previously offered to children, the basic differences which appeared in the books of McGuffey and his contemporaries were their secular content and the fact that they were "graded" in vocabulary and reading difficulty. It appears that these authors were developing a concept of readability similar to that generally used today. They believed readability could be consciously controlled, and several decades before any scientific investigations of readability were begun, they were already identifying and manipulating factors which they felt affected it.

McGuffey and his contemporaries seemed to view vocabulary as the primary determinant of reading difficulty, for as Spache and Spache (1977) have observed

This author (McGuffey) controlled the difficulty of his books, he believed, by the length of words in the stories. The opening book used only two- or three-letter words and longer words were gradually introduced in later books. (p. 42)

Klare (1963, p. 30) notes that this relationship between vocabulary and reading difficulty seems to have been generally agreed upon during this period with much early work focusing on it, and Chall (1958, p. 17) contends that vocabulary has probably always been associated with reading difficulty.



Interest in the relationship between vocabulary and reading difficulty eventually led to the publication in 1921 of The Teacher's Word Book by E. L. Thorndike. This work, which listed words with tabulations of their frequencies in print, was intended to provide estimates of the commonness of words and therefore their relative importance. The list would influence the teaching of vocabulary in schools for generations and would also be a significant event in readability development since it would be used as the basis for many later readability formulae.

Klare (1963, p. 32) cites two additional events for their significant contribution to the development of modern readability theory. One was the formation in 1935 of the Sub-committee on Readable Books of the Commission on the Library and Adult Education. This committee consolidated the efforts of scattered individuals and gave recognition to the problem of readability in general. The second event was the publication, during that same year, of W. A. McCall and Lelah Mae Crabbs' Standard Test Lessons in Reading. This set of graded reading passages would later become the most often used criteria for the construction of readability formulae (Klare, 1984, p. 685).

Development of Readability Formulae

Early in this century interest in readability mounted dramatically. Literacy had become commonplace and the purposes for reading had expanded beyond religion into

information and pleasure. Readability was no longer simply a matter of importance to educators. Publishers of newspapers, magazines and best-sellers, and authors of government bulletins, industrial communications and military manuals were forced to write for a much larger group of readers more diverse in their reading abilities. At the same time as the need to understand and control readability was expanding, improved statistical procedures gave researchers better tools with which to work, and some of these methods, especially those in factor analysis and multiple correlation techniques, were particularly suited to readability study.

By 1920 researchers were conducting systematic and scientific investigations of readability, and Chall (1958, p. 17) credits Bertha A. Lively and S. L. Pressey in 1923, with developing the first procedure which approached the modern concept of a formula. Their work as well as that in other early studies generated much enthusiasm, motivated other researchers, and eventually led to the development of a host of formulae and other techniques which claim to predict the reading difficulty of a passage. This abundance of measures in turn produced an even greater proliferation of literature regarding the validity of such devices and controversies surrounding their use. To review all of the studies on readability would be a formidable task. Fortunately, two notable authors, Jean Chall and George Klare have provided comprehensive reviews of the most

significant early research concerning formula development and validity. Chall's books, Readability, An Appraisal of Research and Application (1963) and Klare's book, The Measurement of Readability (1963) are cited in nearly every article, book or dissertation concerning readability. They have become virtual classics in the field.

Exactly how many readability formulae have been developed is somewhat controversial. As Klare (1963, p. 33) has explained, the term formula has been used loosely to include both true formula based on regression equations as well as other devices for measuring readability. For this reason authors have defined the term differently and have therefore reported varying numbers of formulae as having been developed. No matter what definition is used, however, the number seems more than substantial. Chall (1958), for instance, tallied 29 up to 1954, Klare (1963) estimated 39, while one of Klare's students, Carolyn Dunlap (1954), listed 56 (Klare, 1963).

Historic Trends The general trend in formula development has been first one toward greater and greater complexity and then a sharp reversal toward increasing efficiency and simplicity. Chall (1958, p. 27) credits Irving Lorge (1939) with beginning this trend of simplification in 1938, while Klare (1963, p. 37-80) notes the same pattern but distinguished four historical periods. "Early Formulas", 1921-1934, used vocabulary primary as the predicting

factor and there was great dependency on Thorndike's Teacher's Word Book (1921). The criteria used was relatively crude. The next period, "Detailed Formulas", 1935-1938, saw an ever increasing tendency to use more and different predicting factors with less emphasis on Thorndike's work. There was also an increased concern for adequate criterion. The following period, 1938-1953, is termed "Efficient Formulas" by Klare, since the emphasis shifted during that time to increased efficiency and simplicity of use. The period from 1953-1959, a period following the publication of Chall's work, Klare labels "Specialized Formulas", since the tendency was to develop formulas based on a particular aspect of readability or a special audience level rather than wide applicability. Forbes and Cottle's formula (1953), for instance, was designed for use with psychological tests, while Bloomer (1959) was interested in measuring "the level of abstraction as a function of modifier load" and Spache (1953), Stone (1957) and Wheeler and Smith (1954) all authored formulae intended specifically for materials at the early elementary grade levels.

Methodology While individual formulae have varied, both Chall (1958) and Klare (1963) agree that the basic methodology by which most have been developed has been virtually the same, and generally proceeds according to the following steps: (a) A list of possible elements which could



be responsible for differences in readability is compiled. This list is usually based on some survey of reader and/or expert opinion or some analysis of content. (b) A set of criterion passages, representing a range of difficulty, is selected or developed. Methods used to establish the relative difficulty of the passages have varied and include the results of comprehension tests, ratings by readers or experts, publishers grade level recommendations and even other readability formula scores. (c) Once the relative difficulties are established, counts are made of the frequencies with which the identified elements occur in the criterion passages. (d) The frequency counts are correlated with the difficulty indices of the criterion materials. (e) The correlational information is combined in a regression equation which ultimately becomes the final formula.

While differences have occurred in the criterion used and the factors studied, most formula have used the correlational method and virtually all have followed the same developmental procedure.

Limitations of Readability Formulae

Chall (1958, p. 34-56) distinguished the following five components of readability formulae which are useful for evaluation and comparison: (a) The criterion on which the formula is based, (b) the range of difficulty of the criterion materials, (c) the method used for determining that difficulty, (d) the internal factors studies, and (e)

the method used to compare the occurrences of the factors studied with the difficulty indexes of the criterion materials.

While a few early formulae used an inspection method to compare the occurrences of the factors studied with the difficulty indexes of the criterion materials, generally all others have used the correlational method. Aside from this, however, formulae have differed greatly in the criterion used and in the factors studied. Both areas have posed serious limitations for readability prediction.

Limitations in Factors Studied By their very nature readability formulae must be based on an extremely limited set of factors which can affect reading difficulty. Most restrictive is the fact that they can only utilize those aspects of writing which can be quantitatively measured, and generally only stylistic factors have lent themselves to that kind of analysis. While some formulae have attempted to include content factors such as abstractness of words or analysis of ideas, Klare points out that they only touch on content in a very indirect way. Chall (1958, p. 12) and Klare (1963, p. 24) both caution however, that content, an aspect of writing that is difficult to measure quantitatively, is frequently thought to be as important as, or even more important than style, in determining the ease with which a selection can be read and understood. In fact, a classic study by Gray and Leary, reported in their book

What Makes a Book Readable in 1935, found content to be judged most important by their sample of librarians, publishers, teachers, and adult education directors. Style was ranked a close second, format a distant third and general factors of organization judged least important in determining passage difficulty. Klare (1963, p. 24-25) further warns that only one aspect of style, that of difficulty, can be measured by the formula. Formulae cannot measure the effectiveness or quality of that style, and moreover, they cannot even measure style difficulty with perfection.

In addition to being limited to measurement only of style, the style elements included in the formulae are also very restricted. While hundreds of factors of style have been studied, ultimately only two, some measure of word difficulty and some measure of sentence complexity, have emerged as being significant enough, or measurable enough, to become common elements in most formulae. Chall (1958, p. 54) explains that although other factors have been found to be significantly related to the criterion, they are also highly related to other factors in the formula and consequently add little by themselves to the final prediction. Their contribution is so meager that it is not worthwhile to include them. "The law of diminishing returns," she notes, "sets in early in readability prediction."



Counts of words which appear (or do not appear) on various word lists of presumably "easy" or "hard" words has been a favorite means of assessing the vocabulary element.

The general premise has been that the frequency with which a word appears in print, or its "commonness" is related to its difficulty. Thorndike's list (1921) has often been used for this purpose and was especially popular with early formula authors. Early authors also assessed vocabulary difficulty through some count of the number of different words in a selection. This method has sometimes been termed "word range" or "vocabulary diversity". Determining either frequency or diversity, however, required cumbersome, time consuming word counts, especially difficult to make in a pre-computer age. In attempting to find faster and easier to use methods, Lewernz (1922) and later Dale and Tyler (1934) used words beginning with certain letters. W, h, and b words were considered easy while words beginning with e and i were considered hard. Eventually the number of affixes, number of syllables in words and word length in terms of the number of letters were all found to be highly related to the commonness of words also. Since these factors provided simpler, faster and more reliable means for assessing vocabulary difficulty, they would become increasingly common elements in later formulae.

While the very first readability measures reported by Chall (1958) concentrated primarily on vocabulary factors, measurements of sentence complexity were soon being

incorporated. Sentence structure and the number of clauses were generally considered related to sentence difficulty. These in turn, however, were found to be highly related to certain types of words. Counts of the number of prepositions and prepositional phrases, and content words (nouns and verbs) have all appeared in various formulae. Ultimately, however, there is an obvious relationship between these factors and sentence length in general. Since length is a factor which can be counted simply and reliably, it became a common element in later formulae, usually in terms of average number of words per sentence.

Limitations in Criterion While at first it appears that formulae have varied widely in the internal factors studied, in reality the factors involved are all highly interrelated. All formulae have included some measure of vocabulary, most have included some measure of sentence complexity, and few have included much more than that. Therefore, in respect to the factors studied, the differences between formulae tend to be in the methods used to measure the factors and not in the factors themselves. Real differences have occurred, however, in the criterion used to construct various formula, the range of difficulty it represents and the method used to establish that difficulty. These differences in criterion are particularly important since they greatly limit the generalizability of any one formula, for as Chall (1958) has noted



Judged by strict scientific standard, each of the formulas is applicable only to material similar to the criterion on which it is based. Too often this is forgotten.....This has led to criticism of the formulas when actually the fault lay in their application to a type of material for which they were not designed. (p. 35)

Some criterion materials have been highly specialized. Ojemann (1934), for instance, used only parent education materials and Dale and Tyler (1934) used health brochures. Some authors have used general adult selections while others have concentrated on children's literature or textbooks used at particular grades. The McCall-Grabbs passages (1925) have been popular with several authors including two of the most well known formulae, the Dale-Chall (1948) and Flesch (1948).

The range of difficulty used for the criterion has also varied. In some formulae the range has included grades primer to adult, while others have been confined only to adult materials or a limited number of grade levels such as primer through third.

The methods of establishing the difficulty of the passages have also been diverse. Some authors have used various measures of comprehension of the test while others have favored more informal means such as ratings based on "expert" judgment. The grade level recommendations of publishers and later even other readability formulae have also been used.

Initially it would appear that tested comprehension on the material is the best possible method for establishing

the relative difficulty of criterion passages, since ultimately this is what the user of a formula wants to predict. Using comprehension scores for this purpose, however, has presented particular problems. As Chall (1958) notes

The major weakness.....lies in the fact that the difficulty of the passage can be changed by the ease or difficulty of the question asked.....Easy questions based on hard passages will result in underestimates of passage difficulty. (p. 40)

A study by Irving Lorge (1949) emphasized this point. Lorge applied the Gray-Learly formula (1935) to both the McCall-Crabbs passages (1925) and the questions on the passages. He found the correlation coefficient to be .6156, suggesting "there are factors in the passage which are unrelated to factors in the structure of the questions."

Determinations of passage difficulty become even more controversial when they go beyond providing indexes of relative difficulty to providing a grade level score. The later implies not only a comparison of the passages with each other, but a comparison of the passages to the performance of readers of a given ability. Typically determinations of that ability have been derived in an indirect manner, often using the standardized test scores of persons who have performed in a prescribed way on the criterion passages. Ojemann (1934), for instance, in developing his criterion, used the reading grade equivalent on a standardized test of the readers who were able to

answer correctly 50% of the comprehension questions on the criterion passages. Washburne and Vogel (1925) used children's books and the median of the score on the Stanford Paragraph Meaning Test of children who "read and liked" the book. Later authors, such as Spache (1953), however, simply accepted the grade level placement of materials as recommended by the publisher.

Validity data on grade level indexes when compared with external criteria and other formulae has been contradictory, leading Chall (1958) to conclude

...it is questionable whether the grade placement arrived at by the application of any one of these formulas can be used to make a definitive statement about the suitability of a particular piece of reading matter for a specific level of reading ability, even if only in terms of expressional difficulty. (p. 96)

Klare, (1963) has reached a similar conclusion and writes

The various formulas do not necessarily give comparable grade-level results even though they frequently show high intercorrelations. This indicates that attempting to place materials within a grade level by means of formula score is certainly questionable. (p. 120)

Validity of Readability Formulas

Nearly all readability validity studies can be classified using the following five categories: (a) Original presentation of the readability method, (b) original criterion prediction, (c) correlation with other

readability formulas, (d) experimental validation studies and (e) validation against outside criteria.

Generally Chall (1958) and Klare (1963) both reviewed studies in all of these categories but their classification systems and category titles differed somewhat.

Original Presentation of the Readability Method Studies in this category involve evaluation of a formula's validity based on logical grounds or on the evidence provided by its author. As Chall (1958, p. 70) points out some investigators merely assume the validity of their techniques, however two-thirds of the authors she studied provided some kind of empirical evidence such as correlation with test scores, basal reading series, or other formulae. On logical grounds the validity of a formula can be assessed based on such considerations as the way it was developed, the materials that were used, or the factors involved. For instance, a formula based on factors which have previously shown a strong relationship to reading difficulty would be considered more valid than one based on factors for which a relationship to reading ease has not been determined.

Original Criterion Prediction Studies in this category consider how successfully a formula will predict the scores of the original criterion passages from which it was developed. Klare (1963, p. 111) has characterized this as being "almost analogous to pulling oneself up by the

bootstraps", and cautions, that while it is an important consideration, it is not sufficient by itself. Particular factors used in a formula are usually selected because they are the most highly related to the criterion.

Klare (1963, p. 113) found that in original criterion prediction, the criterion coefficient for most recent formulas in 1963 was about .70. He explains that this in turn means roughly one-half of the variance in readability can be accounted for by the formula, a level of validity somewhat higher than the relationship usually found between psychological test scores and college course grades. Thus, he concludes, "these readability formulas can be considered of relatively high validity in a general sense".

Correlation with Other Readability Formulae Studies in this category examine the amount of agreement which exists between and among formulae. The assumption is that if readability formulae are all measuring the same thing then there should be a great deal of agreement in their results. Although a large number of comparative studies have been done, Klare (1963, p. 119) found the data difficult to interpret for the following reasons: Different investigators have used different materials and different formulae; some formula have yielded grade level scores while others have required corrections; different studies have used different criteria, with some studies based on the level at which 50% of the questions on a given passage could

be answered, while others have used 75%, etc.; and some studies have used a rank order correlation while others have used product-moment correlations.

In spite of the numerous disagreements in the data, the following are among the conclusions Klare (1963, p. 120) felt could be justifiably drawn from the comparative studies he examined: (a) The Dale-Chall (1948) and Flesch Reading Ease formulae (1948) have provided the most consistently comparable results in terms of both correlational and grade-placement data, (b) more of the high intercorrelations have involved Dale-Chall scores than those of any other formula and, (c) the various formulae do not necessarily give comparable grade-level results even though they frequently show high intercorrelations.

Earlier Chall (1958, p. 96) had also concluded that "at all ranges of difficulty the Flesch and Dale-Chall formulas tend to assign similar grade-levels". She also noted inconsistencies in grade level designations and expressed a need for additional comparative studies in specific subject area fields in order to interpret the meaning of the grade placements of one formula in relation to those of another.

These conclusions by Chall and Klare probably had considerable impact on later formula development, since they led to a generalized belief that the Dale-Chall is the "best" formula. Many later formula authors would justify the validity of their device by how well it correlated with the Dale-Chall.

Experimental Validation Studies Studies in this category involve rewriting material in easier and harder versions. These versions are then read by groups of readers presumed to be equivalent in reading ability. Most often comprehension (or learning, or retention) has been used as the criterion in such studies, although readership and reading speed (or efficiency) have also appeared.

Even though experimental studies offer the best opportunities for controlling variables, results from those using comprehension and readership criteria have been contradictory (Chall, 1958, p. 111; Klare, 1963, p. 133). Among other factors, Chall (p. 111, 112) speculates that the differences in effect may be related to the magnitude of the difference in readability between the two versions. Using a version that is greatly simplified is more likely to show a difference in comprehension than one that is only slightly easier. Moreover the relationship of difficulty of the passages to the reader's ability may be very important. If both versions are above or below the reader's ability, there may be little difference in comprehension, but if the original is beyond the reader's ability and simplifying brings the difficulty of the passage down to the reader's level, the effects might be considerable. Differences may also depend on the importance of the factor being manipulated. Vocabulary and sentence length, for instance, have shown more effect than human interest factors (Allen, 1952). The number of factors involved may also be important

since later studies using multifactor formulae tended to show more positive results than earlier studies based on vocabulary changes only.

Reading speed as a criteria appears relatively late in readability research studies, with Rudolf Flesch (1949) being credited as the first to suggest a relationship between the two (Klare, 1963, p. 135). Of seven studies reported by Klare (1963, p. 137), all were judged positive in results when speed was measured in terms of words per minute. Klare (1963, p. 137) concluded that "the general results indicate clearly that readability and reading speed are related. This measure appears to be both a sensitive and consistent criterion." More recent studies (Miller and Coleman, 1971; Coke, 1974), however, have indicated that reading rate for both oral and silent reading, remains constant over a wide range of readability, when rate is measured in units smaller than words per minute, namely syllables per minute. The word rate in these studies decreased with passage difficulty, but the syllable rate remained constant. Coke (1974) explains

Since subjects read at a constant syllable rate, words containing more syllables took longer to read than words with fewer syllables. Therefore, the harder passages, which had a larger proportion of longer words, were read more slowly when rate was measured in word units. (p. 407)

Coke cautions that "the almost universal practice of measuring rate in words can lead to spurious conclusions about the relationship between reading rate and

readability."

Validation Against Outside Criteria Studies in this category involve comparing formula results with results obtained from other sources. Judgments and reading performance have been the most common types of outside criteria used.

In judgment studies, readers, librarians, teachers, publishers or other experts are asked to rank or assign grade level designations to the research passages. These are then compared with the formula results. In studies reviewed by Klare (1963, p. 155), 12 of those using judgments showed positive results, 2 were negative and 3 were considered indeterminate. Klare concluded that more readable material as measured by formulae can be judged more readable by experts and readers.

Comprehension, also referred to as learning or retention, was one of the first and most frequently used criteria for comparing formula results with reading performance. However, most of the studies using comprehension criteria were of the experimental type described earlier. Klare (1963, p. 133) reviewed five studies, however, which gave some indication of the formula's ability to predict performance for a particular grade level. These studies by Stadtlander (1938), Miller (1946), Latimer (1948), Dunlap (1954) and Peterson (1956) generally involved estimating the subject's reading ability

either by test scores or their grade placement, administering comprehension questions to them after they had read the research passages and then comparing the results with the formula scores of the passages. All of these studies were considered positive in results although many questions have been posed concerning their validity.

Oral Reading Criteria in Readability Research It should be noted that when reading performance has been used as the criteria for constructing or validating readability formulae, the type of performance used has been almost exclusively silent reading comprehension, the only exceptions being those few studies using reading speed. There is no evidence of oral reading as a criteria being used in either formula development or in validation studies (Klare, 1963; Klare, 1984, p.688-699). This is probably not accidental. At the time early readability theory was developing, oral reading, either as a testing or teaching procedure, was very much out of favor in education (Allington, 1986, p. 830-831, 835; Smith, 1986, p. 158-195). This helps to explain the early researcher's preoccupation with readability only as it relates to silent reading. Oral Reading as an evaluation tool, did not appear until the late 1940's (Durrell, 1937; Betts, 1946) and no doubt did not become well established in practice until much later. Likewise, formulae specifically intended for primary level materials (Spache, 1953; Stone, 1957; Wheeler and Smith,

1954) did not appear until much later in the development of readability prediction. Even then, the earlier traditions seem to have prevailed, for the use of oral reading, either in the development of formulae or in studies of their validity, has been almost totally ignored even to the present day.

Development of the Fry Graph

By the time Klare's book was published in 1963, research in readability had fairly well run its course. There seemed to be very little additional information that a new formula could add to the existing body of knowledge of readability, nor did there seem to be much need for developing another readability measuring device considering the abundance of procedures already available. Yet in 1968, Edward Fry, a professor at Rutgers University, published still one more method for predicting the ease with which a selection could be read and understood. Surprisingly, it would become one of the most popular methods ever developed.

Kistulentz (1967) found Fry's procedure to show high correlations with other readability methods. Its added appeal, however, lies in its simplicity and ease of use and it was on this basis that Fry (1968) justified its publication. It is not a formula based on a regression equation as such, but rather it utilizes a nomograph. It, therefore, does not require the user to make any mathematical calculations, a definite advantage in the days before inexpensive calculators were commonplace. Instead

the user simply makes counts of two variables, one of word length and one of sentence length, and then plots these on the Fry Graph. Fry published the graph with a specific statement that it was not copyrighted, thus assuring its continual easy and widespread availability.

Simplicity is not only evident to the user of Fry's procedure but also appears to be a keynote in its development. The author chose to capitalize on previous research which suggested that only two factors, word difficulty and sentence complexity would consistently emerge as the most significant elements in prediction of readability. He uses a count of the total number of syllables and a count of the total number of sentences (estimated to the nearest tenth) in one hundred word samples as measures of these factors.

Fry (1969) cites research by Stolurow and Newman (1959) and Brinton and Danielson (1958) as support for his use of sentence length as the measure of syntactic complexity and a syllable count as the measure of word difficulty. The former study found a high correlation (.90) between reading difficulty and polysyllables, difficult words and the percentage of different difficult words, a high correlation (.90) between reading ease and easy words and monosyllables, and a relatively high correlation (.86) between average sentence length and difficulty. The researchers concluded that "any yardstick which gave primary weight to the so-called word factor and a lesser but almost equal weight to

the sentence factor would account for a good deal of the variance in readability." Similarly, after studying twenty language elements, Brinton and Danielson concluded that their investigation "confirms the importance of word length and sentence length" in readability measurement.

Some researchers (Bormuth, 1966, 1969, 1975; Stolurow and Newman, 1979, p. 250) have suggested that a curvilinear relationship may exist among readability factors with sentence length having a greater affect on readability at lower reading levels and word difficulty being more significant in upper grades. The values on the Fry graph support this contention (Entin and Klare, 1979, p. 288). Because of the way the graph was constructed, it may automatically take such a curvilinear relationship into account.

When the syllable count and sentence count form a sample are plotted on the Fry graph, they fall into various areas which indicate the difficulty of the passage in terms of a grade level score. Fry (1968) explains how he arrived at these scores as follows:

Grade level designations were determined by simply plotting lots of books which publishers said were 3rd grade readers, 5th grade readers, etc. I then looked for clusters and "smoothed the curve." After some use of correlational studies the grade level areas were adjusted. (p. 515)

There are indications in the literature that the Fry method assigns higher grade levels to primary level materials than other methods do, leading some to regard his

formula as being too easy. Harris and Jacobson (1980), for instance, measured numerous samples from the Economy (1980) and Houghton Mifflin (1981) reading series for levels pre-primer through third grade using the Fry graph (1968), the Spache formula (1974) and a computer version of the Harris-Jacobson formula (1974). A comparison of the results found that the Fry method assigned much higher grade levels to materials beyond the second grade than did either of the other two formulae or the publisher's designations. This led Harris and Jacobson to conclude that the Fry graph seriously overestimates the difficulty of second and third grade reading materials (Harris and Jacobson, 1980).

Fry (1980) has not considered this overestimation a serious problem since it would mean assignment of easier to read books, a better alternative than assigning books that would be too difficult for a reader. Fry also contends that the differences between his results and those from other procedures are not as great as the Harris and Jacobson study indicates. He cites studies by Britton and Lumpkin (1977), who found good correlations and good grade level agreement among publisher's designations and Fry, Spache and Harris-Jacobson results. He also notes a second study by Fox (1979), who used the Fry measure on "almost every basal reader in America being sold in 1978". Fry (1980) reports that the Fox study found his formula "correlates quite well at grades 1 and 2 but that it is a little high in third grade, but not as far off as Harris and Jacobson found in

the average of the two series that they analyzed".

Recent Trends in Readability Prediction

In spite of their inherent weaknesses, and contradictory evidence as to their general usefulness, methods for estimating readability continue to receive a great deal of attention from publishers and educators today. Most recently published pedagogical texts, especially those intended for secondary and content area fields, at least mention readability, if only to warn readers of the limitations involved, and many such books devote entire chapters to the subject, along with descriptions and instructions for using several readability techniques (Singer and Donlan, 1980; Vacca, 1981, 1986; Criscoe and Gee, 1984). Articles on readability also continue to appear in professional journals and new formulae continue to be developed regularly.

Two trends are noticeable in most recently published readability literature. First the need for increasing ease and speed in making readability estimates still continues to be a primary consideration. Secondly, there appears to be a trend toward finding more valid methods for developing criterion passages. These include methods which do not ask comprehension questions and which thereby circumvent the problems of relative difficulty inherent in the questions themselves, and also the use of specialized criteria for formulae to be used in specialized areas. In addition,

Klare (1984), in a recent review of current readability research, identifies two other trends: (a) Use of new approaches and (b) work in languages other than English. Of these, only the new approaches seem to be pertinent to this study. Most of these new approaches, however, are closely related to the trends previously mentioned since they are being developed, it appears, because of dissatisfaction with the ease of use and the adequacy of criteria in existing formulae. They will be included, therefore, as a part of the following discussion which will concentrate on efforts to improve ease and speed of use and to develop better criterion procedures.

Ease and Speed of Use Most readability prediction procedures currently popular, are similar to the Fry method, but claim to improve ease and speed of use even further, primarily through easier counting of factors or the use of graphs, charts and manual or machine aids. The Flesch procedure (Flesch, 1949), for instance, uses a scale instead of a graph. The Raygor method (Raygor, 1977) employs a graph similar to the Fry, but utilizes counts of long words (six letters or more) rather than syllables. The SMOG method (McLaughlin, 1969) involves counting only words of three or more syllables, estimating the square root of this number and adding a constant of 3. No graph is necessary.

In addition to these modifications of current methods, the quest for increasing ease and speed in calculating

readability has taken at least two new directions: (a) The appearance of several new subjective, rather than objective methods, and (b) the publication of numerous computer based formulae.

The trend toward developing faster, easier to use subjective techniques, rather than objective measures, is noticeable in the guidelines offered by Vacca (1986, p. 40-41), in the Irwin and Davis checklist (1980), and in the attention that has been given recently to the SEER technique (Singer Eyeball Estimate of Readability, Singer, 1975) and the Rauding Scale (Carver, 1974, 1975, 1976). These latter two techniques are based on the ratings of trained judges. They involve taking a passage of unknown readability level and comparing it with a set of scaled passages, the reading levels of which have already been determined. While at first such subjective techniques may appear to have little merit, their authors claim they are faster and just as accurate as other procedures. The positive findings of early studies using judgmental criteria reported by Klare (1963), would suggest that human raters can be as accurate as mechanical devices, however, more recently, Klare (1984, p. 702) has questioned the purported speed advantages of judgmental techniques, especially when the time to train and qualify the raters is considered.

While some researchers have been pursuing subjective techniques, others have concentrated on the development of computer based formulae, and as inexpensive, desktop

computers became available, they quickly capitalized on this tool as a means for making readability prediction more efficient. A host of readability estimation programs, both commercial and non-commercial, have recently appeared (Danielson and Bryan, 1963; Carlson, 1980; Gerbens, 1978; Goodman and Schwab, 1980; Irving and Arnold, 1979; Keller, 1982; Schuyler, 1982). Some of these programs use new formulae, while others are simply automated versions of present, well known readability procedures. Even the new methods, however, still seem to be based on measures of word and sentence difficulty similar to that used in existing formulae. For this reason Geoffrion and Geoffrion (1983), in reviewing such programs, have concluded

Existing computerized readability software make inefficient use of a computer's capabilities.

They note that current formulae

are limited to measures such as sentence length and syllable counts because these are easy for human evaluators to judge quickly. More complex aspects of a passage are ignored because they are too tedious for rapid manual calculation. Yet the computer's speed and accuracy make feasible much more complex calculations. (p. 104-105)

As yet there seems to be little research aimed at producing computer programs designed to take full advantage of the computer's capabilities in producing more accurate readability estimates, although the authors do note work in this direction underway at Bell Laboratories (Geoffrion and Geoffrion, 1983, p. 105). Klare (1984), however, questions

whether or not more complex formulae can really add any greater precision. In a study by Bormuth (1969), "unrestricted" formulae with up to 20 variables gained slightly in predictive power over simpler formulae in the validation process but dropped considerably in cross-validation. Klare (1984) concludes that

This yielded an unexpected answer to those who felt that the availability of computers would lead to more complex and, therefore, necessarily more powerful predictors. (p. 687)

Geoffrion and Geoffrion (1983, p. 105-106) also caution that practical problems greatly limit the usefulness of presently available readability measurement software. They note that current formula programs lack a convenient means for entering text samples and the less sophisticated ones also lack an easy way to correct typing errors. The MECCA program used in this study is typical. Once text is entered, this program can give readability estimates based on several popular procedures as well as syllable counts, word and sentence length and other such information. Text, however, must be entered line by line. Corrections can be made by backspacing only within the line currently being typed for entry. Once the line is entered, corrections can only be made by calling up another program for text editing. Correcting even the simplest one letter error, entails designating which line is to be changed, indicating the type of change desired (add a line, delete a line or edit a line), and then actually replacing the line by retyping and

reentering it. This is a tedious and time consuming process.

Computers capable of recognizing written text in a variety of print fonts, styles and layouts have already been developed for use by the visually impaired in oral reading machines. While this capability is still too expensive for general use, Geoffrion and Geoffrion (1983, p. 106) see it as having great potential for future readability programs. The development of computers that can recognize human speech and voice commands also holds promise for facilitating text entry. Until such machines are available, however, computerized readability measurement is not as easy as it may first appear.

Criterion Development Accompanying recent efforts to increase ease and speed of use in readability prediction are developments aimed at improving the criteria on which the procedures are based. Klare (1984, p. 691) notes the following trends in this regard: (a) Improvements in existing criteria primarily through renorming of the McCall-Crabbs passages (Harris and Jacobson, 1976; Jacobson, Kirkland and Selden, 1978), (b) specialized criteria for use in special areas (Caylor, Sticht, Fox and Ford, 1973; Kincaid, Fishburne, Rogers and Chissom, 1975), and (c) use of the Cloze procedure in criteria development (Coleman, 1965; Miller and Coleman, 1967; Bormuth, 1969, 1975). Of these the latter is particularly significant and will

therefore be examined in more depth.

In order to understand the role of the Cloze procedure as it relates to readability, it is necessary to understand the distinction between readability prediction as opposed to readability measurement. As Klare (1984, p. 701) and Vacca (1986, p. 53) point out, formulae are predictive techniques. They hypothesize about text difficulty based on an analysis using selected variables that have been statistically found to correlate with comprehension difficulty. The reader is not a variable. In contrast the Cloze technique, like oral reading assessment, is a readability measurement procedure. It measures readability by using actual reader performance in the material without making any predictions concerning that reader's performance in any other material.

The Cloze technique (Taylor, 1953) involves the systematic deletion of words from a passage, usually every 5th or 7th word. The reader is then asked to fill in the blanks with the word they think appeared in the original text. Using criteria established by Bormuth (1966), identifying between 40% to 60% of the words correctly would indicate the passage is at the reader's instructional reading level. The Maze procedure (Guthrie, 1974) is similar to the Cloze test, except that it uses a multiple choice format instead of blanks, and, since it is an easier process, it uses more stringent criteria of 60% to 85%.

The Cloze and Maze procedures appear to have more validity than readability prediction procedures because

they, like oral reading assessment, use actual reader performance. However they also have some of the same drawbacks. They do not assign a readability index number or grade level to the passage as such, but simply indicate if the text in question is of suitable difficulty for the particular student or group of students involved. Every time a new text or a new group of students is encountered, the procedure must be repeated. Unlike oral reading assessment, however, which is a one to one process, the Cloze and Maze tests can measure readability for many readers at the same time.

Recently the cloze format has gained popularity as the means for establishing the relative difficulty of the criterion passages on which new readability formulae are based (Bormuth, 1969, 1975; Coleman, 1965; Miller and Coleman, 1967). In this manner it goes beyond readability measurement and becomes a prediction device.

In determining the difficulty of criterion passages, the Cloze procedure has a distinct advantage over traditional methods of assessing comprehension through questioning. It by-passes the problem of variations in difficulty inherent, not in the passages, but in the questions themselves. Because of this, it holds particular promise for formula authors, and Klare (1984, p. 687) has found the use of the Cloze measure to be one of the important new directions in criteria development.

Readability in the Early Elementary Grades Although several formulae have been designed for use at the lower grade levels (Spache, 1953; Stone, 1957; Wheeler and Smith, 1954), basal programs generally prevail in these grades and the demands of content area reading are far less than those at intermediate and secondary levels. Therefore, it is not surprising to find relatively less emphasis on readability theory and measurement in training materials intended for teachers at these levels. There are two notable exceptions however. One is in the preparation of materials for use in constructing Informal Reading Inventories (IRI) and the second is in the development of materials for use in individualized reading instruction.

An informal Reading Inventory estimates the level of a student's reading ability by listening to the student read in materials of increasing difficulty and noting the errors that are made. Although many commercially prepared inventories are available, the procedure is considered most valid when the paragraphs to be read are prepared by the teacher using materials the student will be reading in the classroom (Betts, 1946, p. 454; Ekwall, 1976, p. 271; Bader, 1980, p. 206). Authors usually suggest that the teacher use basal text passages for this purpose, but since basal materials can vary considerably in difficulty, many also suggest using a readability formula to check the difficulty level as well. Some, like Harris and Sipay (1980, p. 58), even give detailed, step by step instructions for doing so.

In addition to Informal Reading Inventories, Individualized Reading Instruction Programs have created a particular need for determining readability of materials for elementary students. Such programs typically do not use a basal text series through which all students in a class proceed together. Instead, such programs are usually centered around an extensive collection of children's literature and trade books. Since self selection of reading materials is usually emphasized, a large collection is necessary in order to provide for the varying interests and reading abilities of a classroom. A readability formula could be used to determine the difficulty of the selections. However, the great number of books involved usually makes using even the simplest procedure impractical. Therefore the problem of helping students find books they both want to read and are capable of reading has been a major problem in implementation of this type of program. As a result, advocates of individualized reading instruction have suggested some interesting solutions. One of the most popular is the "Rule of Thumb" (also called the Five Finger Exercise) developed by Jeannette Veatch (1966, 1978), whose leadership has been foremost in popularizing the concept of individualized reading instruction.

The "Rule of Thumb" is meant to be taught to children. They are then to use it by themselves to decide if a particular book is easy enough for them to read. When the student has found a book he wants to read, Veatch instructs

the teacher to say the following to the student: (Veatch, 1978)

Riffle the pages and stop on one page in the middle of the book. Start to read it to yourself. If you come to a word that you don't know, put your thumb on the table. If you come to another word you don't know, put down your first finger. Another unknown word, another finger, and so on. If you use up all your fingers, then the book is too hard for you. Put it down and find another.

If you find a book that has no unknown words, it is probably too easy for you. Save it for a free time, and choose another book to bring to me for your conference. (p. 55)

Veatch does not indicate how many words the student should read before deciding if the book is of suitable difficulty. Cunningham (1977, p. 191), however, modifies the procedure by having the student choose a paragraph of about 100 words. Five unknown words in 100, would then suggest 95% word recognition accuracy. This corresponds roughly to the criteria Betts (1946) established for determining a reader's instructional reading level.

The reliability of the Rule of Thumb method appears to be questionable. As Cunningham (1977, p. 191-192) warns, some children may not be able to handle this procedure, either because they cannot admit, even to themselves, that they do not know a word, or because they may be unaware that they have made an error. Validity data to support the practice is not offered and Veatch herself (1976, p. 55) calls the method a "rough measure, to be sure, but the only one in which the choice of material is the pupil's."

ARRF (Average Reader Readability Formula), proposed by Patricia Cunningham (1976), appears to be another rather crude method for making a quick assessment of the difficulty of a large number of books to be used for a particular class. It involves identifying a student whose reading ability is considered about "average" for the classroom in question. This student is then asked to spend a couple of hours with the teacher reading short passages from each of the books to be used in the program, and deciding if each selection is too easy, too hard or just about right for the average readers in this class, of whom this reader is supposedly typical. The books are then codified as being "easy", "average" or "hard". Below average readers in the room may choose books from the "easy" group, while above average readers may choose from any group.

The validity of a procedure such as ARRF is also obviously questionable. The method assumes that an "average" reader can be identified and that this reader's performance can be generalized to other readers. With the large amount of reading involved, variations in the reader's performance while classifying the books must also be considered. It is possible that books read later in the classification process may appear easier due to the practice effect or more difficult due to reader fatigue. Cunningham does not offer validity data.

Oral Reading in Readability Measurement and Prediction

Although oral reading is currently a commonly used method for measuring readability, it seems to have been virtually ignored in either the development of readability formulae or studies of their validity. As previously noted, oral reading, either for instructional or assessment purposes, was very much out of favor at the time that early readability research and formula development were occurring (Allington, 1985, p. 829-835); Smith, 1986, p. 158-195). It is not surprising to find, therefore, that the early readability researchers' focus was on silent reading comprehension, and oral reading as a criterion measure or in validation studies was apparently disregarded. That disregard seems to have persisted to the present time, for this literature search revealed no evidence of oral reading having ever been used in the development of any formula, and only two validity studies using oral reading could be found.

Fry (1969) used oral reading errors along with cloze procedure errors and Fry and Spache readability procedures to make rank order comparisons of the readability of seven selections. Fry found that all four methods ranked the difficulty level of the passages quite well, but the cloze procedure seemed to be the most accurate and made the finest distinctions. The oral reading scores were not as accurate or as fine grained as the cloze scores. However Fry views oral reading as an interesting method for judging readability and one not often used. It has the advantage of

being objective, independent and a different validation procedure. Fry (1969) notes that

Readability formulas are often validated on such non-objective criteria as subjective judgment or publishers' recommendations. Or they are validated by comparing them with other formulas. These methods are not wrong, but we must continually keep in mind the real basis for readability is whether a child can read the material. Therefore, validity measures that use children should receive high priority. For this reason cloze and oral reading errors should be used increasingly in research to validate readability formulas, although the time factor limits their use as practical methods of determining readability. (p. 536)

In another study, Paolo (1977) compared Fry readability scores and reading errors in "Easy-to-Read" trade books for children. She found that eight of the ten books she studied were at frustration level for her first and second grade subjects, and that a positive and significant correlation (.78) existed between the oral reading errors and the readability scores. Paolo's study was well designed. However, it involved only five subjects, which seriously limits the impact of its findings.

Part II

Determining Reading Ability

Introduction

While the concept of readability was being developed in the late eighteenth and early nineteenth centuries, methods for measuring the readers' ability were also under investigation. Ultimately these efforts lead first to the development and widespread use of standardized tests, and later to the appearance of the Informal Reading Inventory and oral reading assessment procedures. Each of these methods will be examined in turn.

Standardized Tests

Standardized tests are probably the most frequently used means of measuring student reading performance. They are designed to be administered and scored in a uniform manner so that any variation in test scores can be attributed to differences in the students taking the test and not to the conditions of testing. Generally such tests are put out by major publishing companies with items written by professional test specialists and revised through many try outs and item analysis. This process has resulted in tests of exceptionally high reliability. The general range is from .80 to .95 and standardized tests with reliabilities over .90 are not uncommon (Borg and Gall, 1979, p. 218).

Scores on standardized tests are generally based on relative performance. An individual score has meaning only in relationship to the scores obtained by others who have taken the same test. Norms, or scores which indicate "average" or "normal" performance, are developed by administering the test to a standardization sample. The sample itself is chosen from persons who are representative of students for whom the test is intended. Usually this sample is large, with 1000 or more subjects. Thus developing the norms for a standardized test is an expensive procedure.

Publishers of standardized tests typically supply detailed information concerning the norming procedures used and descriptions of the social, educational, economic, ethnic and racial characteristics of the standardization sample. This information is important to persons using the test since the test has its greatest validity for students with backgrounds most similar to those of the persons in the sample. Various tables and instructions for converting and interpreting raw score data are also included.

Standardized tests can be individual measures and can assess achievement or abilities by sampling many behaviors in diverse ways. In practice, however, the standardized tests of reading achievement being used today are almost exclusively objective tests of silent reading comprehension designed for group administration. Johnston (1984) notes that this group focus and silent reading emphasis are not

accidental, but are rather related to the historical climate prevalent during the development of such testing procedures. Standardized reading tests are a direct outgrowth of the turn of the century psychological testing movement in general. Johnston identifies two driving forces of this movement. The first was the intention of making psychology worthy of the term science which seemed to indicate quantification and "objectivity". The second was the press for educational accountability that accompanied a dramatic rise in school enrollments brought on by immigration and population growth, child labor and compulsory education laws and increased literacy expectations in society. These forces, along with the almost universal emphasis on silent reading during the period produced a climate in which Johnston (1984) concludes only a certain kind of test could survive.

Thus, while diverse approaches were developed initially, the fittest in terms of efficiency soon surfaced. Reading tests came to consist of the silent reading of a passage, followed by the solving of brief, generally text-related, problems; usually questions. (p. 149)

The efficiency of administering and scoring standardized tests, along with their high reliability have made them popular measuring devices in educational research and program evaluation. However, while these qualities of efficiency and reliability are generally accepted, the question of test validity, or the ability of such tests to measure what they claim to measure, along with

questions concerning their proper use, remain controversial.

Farr (1969, p. 85) lists two valid uses of standardized tests. First they are reliable for comparing students in terms of general reading achievement. Secondly, the tests are useful as screening devices in determining if further assessment through individual reading tests and informal testing procedures is needed.

The major weakness in using standardized tests seem to center around the tests inability to identify specific areas of reading strength or weakness, and in the use, or misuse, of grade equivalency scores.

While most standardized tests are made up of subtests such as "phonetic analysis", "vocabulary", and "comprehension", Farr finds that such tests are unable to measure distinct skills or abilities (Farr, 1969, p. 82). Because of this lack of discriminant validity, the tests are of little value in reading diagnosis or for planning specific instructional programs.

The inclusion of grade equivalency scores, which are provided by most test publishers in addition to other norming information such as percentile ranks and standard scores, raises further questions concerning the valid use of standardized testing. Grade equivalency scores are popular with both teachers and the general public because they provide an easy point of reference. However, the term "equivalent" is probably very misleading and such scores should be interpreted with great care.

The term "equivalent" implies that, regardless of their grade placement, students receiving the same grade equivalency scores, have comparable reading abilities. Glaser (1964), however, found that while a group of seventh graders and a group of third graders had the same scores on the Gate's Survey (between 5.0 and 5.9), their performances on an informal reading inventory differed considerably. He concluded that this was because the standardized test compared individual performances to that of other students, while the informal inventory compared individual performances to a set of criterion tasks.

Another problem with grade level norms lies in the between grade scores. Usually such scores are reported with a decimal. The number before the decimal indicates the grade level, while the number after the decimal indicates the month in that grade, with 0 to 9 standing for the months of September through June. No credit for progress is given for the months of July and August. Usually, in the course of being normed, a test is administered only once during a year. Between grade norms are interpolated from these "empirical norms". Using scores derived in this manner assumes that learning within a year proceeds at a uniform pace, however, studies by Bernard (1966), Lennon (1951) and Traxler (1950) suggest that this is not the case.

For these reasons, grade level scores are considered to have their greatest validity when time of year of testing corresponds as closely as possible to the time of norming.

In addition, scores considerably above or below the student's grade placement should be interpreted only as above or below the norm for that grade. The student should not be considered to have the same reading ability as the average student in the grade indicated by the grade level score.

Finally, several studies (Betts, 1940; Killgallon, 1942; Sipay, 1964; Glaser, 1964; McCracken, 1964 and Leibert, 1965) have compared standardized test results with results from Informal Reading Inventories. Generally these studies found that standardized tests gave higher grade level scores than the IRI, indicating that they cannot be used to place students in materials at their functional reading levels. Moreover, Farr (1966, p. 108) notes that most publishers of standardized tests do not suggest that the grade score norms be used as indicators of the levels at which reading instruction should be provided.

The California Achievement Tests The California Achievement Tests, 1977 edition, published by McGraw-Hill, have been a well known, widely used and highly regarded series of test batteries designed to measure achievement in basic skills. Ten different levels of the tests are available for children in grades kindergarten through twelve. The upper seven levels also have alternate forms for use in pre and post testing situations or when multiple administrations of a level are necessary. The test

is nationally normed and adheres to the standards of the American Psychological Association to assure that the standardization group is a representative national sample. Derived scores are provided in the form of percentile ranks, normal curve equivalents, stanine scores, grade equivalents and scale scores.

Unlike many standardized tests which are administered to the standardization sample only once during the norming process, the California Achievement Tests have had two administrations, and therefore empirical norms are available for both spring and fall testing. This improves the validity of the test results and is probably a major reason why the tests enjoy high regard.

The Reading Test of the California Achievement Tests consists of four subtests at Levels 13 and below: Phonic Analysis, Structural Analysis, Reading Vocabulary and Reading Comprehension. Beginning with Level 14, which is the test usually used with fourth grade students, there are only two subtests: Reading Vocabulary and Reading Comprehension. Thus the tests are heavily dependent on silent reading comprehension at the primary grades and become almost totally a test of silent reading comprehension at grade four and beyond.

While Grade Equivalency scores are provided, the publisher also includes cautions for interpreting these scores (California Achievement Tests-Norms Tables, 1977, p. 4). Among them they warn that (a) grade equivalents do not

mean that a student has mastered all of the objectives taught in the school district up to the grades corresponding to the grade equivalent score, (b) grade equivalents should not be used in placing students in school grades corresponding to the test score and, (c) because grade equivalent scores can be easily misinterpreted, it is strongly recommended by the publisher that they not be used in reporting a student's score to parents or other persons with little or no training in testing.

Oral Reading Assessment

In addition to standardized tests of reading achievement, procedures for assessing oral reading performance are frequently used to determine reading proficiency.

The first formal assessment of reading through observation of oral reading performance probably appeared in 1915 with the publication of Gray's Standardized Oral Reading Paragraphs (Allington, 1984, p. 835). These test passages, arranged in order of difficulty, were to be read aloud while the examiner recorded such errors as mispronunciations, omissions, additions, and repetitions. The test received very little attention at the time of its publication, however, probably because it coincided with widespread criticism of oral reading and vigorous expansion of silent reading practices in instruction brought on by expanding literacy, changing needs in society and research

reports (Piner, 1913; Thorndike, 1917; Judd and Buswell, 1922) which stressed the superiority of silent reading over oral reading in developing fluency and comprehension. These studies have been summarized by Huey (1908; 1968).

Development of Traditional Practices Moderation between oral and silent reading positions eventually led to renewed interest in oral reading. This interest was no doubt prompted by growing dissatisfaction with standardized testing, which offered no opportunity to observe reading behaviors directly. During the 1930's several authors developed descriptions of oral reading errors (Duffy and Durrell, 1935; Daw, 1938) and oral error classification systems (Payne, 1930; Monroe, 1932). It is Emmett Betts, however, who is generally credited with defining and popularizing the practices of oral reading assessment. The principles underlying the Informal Reading Inventory, and the concepts of independent (basal), instructional, frustrational and "capacity" reading levels, with the criteria for establishing them, were presented by Betts in one chapter of his book, Foundations of Reading Instruction, published in 1946. This work had a profound affect on the development of modern reading diagnostic theory and practice, and while the Betts' criteria is often challenged, it remains widely used and commonly accepted by practitioners today.

In determining placement of students in materials, Betts identified four levels of functioning for a reader in relationship to the readability of materials at various grade levels. The first level he called the basal level. This is generally referred to as the independent level today, since the basal level "approximates the level at which "free," supplementary, independent, or extensive reading can be done successfully" (Betts, 1946, p. 446).

The second level, the instructional level, is the place "where learning begins". It represents that level where the learner is "challenged but not frustrated" by the material (Betts, 1946, p. 447).

The third level, the frustration level, is "the lowest level of readability at which the pupil is unable to comprehend printed symbols to a reasonable degree....the individual is inadequate to deal with the reading matter" (Betts, 1946, p. 451).

Betts also identified a fourth level, the capacity level which is sometimes called the listening comprehension level today (Durrell, 1955; Kress & Johnson, 1965; Ekwall, 1976, 1979). Betts (1946, p. 452) describes this level as "the highest level of readability of material which the learner can comprehend when the material is read to him."

Betts (1946, p. 446) included the following in his criteria for establishing a reader's basal level: Accurate pronunciation of more than 99% of the words; freedom from tension and finger pointing; acceptable reading posture;



oral reading characterized by proper phrasing; accurate interpretation of punctuation; and use of conversational tone.

The criteria for oral reading at the instructional level (Betts, 1946, p. 449) included the following: Accurate pronunciation of 95% of the running words; ability to anticipate meaning; freedom from tension, finger pointing and head movement; and acceptable reading posture.

At the frustration level (Betts, 1946, p. 451), the criteria for oral reading included: Inability to pronounce ten percent or more of the running words; frequent or continuous finger pointing; distracting tension, such as frowning, blinking, excessive and erratic body movements; unwillingness to attempt the reading; attempts to distract the examiner's attention from the problem; word-by-word reading; failure to interpret punctuation; high-pitched voice; meaningless word substitution; repetition of words; insertion of words; partial and complete word reversals; omission of words; and practically no eye-voice span.

The criteria as presented by Betts contains several contradictions which have caused considerable controversy and variation in the way the criteria has been interpreted in practice. First, Betts established word recognition scores of 95% for the instructional level and 90% for the frustration level leaving a gap of 5 percentage points not designated as being at any level. Secondly, the way Betts originally presented the criteria left it unclear if silent

reading should precede the oral reading in an IRI, or if the oral reading should be at sight without the benefit of preparation. Finally, using the Betts' criteria is further complicated because, although Betts gave definite percentages for judging word recognition at each level, he did not clearly define what deviations from text should be considered in determining these percentages. He simply refers to "accurate pronunciation" of a given percent "of the running words" (Betts, 1946, p. 446, 449, 451). Consequently, what determines a mispronunciation has been left largely up to interpretation.

Most authors have dealt with the gap in percentages between levels by using other information gathered during the reading, to decide if a score falling in the range of 95% to 90% should be designated at the reader's instructional level or frustrational level.

The question of silent reading proceeding oral reading is a more serious one, since the number of errors made in oral reading falls dramatically when the reader is first allowed to prepare silently (Brecht, 1977).

The confusion concerning silent reading preparation seems to stem from the fact that Betts introduced the principles of the informal reading inventory, a form of testing, with the principles underlying a directed reading lesson, a form of instruction, simultaneously, in the same section of his text. In this regard, Betts (1946) writes

There is general agreement on one basic principle regarding directed reading instruction...namely, silent reading should precede oral reading. (p. 449) (Emphasis added)

A few pages later, in giving the principles underlying an Informal Reading Inventory, he states

In general, the procedure for the administration of an informal reading inventory for the systematic observation of performance in controlled reading situations is based on the principles governing a directed reading activity. (p. 456) (Emphasis added)

As one of these principles, he notes that "silent reading should precede oral reading", but a few lines later he writes

An exception to the principles basic to a directed reading activity is that of using oral reading at sight (i.e., without previous silent-reading preparation) as one means of appraising reading performance. (p. 456) (Emphasis added)

On the next page (p. 457), in giving a description of the "procedure for appraising reading achievement by means of an informal reading inventory", he lists "Oral Reading at Sight" as the first step and explains that this is done to "appraise reading behavior in a situation where the pupil is without benefit of preparation".

It appears that Betts clearly intended oral reading at sight to be the first step in administering an IRI, a form of testing, and that prepared oral reading was to be used in a directed reading lesson, a form of instruction. But the criteria for the reading levels is listed with descriptions

of the directed reading lesson. Thus it is unclear if Betts meant this criteria to be used with unprepared oral reading in a testing situation or some kind of continuous evaluation of reading progress made during the directed reading lesson.

Generally the criteria established by Betts has been used with unprepared oral reading. Several authors, however, debate this practice, especially since it appears that Betts based the criteria on the results of a study conducted for a doctoral dissertation by a student under his direction, Killgallon (1942). In the Killgallon study the subjects preread the research passages silently.

Since Betts did not specify what deviations from text should be considered as errors when using his criteria, considerable variation has resulted in interpretation and practice. Generally, counting the following deviations has been widely agreed upon: omissions; substitutions; insertions; gross or partial mispronunciations; and words aided. This agreement, however, may be due more to the high interscorer reliability found on these items, rather than their demonstrated relationship to frustration. Most authors also consider hesitations and lack of regard for punctuation important as well, but do not count them in computing the percentages, probably because they are difficult to score objectively (Ekwall, 1976, p. 266).

Whether or not to count repetitions as errors has been one of the more controversial issues (Ekwall, 1976, p. 267). Some writers feel repetitions should not be counted since

recent psycholinguistic research suggests the repetition or regression is frequently the student's means of reprocessing a selective bit of data necessary to the emerging story line (Guzak, 1970, p. 667). Other authors recommend counting only the first repetition but not subsequent repetitions of the same word or group of words. Some suggest counting only repetitions of more than one word, while others, like Ekwall, insist that all repetitions should be counted as errors.

Ekwall bases his insistence on research studies (Ekwall and English, 1971; Ekwall, Solis and Solis, 1973; Ekwall, 1974) that not only give support for his position, but also provide physiological evidence that, as material becomes more difficult, readers really do experience the anxiety associated with frustration. Using polygraph and galvanic skin measurement devices, the researchers found the students actually became physiologically frustrated before they reached the percentage of errors normally recognized as being at the student's frustration level. As Ekwall (1967) explains

...students become so concerned about their reading performance that their hearts beat faster, they begin to perspire, etc. just as one does when he is frightened or extremely nervous. With this sort of empirical research available it seems that there should be no doubt that using the normally recognized criteria, all repetitions should be counted as errors. (p. 267)

While instructions for preparing teacher made oral reading tests generally suggest use of the Betts' criteria,

the authors of commercial oral reading tests have usually developed their own standards (Powell and Dunkeld, 1971; Allington, 1984, p. 838). Criteria has differed from author to author but has generally allowed more errors at the lower grade levels. Other research studies have also suggested that the criteria for establishing reading levels should differ with the ability of the reader. Ekwall, Solis and Solis (1973), for instance, found that it seems to take fewer oral errors to frustrate good readers than poor ones. Studies by Cooper (1953) and Powell (1969) suggest that children in lower grades seem able to tolerate a greater percentage of oral errors while maintaining a given level of comprehension.

In reviewing studies concerning the reading levels criteria, it should be noted that various researchers are actually defining the frustration level differently. Powell (1969), for instance, is viewing it as the point where comprehension breaks down, while others like Betts (1946) and Ekwall (1967), are considering it as the place where difficulty in reading begins to produce an anxiety reaction in the reader. Still other researchers (Cooper, 1952; Dunkeld, 1970) have been concerned with validating the instructional level in terms of the relationship of error rate to achievement. Studies by Cambrell, Wilson, and Gantt (1981), Berliner (1981) and Jorgenson (1977) have suggested that achievement improves when students are placed in materials which produce error rates of 5% or less, and that

readers placed in materials which produced error rates greater than this tended to spend more time off task.

In the final analysis, while the traditional oral reading assessment practices described here appear to be very pervasive in both educational practice and the pedagogical literature, they remain a highly diverse and subjective matter, using varying standards and criteria, with amazingly little empirical evidence to support their widespread acceptance. On the other hand, traditional practices seem to prevail because, as yet, although efforts may be increasing, no one has presented conclusive evidence for anything better (Pikulski and Shanahan, 1982).

Traditional Versus Psycholinguistic Diagnosis During the late 1960's and early 1970's, researchers at Wayne State University, under the leadership of Kenneth Goodman, conducted a series of investigations in which they studied the oral reading "miscues" of children and adults. This research has provided new insights into the reading process and has led to the development of new theories and models of reading as well as a new approach to reading diagnosis. The Reading Miscue Inventory (Goodman and Burke, 1970) was developed as a diagnostic procedure based on principles generated by miscue research.

In the miscue analysis studies, a miscue was defined as the deviation between the oral response of the reader and the expected response of the text. Allen (1976) notes it

was a basic assumption of the studies that every response a reader makes is cued in some way by the reading situation and these responses will vary qualitatively.

Goodman (1967) has characterized the reading process as a "psycholinguistic guessing game" in which the reader is constantly sampling cues from the material, predicting what will come next and verifying those predictions by sampling more cues. The Goodman model is based on three cue systems which the readers in the miscue studies seemed to be using: (a) Grapho-phonetic (sound-symbol relationship) cues, (b) syntactic (grammar) cues, and (c) semantic (meaning) cues.

A basic assertion made by Goodman is that readers rely as little as possible on grapho-phonetic cues. Instead they tend to use higher order language and meaning cues, and their miscues are most often affected by semantic and, even more importantly, syntactic constraints. Authors have recently begun to characterize this type of model as "top-down" processing (DeBeaugrande, 1981). In contrast, traditional diagnosis has viewed reading as a "bottom-up" process, proceeding from letters to sounds to word recognition to meaning.

Both top-down and bottom-up processing models have had problems explaining, from a theoretical position, apparent contradictions which have appeared in particular research studies, especially differences in strategies used by good and poor readers and differences between recognition of words in isolation as opposed to recognition in context. In

response, Stanovich (1980) has proposed an "interactive-compensatory model" which suggest readers use both types of processing. Samuels and Kamil (1984, p. 213) explain that a poor reader, who may be inaccurate or slow at word recognition but who has knowledge of the text topic, may use top-down processes to compensate for the weakness in decoding. On the other hand, if a reader is skilled at word recognition but does not know much about the text topic, he may find it easier to simply recognize the words on the page and rely on bottom-up processes.

While many controversies still surround the interactive view of the reading process, Spiro and Myers (1984) have concluded that

By most accounts, the dominant view of reading today is that of an interactive activity (Rumelhart, 1977). Processing goes on from the bottom-up and from the top-down (either simultaneously or alternately). (p. 483)

Essential both traditional and psycholinguistic diagnosis consider the same reading behaviors as errors or miscues, but they have differed sharply in how those behaviors are viewed. Traditional diagnosis has treated all errors as undesirable behaviors to be eliminated. The purpose of error analysis is to determine the best instructional procedure to accomplish this. Psycholinguistic diagnosis considers miscues as a natural aspect of the reading process, and the term "miscue" is used instead of the word "error" to denote this distinction. Not

all miscues are considered undesirable. Qualitative rather than quantitative analysis of miscues is carried out to determine the seriousness of the miscue and to gain insight into the strategies being used by the reader.

While the goal in traditional oral reading assessment has been both diagnosis and placement in materials, placement has not been a goal of miscue analysis. Rather the reader is purposely given difficult material in order to elicit a sufficient number of miscues for making the analysis. Many subsequent studies, however, have examined the relationship between miscues and material difficulty as well as reader's proficiency (Christie, 1981; Christie and Alonso, 1980; Kibby, 1979; Leslie and Osol, 1978; Schlieper, 1977), and many of the findings of miscue research have implications which challenge assumptions underlying current readability theory.

Laura Smith (1976, p. 146), as a part of the reading miscue research project, was involved with testing materials being considered for inclusion in a new basal reading series. Based on the oral reading miscues and retellings by many children, she reported that the researchers found many factors that seemed to be ignored by readability scales. The factors could be categorized as either language related or concept related factors. Among language related factors, even though traditional readability theory asserts that short sentences are easier to read than long ones, the researchers found that very long sentences could be read

easily under the following conditions:

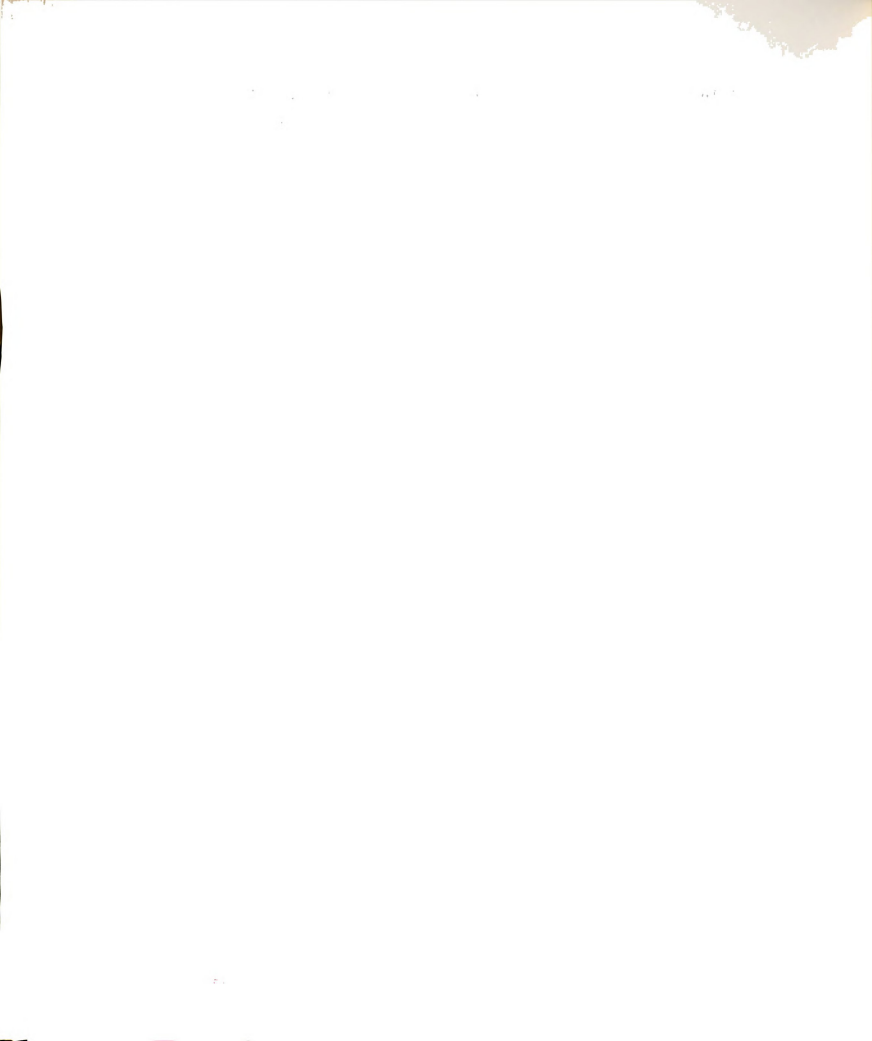
1. When the grammatical function of words and their meanings were familiar in a long sentence. The word brown, for instance, might be easily identified when used as a color word, but presented difficulty when used as someone's name. The word saddle was not a problem when it appeared as a noun, but was more difficult when used as a verb.
2. When the phrases in a long sentence were familiar. Phrases such as "she walks in such a way" and "Charlie turned his attention" were difficult for many readers.
3. When the tense choices in a long sentence were familiar to the reader and predictable in the story. Subtle changes in tense made by an author, usually to emphasize a point, were difficult for the readers. Frequently they would change the tense to the one they expected.
4. When the word order in a long sentence was predictable. Questions and negative statements were consistently not anticipated and readers frequently changed the construction into positive statements. Sentences beginning with the words what, where and when usually suggested a question and if the sentence was not a question the readers would often change the structure to make it a question. Dialogue and dialogue carriers presented problems. Dialogue carriers appearing at the beginning of the sentences were the easiest to read and those in the middle were the most difficult.

Dialogues containing a name were even harder. "We must hurry, John," said Mother. "We will be late." was often read as "We must hurry." John said, "Mother, we will be late." Unusual dialogue carriers, such as shouted, cried or screamed and additions to carriers such as gloomily, anxiously and briskly presented problems for the readers. In addition, the word order in directions and descriptions of processes caused more miscues than stories with a plot.

Three concept related factors were also found to be important: (a) The amount of specialized vocabulary, (b) the amount of vocabulary that was unfamiliar to the reader and (c) the complexity of the concept and how thoroughly it was developed.

In spite of their differences, both traditional and psycholinguistic diagnosis share some common weaknesses. Both generally assume oral reading can indicate silent reading processes, an assumption not universally agreed upon. Both diagnostic procedures also suffer from lack of empirical evidence of their validity and both rely heavily on judgments made by the examiner.

The assumption that oral and silent reading processes represent a unitary phenomenon is a position held by K. S. Goodman and implied in the Reading Miscue Inventory. Some studies (Fairbanks, 1937; Gillmore, 1947) have found a high correlation between silent and oral reading which would justify the use of oral reading performance to assess reading achievement in general. Other researchers (Wells,



1950; Mosenthal, 1976-77, 1978), however, found evidence supporting a contrary position.

As an indicator of silent reading response, oral reading may have its greatest validity when used at the primary level or in the beginning stages of reading development. It appears that, at these levels oral and silent reading tend to be very similar processes, but they soon begin to diverge, until finally, in the mature reader, they may become two totally different aspects of language. This position has been supported by Gray and Reese (1957), who found that a student's reading rate for both types of reading was virtually the same at the first grade level, but by second grade, silent reading was becoming faster and it continued to do so every year thereafter.

The problem of subjectivity continues to be a major concern in both traditional and psycholinguistic diagnosis since several studies have indicated that oral reading assessment can be a highly diverse matter with little, if any agreement among diagnosticians. Weinshank (1980), for instance, found agreement between any two practitioners (reading specialists, learning disabilities specialists and classroom teachers) concerning the reading diagnostic statements they made regarding the same case, was virtually nil (0.00). Moreover, she also found that when a clinician was presented a virtually identical replica of a case they had diagnosed at an earlier time, the mean agreement with their own previous statements was less than 0.23. Studies

by Sherman, Weinshank and Brown (1979), by Gill, Polin, Vinsonhaler and VanRoekel (1980) and by Polin (1981) have demonstrated, however, that practitioners can agree on what they find if first they agree on what they are looking for. These studies found diagnostic agreement could be improved dramatically through training, especially when decision aids were employed.

Summary of the Literature Review

Both the development of readability formulae and methods for assessing reading achievement through standardized testing occurred simultaneously but independently during the early part of this century. Both appear to have been influenced heavily by the almost universal emphasis on silent reading in instruction during the time, an emphasis which was prompted by the changing needs of society and supported by research studies indicating the superiority of silent reading in developing comprehension and fluency (Pitner, 1913; Thorndike, 1917; Judd and Buswell, 1922). It was perhaps because of this, that the use of oral reading in formula development or validation studies, seems to have been largely ignored, and oral reading as an assessment procedure didn't become popular until the middle of the century, prompted no doubt by dissatisfaction with standardized testing.

Formula Limitations While a vast number of readability formulae have been developed, virtually all have used the same methodology, and have encountered similar problems. These problems have centered on the factors studied and the criteria used in formula development. The factors studied have been seriously limited since only quantitative, rather than qualitative elements, can be used in the prediction, and generally only factors of style difficulty have lent themselves to that kind of analysis. Moreover, only two elements of style difficulty, some measure of vocabulary load and some measure of sentence complexity, have consistently emerged as significant enough, or measurable enough, to be included in the final formula.

The criterion materials used in formula development have varied widely in content, the range of difficulty of the criterion passages and the methods used to establish that difficulty. This greatly limits the generalizability of any one formula, for in a strict scientific sense the formula is only applicable to materials similar to those on which the formula was based. Moreover, the use of comprehension questions for establishing passage difficulty of a selection can be effected by asking more or less difficult questions. Finally, while formulae may be useful for establishing the relative difficulty of passages, their ability to relate this difficulty to the reading accomplishment needed by students in various grade levels is questionable.

The Fry Graph Historically, early formulae, after a short period of increasing complexity, showed a sharp reversal toward greater simplicity and ease of use. The Fry procedure (1968) is directly related to this continuing trend. Only two elements, word length measured in syllables and sentence length measured in words, are used, since previous research has repeatedly found these two factors account for a great deal of the variability in reading difficulty. It appears that the criterion materials used in developing the method were taken directly from basal readers or other materials intended for children. Apparently Fry has simply accepted the publishers grade level recommendations of the passages in establishing their relative difficulty. He has then circumvented the task of making tedious calculations by developing a nomograph, rather than a regression equation.

Assessment of Reading Ability Efforts to assess reading ability have basically taken two directions: (a) The development and use of standardized tests and (b) the development of oral reading assessment procedures.

Standardized tests have proven to be highly reliable measures useful for comparing students' reading performances and for screening students to determine if more extensive testing is needed. Such tests have been unsatisfactory, however, for providing direct observation of reading behaviors, for diagnosis of specific reading difficulties or

for placing students in reading materials.

Oral reading performance is frequently used as a readability measure and in informal reading assessment. Traditionally some variation of procedures and criteria described by Betts (1946) have been used for this purpose. While the Betts' criteria is frequently challenged, and contains contradictions resulting in much variation in practice, it remains widely accepted and has had great influence on traditional diagnostic theory. More recent psycholinguistic studies, however, are providing new insights into the reading process and a new approach to reading diagnosis. This approach views reading "miscues" as a natural reading phenomenon to be analyzed qualitatively rather than quantitatively. The results of miscue research studies have also held some important implications for readability prediction since the miscues made by the readers in these studies frequently contradicted some of the basic assumptions of current readability theory. Especially challenged are those assumptions concerning the difficulty of reading long sentences.

Oral reading has received very little attention in the development of readability prediction methods or in studies validating their use. This trend has continued to the present, with no indications of oral reading being used in developing criteria on which new formulae might be based, and while Fry (1969) and Paolo (1977) have used oral reading briefly in validation of the Fry procedure, and Fry

encourages the practice, any further use of oral reading for this purpose appears to be rare and obscure or unpublished if it exists at all.



DESIGN OF THE STUDY

Overview

This study was designed to use oral reading assessment procedures to evaluate the oral reading performance of a group of fifty (50) third grade students. The purpose of the study was to assess how effectively the readers' standardized test scores and Fry Readability Graph data would match readers with materials of appropriate difficulty.

The subjects' grade equivalency scores from the Reading Test of the California Achievement Tests were within three months above or below their grade placement at the time of testing, thus suggesting a rather homogeneous group of students of average reading achievement.

Each subject read the same set of five selections, one each with a readability of first, second, third, fourth and fifth grade, as determined by the Fry Readability Graph. The readability scores of the selections thus suggested gradually increasing difficulty from considerably below to considerably above the students' tested reading achievement.

If the students' standardized test scores and the readability graph data provide an effective means for matching readers with materials of appropriate difficulty, then we would expect to observe the following when the students were reading the research passages aloud: (a) The

subjects will make more word recognition errors (miscues) and will read more slowly as the readability of the passages increases, and (b) the subjects will read the passage with first grade readability with ease (at their independent reading level), the passage with third grade readability with some difficulty (at their instructional reading level) and the passage with fifth grade readability with great difficulty (at their frustrational reading level). Based on these expectations, the following questions and hypotheses were developed to direct the research.

Questions Guiding the Study

The following questions were generated to be answered by this study when subjects from a group of average third grade readers, as determined by the Reading Test of the California Achievement Tests, are reading aloud from selections with varying Fry determined readabilities.

1. Will the readers' word recognition accuracy, based on their oral reading errors (word miscues) decrease as the readability scores of the selections increase?
2. Will the readers' reading rate, in terms of the number of words read per minute, decrease as the readability scores of the selections increase?
3. Will the readers read the selection with first grade readability at their independent reading level?
4. Will the readers read the selection with third grade readability at their instructional reading level?
5. Will the readers read the selection with fifth grade readability at their frustrational reading level?

Hypotheses

Based on the questions guiding the research, the following hypotheses were constructed.

When a group of average third grade readers are reading aloud from materials with varying Fry determined readabilities

1. The mean of the word recognition accuracy scores for any paragraph will be greater than the mean of the word recognition accuracy scores for any paragraph with a higher readability.

1a. The mean of the word recognition accuracy scores for any paragraph will not be greater than the mean of the word recognition accuracy scores for any paragraph with a higher readability.

2. The mean of the reading rate scores for any paragraph will be greater than the mean of the reading rate scores for any paragraph with a higher readability.

2a. The mean of the reading rate scores for any paragraph will not be greater than the mean of the reading rate scores for any paragraph with a higher readability.

3. On the passage with a first grade Fry determined readability, the greatest percentage of the readers will be reading at their independent reading level.

3a. On the passage with a first grade Fry determined readability, the greatest percentage of the readers will not be reading at their independent reading level.

4. On the passage with third grade readability, the greatest percentage of the readers will be reading at their instructional reading level.

4a. On the passage with third grade readability, the greatest percentage of the readers will not be reading at their instructional reading level.

5. On the passage with fifth grade readability, the greatest percentage of the readers will be reading at their frustrational reading level.

5a. On the passage with fifth grade readability, the greatest percentage of the readers will not be reading at their frustrational reading level.

Population

The subjects for this investigation were selected from third grade students attending five Chapter I identified schools in the Bay City Public Schools System, Bay City, Michigan.

The identification of a school for Chapter I services in this district is based on the percentage of students eligible for free or reduced lunches. Since this figure is determined by family income, it is considered for these purposes to be an index of social economic status for the school's population. A school is determined eligible for Chapter I services in this district if it has a greater percentage of students eligible for free or reduced lunches than does the district as a whole. Thus the subjects in this study were attending schools in the lower half of the district's social-economic scale.

Sample Selection

Third grade students in the Bay City Public Schools had taken the level 12 California Achievement Tests (CAT) as second graders in May of the preceding school year as part of the system's district-wide testing program. After securing permission and support from the district's central administration, the researcher asked principals in the participating schools to first ascertain that third grade teachers in their buildings were willing to cooperate in the data collection, and then to identify third grade students

who had taken the CAT the previous spring and who had grade equivalency scores ranging between 2.3 and 3.3 on the Reading Test from that battery. Parents of these students were then sent letters briefly acquainting them with the study and asking for their permission in order to have their child participate. Those children who had parental permission and were themselves willing to be involved, were then administered the CAT Level 13 Reading Test. The group of fifty subjects were then chosen from those students with grade equivalency scores ranging from 3 months above to 3 months below their grade placement at the time of testing with the CAT Level 13 test.

Instrument Selection and Construction

Measurement of Student Reading Ability The Reading Test of the California Achievement Tests (1977 edition) was used in this study as a measure of student reading achievement. This instrument was chosen because it is a widely used and highly regarded, nationally normed standardized test. It is also the test used by the subjects' school system for its district-wide testing program. Therefore it is probable that instructional decisions affecting the subjects are commonly made based on results from this test.

Passage Selection The passages which were read by the subjects were taken from an SRA Reading Laboratory Ic,



published by Science Research Associates, Inc., 1961. Materials were selected from this source because it provided access to many short selections, similar in style, specifically written for children, and easily available. Furthermore, the material involved was not part of the subjects' regular reading program, and the lab chosen was an older edition not currently being used by teachers in the district. These considerations reduced the likelihood that the subjects may have had previous exposure to the material either as part of their regular reading program or as supplemental reading material. Finally, although the publisher provided readabilities for the selections, these readabilities did not correspond with those determined by the Fry Readability Graph.

The selections used in the study were of comparable length, ranging from 95 to 106 words, and were reproduced on plain white typing paper using the same type size and format and eliminating illustrations. Each selection was titled, however the title was not included in determining the readability of the passages and errors made by the subjects when reading the titles were not included in the error counts for the selections. Five selections were used, one selection each with a Fry determined readability score of first, second, third, fourth and fifth grade. An additional selection, one with a first grade Fry determined readability, was also prepared and read by all subjects as a practice passage.

Determination of Readability Scores from the Fry Readability Graph were used as the measure of passage difficulty of the selections. The Fry procedure was chosen because of its speed, ease of use and great popularity.

Readabilities of many selections were first computed by using the microcomputer text analysis program School Utilities Volume 2, published by the Minnesota Educational Computer Consortium. Three to five selections were then chosen at each grade level of readability. These selections were then plotted manually on the Fry Graph to verify the grade level designations obtained by the computer program. The passages used in the study were then chosen from those paragraphs which both the computer and the Fry graph designated as being at a given grade level.

Data Collection

Subjects were given an orientation session by the researcher, familiarizing them with the location in which they would be reading, the recording equipment that would be used and the task the researcher would be asking them to do. With the cooperation of their classroom teachers, each subject was taken individually to a relatively secluded area in the school to do the readings. All subjects first read the practice paragraph. They were then administered the research passages in random order and asked to read each of these aloud. The readings were audio recorded for later analysis.

Data Recording

Three types of data were recorded from the subjects' oral readings:

1. A word recognition accuracy score for each subject for each paragraph, based on the number of miscues made per 100 words. For instance, if a reader made two miscues per 100 words, the word recognition accuracy score would be 98%.

2. A reading rate score for each reader on each passage based on the number of words read per minute.

3. A reading level designation of "independent", "instructional" or "frustrational" for each passage for each reader based on the reader's word recognition accuracy score and using the Betts' criteria.

From the recordings of this data, the following determinations were made:

1. The means of the word recognition accuracy scores for each paragraph.

2. The means of the reading rate scores for each paragraph.

3. The percentage of readers reading at their independent reading level on each passage.

4. The percentage of readers reading at their instructional reading level on each passage.

5. The percentage of readers reading at their frustrational reading level on each passage.

Data Analysis

This study was primarily descriptive in nature, however the following data analysis techniques were used to aid the researcher in the descriptive process:

1. A repeated measures design with each individual exposed to five treatments was employed. Subjects read the research passages in random order to control for the sustained effects which can occur when taking repeated measures. Analysis of variance was then used to determine (a) if there were differences between the means of the word recognition scores for each paragraph and (b) if there were differences between the means of the reading rate scores. The computational formula used was presented by Winer (1971, p. 261-308) for use in single factor experiments with repeated measures. The formula is given in Appendix D. When analysis of variance indicated that differences did exist, the Scheffe test for post-hoc comparisons was used to determine where differences occurred. The computational formula was taken from Hinkle, Wiersma and Jurs (1979, p. 276-280), and is also presented in Appendix D. In the post-hoc comparisons, each paragraph was contrasted individually with all paragraphs of a higher readability. In other words:

P1 vs. P2 P2 vs. P3 P3 vs. P4 P4 vs. P5
P1 vs. P3 P2 vs. P4 P3 vs. P5
P1 vs. P4 P2 vs. P5
P1 vs. P5

where P1 = the means of the word recognition scores (or the reading rate scores) for the paragraph with first grade readability, P2 = the means of the word recognition scores (or reading rate scores) for the paragraph with second grade readability, etc.

2. A series of bar graphs was also constructed, to show the relative number of subjects reading at each of the three reading levels, independent, instructional and frustrational, for each passage.

CHAPTER IV

PRESENTATION AND ANALYSIS OF RESULTS

Introduction

In this chapter the results of the study and a descriptive analysis of the findings will be presented. The purpose of the analysis will be to determine (a) whether or not the frequency of miscue and the reading rates for each paragraph would suggest that the readers experienced increasing difficulty in the material as the grade level readability indexes of the passages increased, and (b) whether or not the readability grade level scores of a selection were predictive of the students' functional reading levels on that selection.

Reviewing the hypotheses set forth in this study, we would expect to find that the means of the scores of word recognition accuracy and reading rate would both decrease as the readability of the paragraphs increased. We would also expect to find that the greater percentage of the readers would be reading at their independent level on the paragraph with first grade readability, at their instructional reading level on the paragraph with third grade readability and at their frustrational reading level on the paragraph with fifth grade readability.

The results of the study will first be examined as they related to each of the questions originally developed to guide the research, and the hypotheses formulated in

association with each question.

Presentation of Results

Question 1. Will the readers' word recognition accuracy, based on their oral reading errors (word miscues) decrease as the readability scores of the selections increase?

H The mean of the word recognition accuracy scores for any paragraph will be greater than the mean of the word recognition accuracy scores for any paragraph with a higher readability.

Ho The mean of the word recognition accuracy scores for any paragraph will not be greater than the mean of the word recognition accuracy scores for any paragraph with a higher readability.

Table IV-1 presents the means of the word recognition accuracy scores for each paragraph when all miscues were counted. It should be noted that while this data is presented for each paragraph in sequential order, the subjects actually read the paragraphs in random order. Randomization was necessary in order to control for sustained effects.

Table IV-1

MEANS OF WORD RECOGNITION ACCURACY SCORES
BASED ON TOTAL NUMBER OF MISCUES

Paragraph	1	2	3	4	5
Means	93.92	94.4	94.36	93.8	92.92

As the table shows these means were almost the same for each passage. An analysis of variance for repeated measures (Winer, 1971, p. 266), as expected, indicated no significant



differences. (See Appendix D for formula and computational procedures used.)

The decision was to accept the null hypothesis for all contrasts and to conclude that, when all miscues were considered, the readers' word recognition accuracy scores did not decrease as the readability scores of the selections increased.

Question 2. Will the readers' reading rate, in terms of the number of words read per minute, decrease as the readability scores of the selections increase?

H The mean of the reading rate scores for any paragraph will be greater than the mean of the reading rate scores for any paragraph with a higher readability.

Ho The mean of the reading rate scores for any paragraph will not be greater than the mean of the reading rate scores for any paragraph with a higher readability.

Table IV-2

MEANS OF READING RATE SCORES
(WORDS READ PER MINUTE)

Paragraph	1	2	3	4	5
Means	93.976	110.93	99.998	98.172	94.358

Analysis of variance indicated significant differences among these scores. The Scheffe post-hoc comparison test (Hinkle, Wiersma and Jurs, 1979, p. 364-368) was used to make all possible pairwise comparisons between the mean of each paragraph and the mean of every paragraph with a higher readability. (See Appendix D for formula used and computational procedures.) Significant differences were

found between paragraphs 1 and 2, 1 and 3, 2 and 3, 2 and 4 and paragraphs 2 and 5. However, the differences between paragraphs 1 and 2 and paragraphs 1 and 3 were in a direction opposite of what would be expected. In other words paragraph 1 was read more slowly than paragraphs 2 or 3.

The decision was to accept the null hypothesis for 7 of the 10 contrasts and to reject the null hypothesis for the following contrasts: Paragraph 2 vs. paragraph 3, paragraph 2 vs. paragraph 4 and paragraph 2 vs. paragraph 5.

It was concluded from this data that the number of words read per minute did decrease between paragraph 2 and paragraph 3, paragraph 2 and paragraph 4 and between paragraph 2 and paragraph 5, but that the number did not decrease between any other paragraphs. In fact the number actually increased between paragraphs 1 and 2 and paragraphs 1 and 3.

Question 3. Will the readers read the selection with first grade readability at their independent reading level?

H On the passage with first grade readability, the greatest percentage of the readers will be reading at their independent reading level.

Ho On the passage with first grade readability, the greatest percentage of the readers will not be reading at their independent reading level.

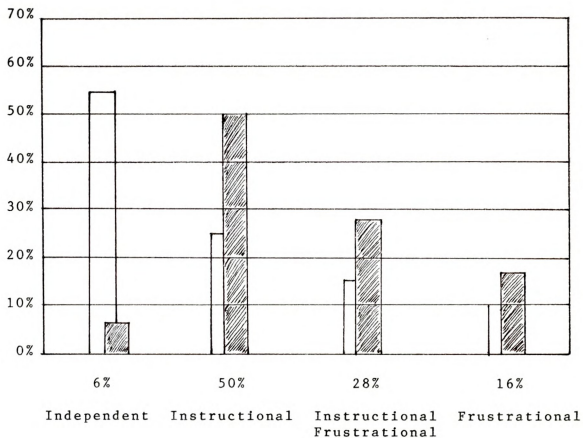
Table IV-3 presents a graph showing the relative percentages of readers at each of the functional reading levels, independent, instructional and frustrational, on the

paragraph with first grade readability. Because of the gap in the Betts' criteria, readers with word recognition accuracy scores from 91% to 94% did not fall into any of these categories. A fourth category labeled "Instructional-Frustrational" was created to accommodate data from these readings.

The shaded bars on the graph represent the actual results obtained in the study. The unshaded bars represent the results that could be reasonably expected if the alternate hypothesis were true. The unshaded bars were included to provide a means of comparison between actual and expected findings.

Table IV-3

PERCENTAGES OF SUBJECTS
 READING AT EACH FUNCTIONAL READING LEVEL
 ON PARAGRAPH #1
 (First Grade Readability)



As the graph indicates 6% of the readers read the first grade passage at their independent reading level, that is with 99% or 100% word recognition accuracy. Fifty percent were reading at their instructional level (95% to 98% word recognition accuracy) and 16% were reading at their frustrational level (90% word recognition accuracy or less). Twenty-eight percent of the readers had scores between 94% and 91% and were placed in the "Instructional-Frustrational" category.

The greatest percentage of readers were reading at their instructional level on paragraph 1, rather than at their independent reading level. Therefore, the decision was to accept the null hypothesis for question 3.

Question 4. Will the readers read the selection with third grade readability at their instructional reading level?

H On the passage with third grade readability, the greatest percentage of the readers will be reading at their instructional reading level.

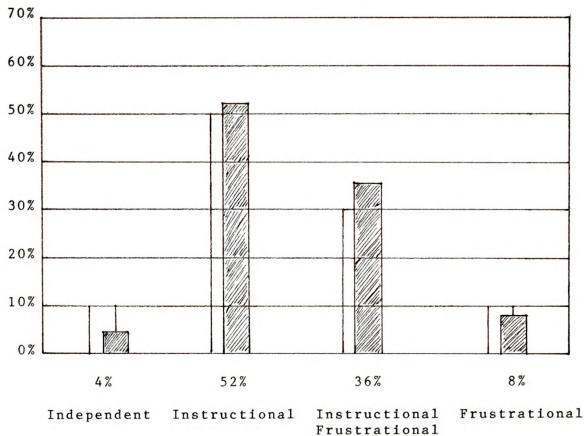
Ho On the passage with third grade readability, the greatest percentage of the readers will not be reading at their instructional reading level.

Table IV-4 presents a graph showing the relative percentages of readers reading at each functional reading level on the third grade paragraph.



Table IV-4

PERCENTAGES OF SUBJECTS
READING AT EACH FUNCTIONAL READING LEVEL
ON PARAGRAPH #3
(Third Grade Readability)



Four percent of the readers were reading at their independent reading level. Fifty-two percent were reading at their instructional reading level and 8% were reading at their frustrational reading level. The other 36% fell into the "Instructional-Frustrational" category.

The greatest percentage of the readers were reading at their instructional level on this paragraph. The decision, therefore, was to reject the null hypothesis in favor of the alternative hypothesis for question 4.

Question 5. Will the readers read the selection with fifth grade readability at their frustrational reading level?

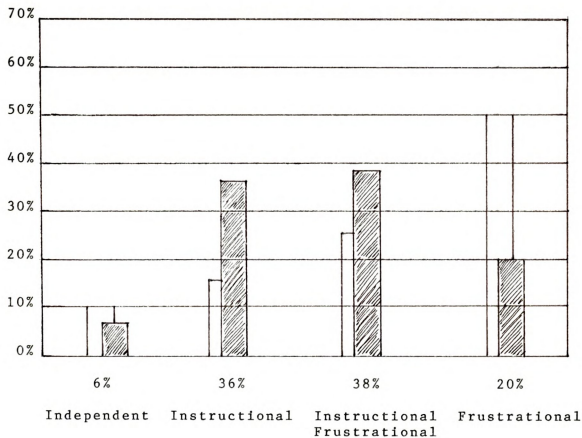
H On the passage with fifth grade readability, the greatest percentage of the readers will be reading at their frustrational reading level.

Ho On the passage with fifth grade readability, the greatest percentage of the readers will not be reading at their frustrational reading level.

Table IV-5 presents a graph showing the relative percentages of readers reading at each functional reading level on the fifth grade paragraph.

Table IV-5

PERCENTAGES OF SUBJECTS
 READING AT EACH FUNCTIONAL READING LEVEL
 ON PARAGRAPH #5
 (Fifth Grade Readability) *



Six percent of the readers were reading at their independent reading level. Thirty-six percent were reading at their instructional reading level and 20% were reading at their frustrational reading level. The other 38% fell into the "Instructional-Frustrational" category.

The greatest percentage of readers were reading at their instructional and instructional-frustrational level, and not at their frustrational reading level on paragraph 5. The decision, therefore, was to accept the null hypothesis for question 5.

Additional Data Analysis

As indicators of whether or not the passages in this study showed evidence of increasing difficulty, the initial results appeared contradictory. The results from the word recognition accuracy scores and the determination of reading levels seemed to indicate that there was no change in difficulty from paragraph to paragraph. The reading rate scores, however, indicated that, while the readability scores did not appear to discriminate well, there were increases in difficulty between paragraph 2 and paragraph 5, with 2 being the easiest and 5 being the most difficult. During the data collection process, however, the researcher made the following additional observations which had not been previously anticipated and which seemed to have implications for the study:

1. Although the readers seemed to encounter difficulty on all of the passages, miscues of a more serious nature, such as words aided and gross mispronunciations, seemed to be occurring more frequently on the paragraphs with higher readability scores. This was particularly true on the fifth grade passage where there were 29 instances of "words aided". There were only 3 occurrences on paragraphs 1 and 4 each and no such occurrences on paragraph 2 or 3.

2. Fluency seemed to be more a reader characteristic, rather than a function of passage difficulty.

3. Miscues did not seem to occur randomly. Instead readers tended to miscue on the same words and at the same places in a passage. Moreover, they frequently made the same or a similar response.

Because these observations had implications for readability research and oral reading assessment practices, and because reporting the original results without further exploration of the data might lead to erroneous conclusions, three additional questions were raised for study.

1. Would the readers' word recognition accuracy scores decrease as the readability of the paragraphs increased when only unacceptable miscues were considered in determining word recognition accuracy?

2. Did the readers read with less fluency as the readability scores of the passages increased?

3. Did the readers' miscues occur randomly or were there identifiable patterns?

In order to better answer these questions, additional data analysis was undertaken as follows:

1. Miscues were classified as either acceptable or unacceptable. An acceptable miscue was defined as any miscue that had no effect or negligible effect on meaning and any miscue that was corrected. Based on this classification, a percent of accuracy score was determined for each reader for each paragraph when only unacceptable miscues were considered. The means of these scores for each paragraph were then tested to see if there were statistically significant differences between means. The procedure was the same as that used with the original word recognition accuracy scores when all miscues were counted.

2. The reading of each paragraph was rated on a scale of 1 to 5, with 1 representing the greatest fluency. The ratings were based on the researcher's general impression of the fluency with which the passage was read. In order to make this determination, the researcher listened to each recorded reading without watching the script and judged the reading on the basis of how it would compare to a television or radio newscast. To receive a high rating the reading had to make sense, be read with appropriate phrasing and intonation and be free of hesitations and repetitions. Readings receiving low ratings had many instances of mispronunciations, omissions, substitutions or improper phrasing which rendered some portion of the reading

senseless or the reading was characterized by monotone, word by word reading, hesitations, stammering, long pauses or requests for aid. The fluency rating scores for each paragraph were then totaled and the means were tested for differences, again using the same procedure as that used for the word recognition scores.

3. A frequency count of the miscues that occurred on each word in each paragraph was made. Graphic representations were then constructed depicting the frequency of miscue for each word in each paragraph.

Additional hypotheses were developed for the first two questions, based on the results that would be expected if the readers were truly experiencing more difficulty as the readability of the passages increased. Hypotheses were not developed for the third question because the question did not lend itself to hypothesis testing. Rather an inspection of the miscue frequency graphs was used to decide if discernable patterns of miscue clustering occurred and if high incidence of miscue seemed to be occurring on specific words.

Each of the additional questions, their accompanying hypotheses and the results of the data analysis associated with them will be presented in turn.

Presentation of Additional Data Analysis Results

Question 1a: Will the readers' word recognition accuracy scores decrease as the readability of the



selections increases if only unacceptable miscues are considered in determining the word recognition accuracy scores?

H The mean of the word recognition accuracy scores for any paragraph, when based on the unacceptable miscues only, will be greater than the mean of the word recognition accuracy scores for any paragraph with a higher readability.

Ho The mean of the word recognition accuracy scores for any paragraph, when based on the unacceptable miscues only, will not be greater than the mean of the word recognition accuracy scores for any paragraph with a higher readability.

Table IV-6 presents the means of the word recognition accuracy scores for each paragraph when only unacceptable miscues were considered.

Table IV-6

MEANS OF WORD RECOGNITION ACCURACY SCORES
WHEN ONLY UNACCEPTABLE MISCUES WERE COUNTED

Paragraph	1	2	3	4	5
Means	99.28	99.58	99.02	99.14	97.36

Analysis of variance indicated significant differences in these scores. The Scheffe post-hoc comparisons indicated significant differences between paragraphs 1 and 5, paragraphs 2 and 5, paragraphs 3 and 5, and paragraphs 4 and 5. It was decided, therefore, to accept the null hypothesis for six of the ten contrasts and to reject the null hypothesis in favor of the alternate hypothesis for the following contrasts: Paragraph 1 vs. paragraph 5; paragraph 2 vs. paragraph 5; paragraph 3 vs. paragraph 5; and paragraph 4 vs. paragraph 5. It was concluded that, when



only unacceptable miscues were considered, word recognition accuracy scores did decrease between paragraph 5 and all other paragraphs, but not between any other paragraphs.

Question 2a: Will the readers' fluency decrease as the readability indexes of the passages increase?

H The mean of the General Impression of Fluency scores for any paragraph will be less (indicating greater fluency) than the mean of the General Impression of Fluency score for any paragraph with a higher readability.

Ho The mean of the General Impression of Fluency scores for any paragraph will not be less than the mean of the General Impression of Fluency score for any paragraph with a higher readability.

Table IV-7 presents the means for the "General Impression of Fluency" ratings.

Table IV-7

MEANS OF GENERAL IMPRESSION OF FLUENCY SCORES

Paragraph	1	2	3	4	5
Means	2.32	2.09	2.28	2.34	2.6

Analysis of Variance indicated there were significant differences in the means of these scores. When the Scheffe Test was used to make post-hoc, pairwise comparisons, significant differences were found between paragraphs 1 and 5, 2 and 5, and 3 and 5. The decision was to accept the null hypothesis for 7 of the 10 contrasts, and to reject the null hypothesis in favor of the alternate hypothesis for paragraphs 1 vs. 5, 2 vs. 5, and 3 vs. 5. It was concluded

that a significant difference in the fluency scores did occur between paragraphs 1 and 5, 2 and 5, and 3 and 5 but not between any other paragraphs.

Question 3a: Did the readers' miscues occur randomly or were there predictable patterns?

Table IV-8 presents a graphic representation of the frequencies of acceptable and unacceptable miscues for the first grade paragraph. Likewise Tables IV-9, IV-10, IV-11 and IV-12 present the same information for paragraphs 2, 3, 4, and 5 respectively. The circled numbers indicate the sentence within the paragraph. The other numbers indicate the position of each word within the paragraph, followed by the word. The solid squares indicate the number of readers who made unacceptable miscues on the word, while a square with an x in it represents the number making acceptable miscues. The first word in the first sentence in Paragraph #1, for instance, was "Something". Four readers made an unacceptable miscue on this word and three others made an acceptable miscue. Squares between words represent insertion miscues.



Table IV-8

FREQUENCIES OF MISCUES
OCCURRING ON EACH WORD
IN PARAGRAPH #1

Unacceptable Miscue

Acceptable Miscue x

① 1 Something	XXXXX	37 around		73 comes	XX
2 is	X	38 you	X	74 out.	
3 all	XX	39 when		⑬ 75 Many	XXXXX
4 around		40 it	XXXXX	76 things	X
5 you	XX	41 blows	X	77 use	XX
6 at	XXXXXX	42 hard.	XX	78 air.	XXXXX
7 all	XX	⑧ 43 You		⑭ 79 Tires.	XXXXX
8 times		44 can	XX	80 windmills	XXXXX
⑨ 9 You		45 feel	X	81 and	
10 cannot.	XXXX	46 it		82 footballe	X
11 see		47 then.	X	83 use	XXXX
12 it.		⑩ 48 Sometimes	X	84 air.	X
⑬ 13 But		49 you		⑮ 85 Sailboats	XX
14 sometimes	XXXXXX	50 can		86 are	XXXXXX
15 you		51 feel	XXXXX	87 pushed	XXXX
16 can	X	52 air	X	88 by	X
17 feel		53 from	XX	89 air.	XX
18 it.		54 a	XXXXX	⑯ 90 Kites	XX
⑬ 19 Without	XXXX	55 balloons		91 are	XXX
20 it	XXXXX	⑩ 56 Blow	XX	92 kept	X
21 nothing	XXXX	57 up	X	93 up	XXXXXXXX
22 can	X	58 a	X	94 by	X
23 live	XX	59 balloons	X	95 air.	XXXXXX
⑭ 24 Do	XXXXX	⑪ 60 Then	XXXXX	⑰ 96 Air	X
25 you	XX	61 let	X	97 helps	X
26 know	X	62 go	XXXXXXXXXXXX	98 to	XXXX
27 what	XXXXX	63 of	XXXX	99 keep	X
28 it	XX	64 its	XXXXXX	100 airplanes	XX
29 is	X	65 south	XX	101 in	X
⑬ 30 It	XXXXXXXXXXXX	⑮ 66 Can	XXXXXXXXXXXX	102 the	
31 is	XXXXXXXXXXXX	67 you	XXXXXXXXXXXX	103 sky	XX
32 air	XX	68 feel	X	104 too.	XXX
⑦ 33 You	X	69 the			
34 know	X	70 air			
35 it	XXX	71 as	XXXX		
36 is	XX	72 it	XX		
	XX		X		

Insertion miscues appear between words.

Circled numbers indicate sentence number.

Uncircled numbers followed by a word identify each word in the paragraph.

Table IV-9

FREQUENCIES OF MISCUES
OCCURRING ON EACH WORD
IN PARAGRAPH #2

Unacceptable Miscue ■

Acceptable Miscue x

① 1 Long	37 rocks. XXX	73 the
2 ago	② 38 He XXXX	74 rocks XXXXXX
3 many X	39 carried XXX	75 had X
4 large XX	40 them	76 been X
5 rocks X	41 to X	77 in ■
6 lay XXXX	42 the XXXX	78 the ■
7 all XXXXXXXXXXX	43 sides XXXXXXXXXXXXX	79 way. ■
8 over XXX	44 of	⑩ 80 but X
9 the XX	45 the XX	81 soon X
10 ground	46 field. ■	82 they XX
⑩ 11 There XXX	④ 47 He XX	83 helped ■
12 was XXXXXXXX	48 made XX	84 the ■
13 a ■	49 a XXX	85 farmer. XXXX
14 farmer ■	50 fence ■	⑪ 86 And XXXX
15 who	51 of X	87 the X
16 wanted	52 the XXXXXXXXXXXXX	88 farmer's XXXX
17 to	53 rocks. X	89 rock XXXXXXXXXXX
18 grow	⑦ 54 Then XX	90 fence
19 things ■	55 all	91 made ■
20 on	56 the XX	92 the XXXXXXXX
21 the XXXXXXXX	57 other XX	93 field
22 land.	58 farmers	94 more
③ 23 But X	59 could XXXX	95 beautiful. X
24 nothing	60 see	
25 would	61 where XXXX	
26 grow X	62 his XXXXX	
27 where XXXXXXXXXXXXXXX	63 field X	
28 the XXXX	64 was	
29 rocks X	⑧ 65 Flowers ■	
30 were. XXXX	66 grew	
④ 31 So X	67 along X	
32 he XX	68 the	
33 started X	69 rock XXXXXXXX	
34 picking XX	70 fence. XXXX	
35 up X	⑨ 71 At XXXXX	
36 the XX	72 first	

Insertion miscues appear between words.

Circled numbers indicate sentence number.

Uncircled numbers followed by a word identify each word in the paragraph.



Table IV-10

FREQUENCIES OF MISCUES
OCCURRING ON EACH WORD
IN PARAGRAPH #3

Unacceptable Miscue ■

Acceptable Miscue x

① 1	Let's	X	37	your	XXXXXXXXXXXXXXXX	73	run	■
2	pretend	XX	38	soo		74	your	XXX
3	you're	XX	39	a	I	75	soo	
4	running	X	40	year	XX	76	you'll	XXXXXX
5	a	XXXXXX	41	ago.		77	look	XXXXXX
6	soo.		⑤ 42	but		78	around	XX
⑦ 7	in	X	43	you're	XXXX	79	for	
8	your	XX	44	unlucky	XXXXXXXXXXXX	80	a	X
9	soo		45	to	XX	81	soo	X
10	you	X	46	have	X	82	that	
11	have		47	only	XXXX	83	wants	X
12	four	X	48	one	X	84	a	
13	tigers		49	polar	■	85	tiger.	
14	but	X	50	bear.		⑧ 86	Maybe	
15	only	■	⑥ 51	it	XXX	87	you	
16	one	XXXXXXXX	52	isn't	XXX	88	can	X
17	polar	XXXX	53	much	XX	89	trade	X
18	bear		54	fun	XX	90	for	XXX
① 19	You're	X	55	for	XX	91	a	XX
20	lucky		56	people	X	92	polar	■
21	to		57	who	XX	93	bear.	
22	have		58	come	XX	⑨ 94	Like	X
23	the	XX	59	to	XX	95	every	XXXXXXXXXX
24	tigers.	XXXX	60	your	XXXX	96	soo	XX
① 25	Tery	XXX	61	soo	X	97	man	XXX
26	few		62	to	XX	98	you	X
27	tigers		63	watch		99	have	X
28	are		64	one	XXXXXXXXXXXX	100	trading	■
29	born		65	lonely	XXXXXXXXXX	101	in	■
30	in		66	polar	■	102	your	■
31	soos	XXXXXXXXXX	67	bear.		103	blood	XXXXXX
32	but	XX	⑦ 68	if				
33	two		69	you	X			
34	were	XX	70	know	X			
35	born		71	how	XX			
36	in		72	to				

Insertion miscues appear between words.

Circled numbers indicate sentence number.

Uncircled numbers followed by a word identify each word in the paragraph.

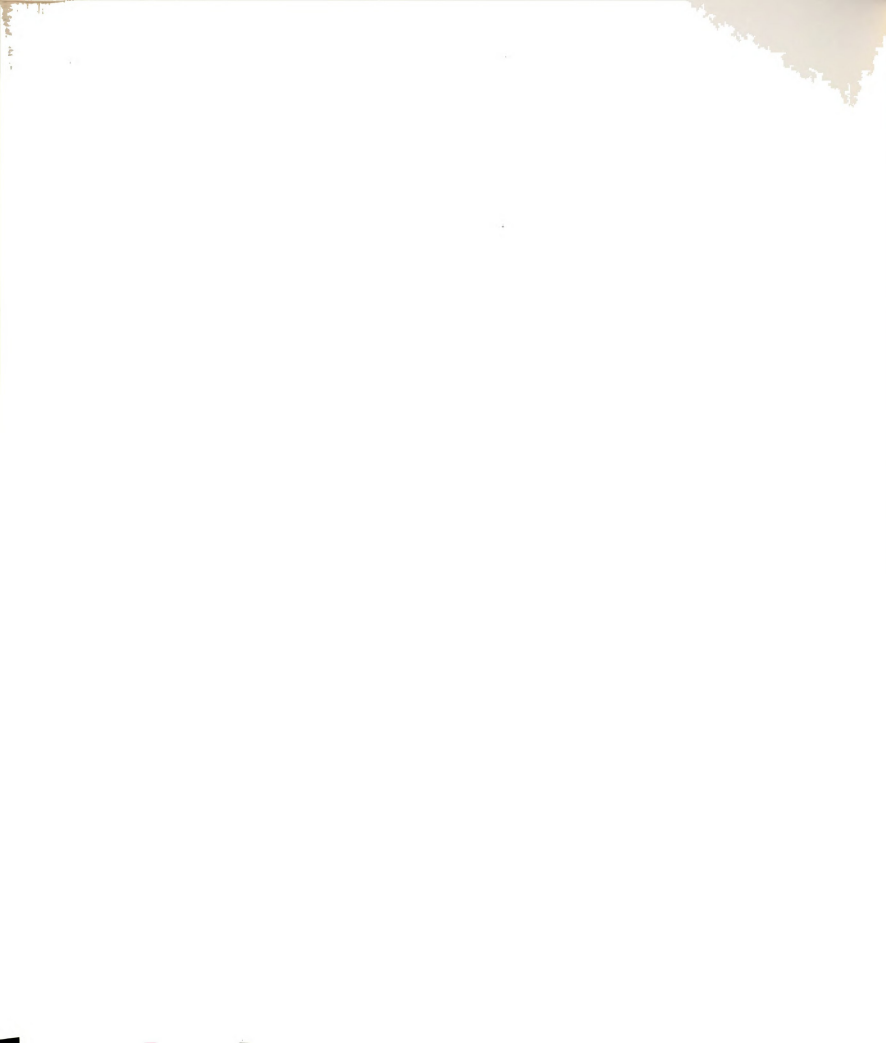


Table IV-11

FREQUENCIES OF MISCUES
OCCURRING ON EACH WORD
IN PARAGRAPH #4

Unacceptable Miscue ■		Acceptable Miscue x	
① 1 Maybe ■■	37 flip	73 a ■■■■■	
2 you ■■■■	38 over ■■	74 clip. ■■■■■	
3 already ■■	39 and ■■■■■■	⑦ 75 Have ■■■■■■■■■■	
4 know ■■	40 stick ■	76 you ■■■■■■■■■■	
5 how ■■	41 to ■■■■	77 ever ■■	
6 magnets ■■	42 the ■■■■	78 cried ■	
7 work. ■■	43 magnet. ■■	79 that? ■■	
② 8 If ■	④ 44 This ■■	⑧ 80 Maybe	
9 you ■■	45 happens ■	81 not ■	
10 were ■■■■	46 because ■■	82 but ■■	
11 to ■■■■	47 paper ■■	83 you ■	
12 hold ■■■■■■	48 clips ■■	84 know ■	
13 a ■■	49 are ■■	85 that ■■■■	
14 magnet ■	50 made ■	86 there ■■	
15 near ■■	51 of ■■	87 is ■■	
16 a ■	52 iron. ■	88 no ■■■	
17 paper	⑤ 53 And ■■	89 iron	
18 clip	54 anything ■■■	90 in ■■■■■■■■	
19 on ■■	55 made ■	91 paper. ■■	
20 your ■■	56 of ■	⑨ 92 And ■■■■■	
21 deck ■■	57 iron	93 because	
22 you'd ■■■■■■■■	58 sticks	94 of ■■	
23 know	59 to ■■■■	95 this ■	
24 what ■■■■■■	60 magnets ■■■■■■	96 you ■	
25 to ■■	⑥ 61 What ■■	97 feel ■	
26 expect ■■■■■■■■	62 would ■	98 sure ■	
③ 27 When ■■	63 happen ■	99 that ■■	
28 the ■■	64 if ■	100 paper	
29 magnet ■	65 you	101 will ■■	
30 got ■	66 used ■	102 not	
31 close ■■	67 a	103 stick	
32 the ■■■■■■■■	68 piece ■	104 to	
33 paper ■	69 of	105 the ■■■■■■■■	
34 clip ■	70 paper	106 magnet ■	
35 would ■■	71 instead ■■■		
36 suddenly ■■■■	72 of ■		

Insertion miscues appear between words.

Circled numbers indicate sentence number.

Uncircled numbers followed by a word identify each word in the paragraph.

Table IV-12

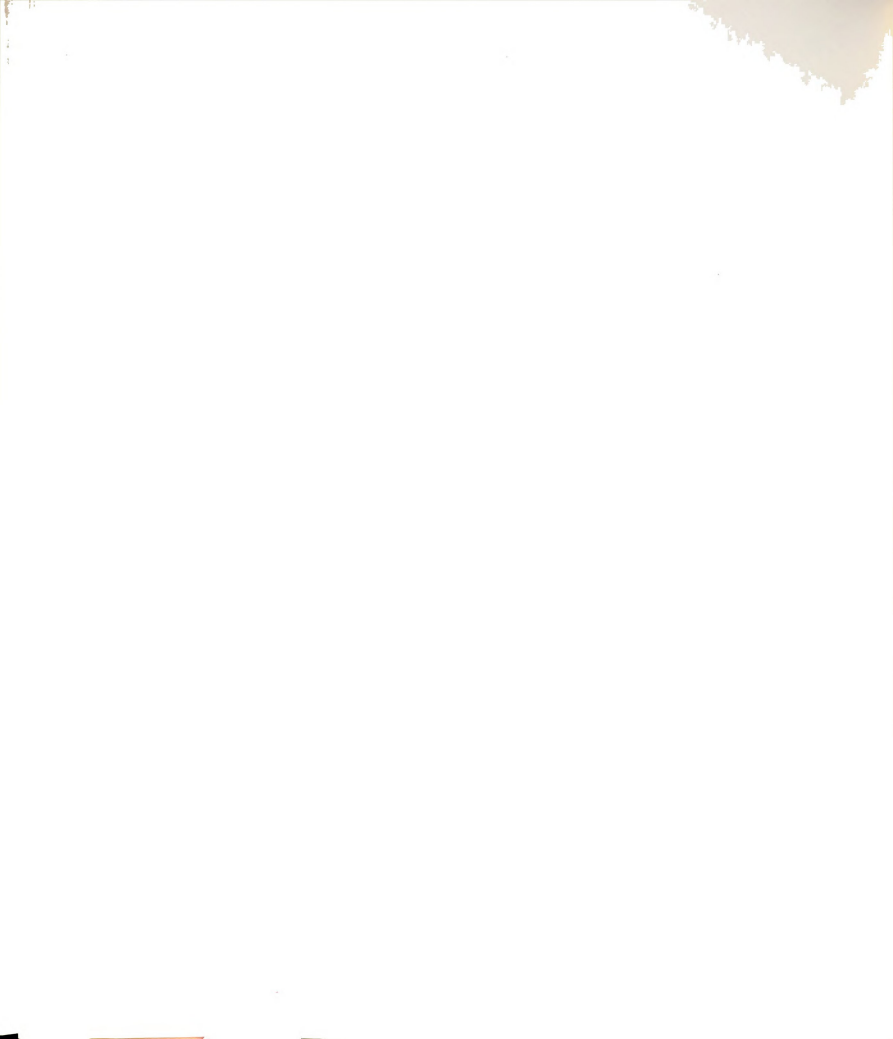
FREQUENCIES OF MISCUES
OCCURRING ON EACH WORD
IN PARAGRAPH #5

Unacceptable Miscue ■			Acceptable Miscue x		
① 1	The	37 facing	XXXXXXXXXX	73 is	X
2	next	38 the		74 is	XXXXXXXXXX
3	time	39 street		75 a	X
4	you	40 or	X	76 new	XXXX
5	are	41 maybe	XX	77 building	XX
6	going	42 none	X	78 there	XX
7	into	43 at	XX	79 way	XXXX
8	a	44 all	X	80 be	XX
9	bank	④ 45 The	XX	81 huge	XXXXXXXXXX
10	stop	46 front		82 windows	XX
11	a	47 of	XX	⑦ 83 Through	
12	minute	48 the	XXX	84 then	
13	before	49 bank	W-	85 you	
14	you	50 will	X	86 can	X
15	push	51 seen	XX	87 see	
16	open	52 a	X	88 the	XX
17	the	53 solid	XXXXXXXXXX	89 bank's	XXXX
18	heavy	54 stone	XXX	90 workers	XX
19	door.	55 wall.	X	91 and	XX
② 20	Look	③ 56 Through	XX	92 customers	XXXXXXXXXX
21	at	57 the	X	93 as	XX
22	the	58 door	X	94 wall	X
23	building.	59 you		95 as	XXXX
③ 24	If	60 way	X	96 a	XXXXXXXXXX
25	it	61 see		97 uniformed	XXXXXXXXXXXXXXXXXXXX
26	is	62 a		98 guard	XXXX
27	as	63 bank		99 who	X
28	old	64 guard	XXXXXX	100 is	
29	bank	65 whose	XXXXXXXXXXXX	101 not	
30	there	66 uniform	XXXXXXXXXXXXXXXXXXXX	102 armed.	XXXXXX
31	will	67 includes	XXXXXXXXXXXXXXXXXXXX		
32	be	68 a	XX		
33	only	69 gun.			
34	a	⑥ 70 If			
35	few	71 your	XXXXXXXXXXXX		
36	windows	72 bank			

Insertion miscues appear between words.

Circled numbers indicate sentence number.

Uncircled numbers followed by a word identify each word in the paragraph.



As the graphs show, some words were never involved in a miscue while other words found 20%, 30%, 40% and even 50% of the readers miscuing. As the graphs also show, miscues tended to cluster at certain places in certain sentences. Based on these observations it was concluded that miscues did not occur randomly, but rather tended to cluster on certain words and in certain parts of sentences in identifiable patterns.

Descriptive Miscue Analysis

It is not the purpose of this study to provide an indepth analysis of the type of miscues made by the readers, however, some of the text conditions that were associated with their miscues and seemed to be triggering them, were strikingly similar to those reported by Laura Smith (1976) as a part of the miscue research project. Because they have implications for readability study, they could not be dismissed without comment.

Generally, the miscues observed in this study could be classified in three categories. In the first category were miscues of little or no consequence. For instance, readers would consistently substitute a contraction for the two words for which it stood. They omitted articles or added them or substituted one for another, and the "s" at the end of a word was often disregarded, with negligible or no effect on meaning. In paragraph two, for example, the sentence "He made a fence of the rocks." was often read "He

made a fence of rocks." or "He made the fence of rock."
The sentence "He carried them to the sides of the field."
was read "He carried them to the side of the field" or "to
sides of the field."

Category one miscues were always acceptable miscues and while the reader's production was not the same as the text, the text could have just as well been written as the reader read it. In fact, in some instances, the reader's miscues actually seemed to produce a better flowing, easier to read version. For example, in the first grade paragraph, in the sentences "Do you know what it is? It is air!", the repetition of the words "it is" gave the reading an awkward and unnatural cadence. Readers consistently substituted a contraction for the second "it is", which produced a smoother flowing text.

In the second category of miscues were those that occurred because, even though it seemed obvious that the reader had sufficient word recognition skills to identify all of the words in the passage, certain conditions in the text, or in the reader's ability to handle those conditions, seemed to repeatedly interfere with the reader's processing strategies. Generally, because they did have sufficient word recognition abilities, the readers were able to recover from these situations with little or no serious damage. These miscues, however, did affect the reader's speed and general fluency and, in some cases, when they were not corrected, they had implications for the reader's

understanding of the text.

Most of the conditions involved in these category 2 miscues were similar to those observed previously by Laura Smith (1976) in her work with the reading miscue research studies. Unfamiliar grammatical function or meaning of a word, unfamiliar phrases, and unfamiliar word order, especially the use of rhetorical questions, accounted for most of these miscues.

In this study there were several instances where a familiar word was used with an unfamiliar meaning, and while the readers did not miscue on the word itself, they made insertions or deletions to make the meaning conform to the one they knew. As Smith has noted, for instance, young readers seem to be more familiar with a word when it is used as a noun rather than a verb. This was evident when children were reading, the latter part of sentence 3 in paragraph 4, "When the magnet got close, the paper clip would suddenly flip over and stick to the magnet". This sentence was frequently read "...the paper clip would suddenly flip over the stick to the magnet" or "the stick of the magnet" or, in one instance "the stick of metal of magnet." Evidently these readers anticipated that "stick" would be a noun, and therefore inserted the word "the" in front of the word "stick" to make it a noun. This of course meant that the rest of the words in the sentence did not make sense and the reader was forced to reevaluate the situation and decide how best to proceed.

The first sentence in paragraph 3 presented readers with the word "running". All of the readers read the word correctly, but, because the word was used with a meaning evidently unfamiliar to some, a preposition was consistently inserted to make the word conform to the meaning more common to the readers. Therefore, "Let's pretend you're running a zoo." was read as "Let's pretend you're running in a zoo" or "running to a zoo" or "running at a zoo", or even, in one case, "running on a zoo". In this situation there was nothing to alert the reader that a miscue had occurred, and the miscue was seldom corrected, however it did have implications for the reader's understanding of the passage.

The unfamiliar phrase "you have trading in your blood" in the last sentence of paragraph 3, also caused problems for many readers. Several students read this phrase as "you have traded in your blood." Most readers seemed to understand the concept of "trading something in". Their families had no doubt traded in cars or appliances. But they did not understand what it meant to have something "in your blood", so they substituted the concept they did understand. It was apparent, however, that they still could not understand why anyone would want to "trade in their blood". This prompted some to reprocess the phrase and sometimes correct the miscue. However, they still indicated a lack of understanding through their hesitancy and questioning tone.



Unfamiliar word order consistently caused readers difficulty. In paragraph 2, in the sentence "But nothing would grow where the rocks were.", one third of the readers read "where" as "there". Not only did this make perfect sense at the time the miscue was made, but the word "where" came at the end of a line of type, making it appear even more likely to be the end of the sentence. The physical position of the word on the page, its graphic similarity to the actual word and its perfect sense undoubtedly accounted for the high frequency of miscue. Once made, however, it left the reader trying to figure out what to do with the words "the rocks were". Most read these words with an intonation that would suggest they thought they were part of another sentence "There the rocks were." Some tried to make "the rocks were" part of the next sentence, but, interestingly, very few went back to correct the miscue.

On the first grade paragraph, the sentences "Blow up a balloon. Then, let go of its mouth." presented readers with an unfamiliar word order. The readers repeatedly read the second sentence "Then let it go." or "Then let go of it." This of course left remaining words which did not make sense and the readers were forced to cope with the situation in various ways.

In paragraph 4, in the sentence "When the magnet got close, the paper clip would suddenly flip over and stick to the magnet...", readers repeatedly ignored the comma and inserted "to" at the end of the opening clause, so it read

"When the magnet got close to the paperclip...". This again left the reader with words that made no sense. Many readers simply went on, their intonations suggesting that they may have made a covert correction, while others fumbled and tried to recover. One reader inserted an "it" to make the sentence read "...it would suddenly flip over". Another inserted "what" to make a question "...what would suddenly flip over..." and one inserted "you", to make it read "you would suddenly flip over and stick to a magnet."

Authors of children's texts frequently insert questions, presumably to increase the reader's involvement and thereby heighten their interest. This seemed to be the case with some of the passages used in this study. These questions, however, usually produced a high incidence of miscue. "Do you know what it is?" and "Can you feel the air as it comes out?" in paragraph 1, and "Have you ever tried that?" in paragraph 4 were often converted to statements. In the first sentence, the "do" was typically omitted, although this miscue was often corrected soon after it was made. In the second sentence, the words "Can you" were usually reversed to make a statement. This miscue was usually not corrected, probably because it made perfect sense as it was. In the third sentence the words "Have you" were usually reversed also to make a statement, but in this case the results did not make sense. Some readers simply went on, while others struggled to recover.

The most interesting question, however, was one that appeared in paragraph 4 and read "What would happen if you used a piece of paper instead of a clip?". This sentence was interesting because, unlike the other questions, it did not produce many miscues. Evidently, the word "what" at the beginning of the sentence provided the readers with a familiar signal of a question and they were better able to predict the text.

In the third category of miscues, readers began to encounter situations where they no longer had the capacity to recover. This usually involved words that they did not know and did not have sufficient word analysis or contextual analysis skills to figure out. The reader either had to ask for help, stop until help was given or make the best attempt possible even though the results generally made little or no sense. These situations became very frequent in paragraph 5. They usually occurred on multiple syllable and/or low frequency words and involved many of the subjects. Category three miscues were always unacceptable.

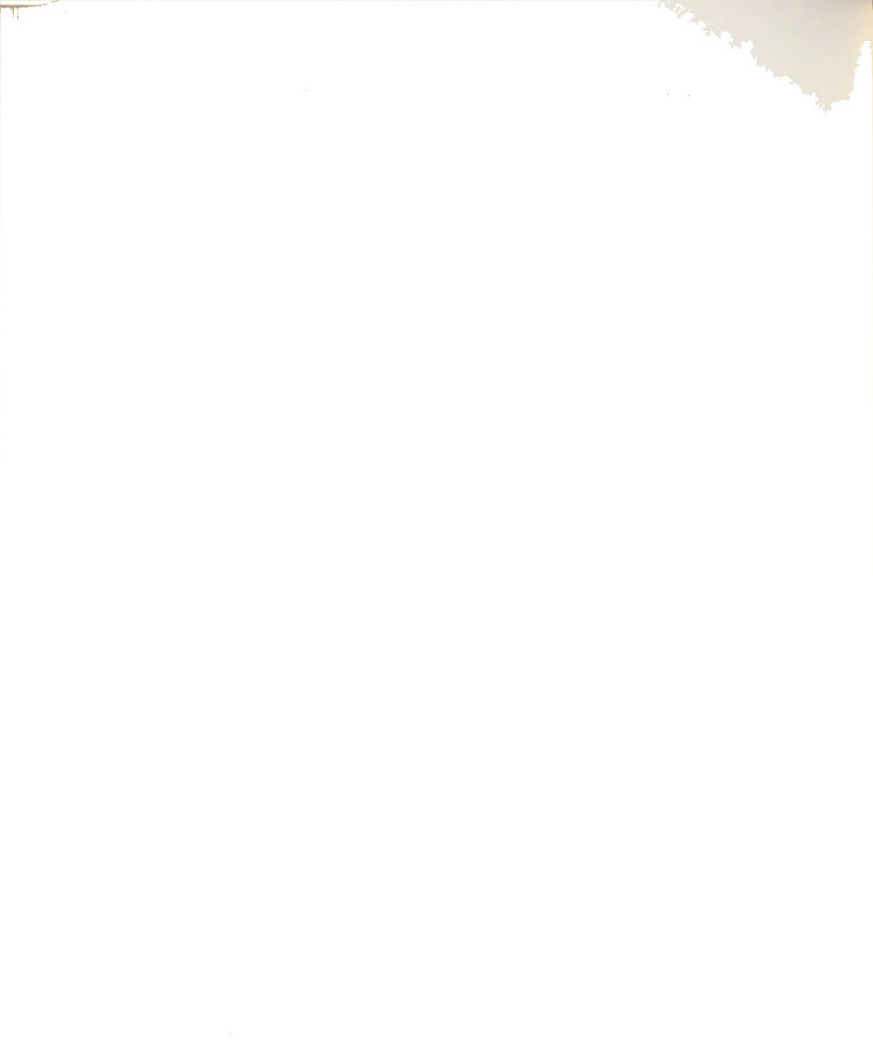
The two words that produced the most miscues, involving 50% of the readers, were the words "uniform" and "uniformed" in paragraph 5. The reader's strategy was almost universally to treat "un" as a prefix, and they either could not abandon this strategy, or they knew of none other. Therefore the "uniformed guard" typically became an "unformed" or "uninformed guard".

Miscues in this category usually involved the insertion, deletion or transposition of a letter or letters to produce another word with a similar visual form even though that word generally made little or no sense. "Facing" often became "facting", "solid" became "soiled", "includes" became "inclouds", "customers" became "costumers" and "armed" became "alarmed". This is not to imply that the subjects in this study had been taught by a sight word method, but simply that they reached a point where the word analysis skills they possessed were no longer adequate to deal with words at this level of complexity.

Summary of Results

The results of this study ultimately took three forms: (a) Statistical analysis of four measures which would seem to be logically associated with passage difficulty - quantity of miscues, rate, quality of miscues, and fluency, (b) graphic analysis of the percentages of students reading at each of the functional reading levels and (c) graphic analysis of miscue frequencies with inspection and descriptive analysis of specific portions of text involved in high frequency of miscue.

Summary of Results from Four Measures of Difficulty Table IV-13 summarizes the significant differences found between paragraphs for the means of (a) word accuracy when all miscues were counted, (b) rate, (c) word accuracy when only



unacceptable miscues were counted, and (d) general impressions of fluency.

Table IV-13
 SUMMARY OF DIFFERENCES
 FOUND BETWEEN PARAGRAPHS
 ON FOUR MEASURES OF DIFFICULTY

	2	3	4	5	
	(1 vs 2)	(1 vs 3)	(1 vs 4)	(1 vs 5)	
1					all miscues
	*X	*X			rate
				X	unacceptable
				X	fluency
2		(2 vs 3)	(2 vs 4)	(2 vs 5)	
					all miscues
		X	X	X	rate
				X	unacceptable
3				X	fluency
			(3 vs 4)	(3 vs 5)	
					all miscues
					rate
4					unacceptable
				X	fluency
				(4 vs 5)	
					all miscues
					rate
				X	unacceptable
					fluency

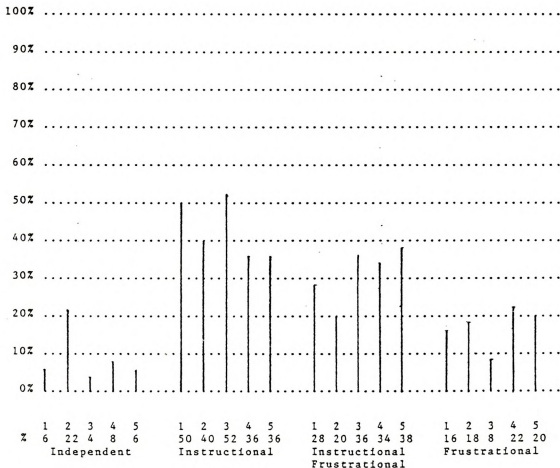
*difference in direction opposite of that expected



As the table indicates, paragraph 5, when contrasted with other paragraphs, showed the greatest number of differences in general. The greatest number of differences in particular occurred between paragraphs 2 and 5. Rate, unacceptable miscues and fluency scores all suggested a definite increase in difficulty between these two paragraphs with paragraph 5 being the most difficult to read and paragraph 2 being the easiest. Other than this, however, there seemed to be very little discrimination of difficulty among paragraphs 1, 3, and 4. There were no significant differences found between any paragraphs when word recognition accuracy was based on quantity of miscues only and the quality of miscues was not considered.

Summary of Results from Functional Reading Levels Table IV-14 presents a graph showing the percentages of students reading at each of the functional reading levels (independent, instructional, instructional-frustrational and frustrational) for each paragraph.

Table IV-14
 PERCENTAGES OF SUBJECTS
 READING AT EACH FUNCTIONAL READING LEVEL
 ON EACH PARAGRAPH



The greatest percentage of students read at their instructional level on the third grade paragraph as expected, however this was also true of all other paragraphs. There seemed to be very little differentiation between paragraphs in terms of the levels at which students were reading. This finding was not surprising, however, when we consider that these levels were established using the traditional Betts' criteria, which in turn was based only on the quantity, and not the quality of miscues. Since there were no differences between paragraphs in terms of quantity of miscue, it would not seem unusual, therefore, to also find no differences between paragraphs in terms of reading levels.

Summary of Results from Miscue Frequency Data The data previously presented in this chapter suggested that miscues did not occur randomly, but rather tended to cluster on certain words and in certain places in a sentence. Further analysis suggested that these miscues could be categorized as either (a) miscues of no consequence, (b) miscues which seemed to be triggered by factors in the text which interfered with the reader's processing strategies, but for which the reader had the capacity to correct, and (c) miscues for which the reader lacked adequate decoding strategies and from which the reader could not recover.

Factors in the text which were associated with high frequency of miscue could be identified. They included



unfamiliar grammatical function, unfamiliar word meanings, unfamiliar phrases and unfamiliar word order. These were not factors traditionally associated with readability procedures, but they were similar to those reported in previous studies of miscue analysis.

CHAPTER V

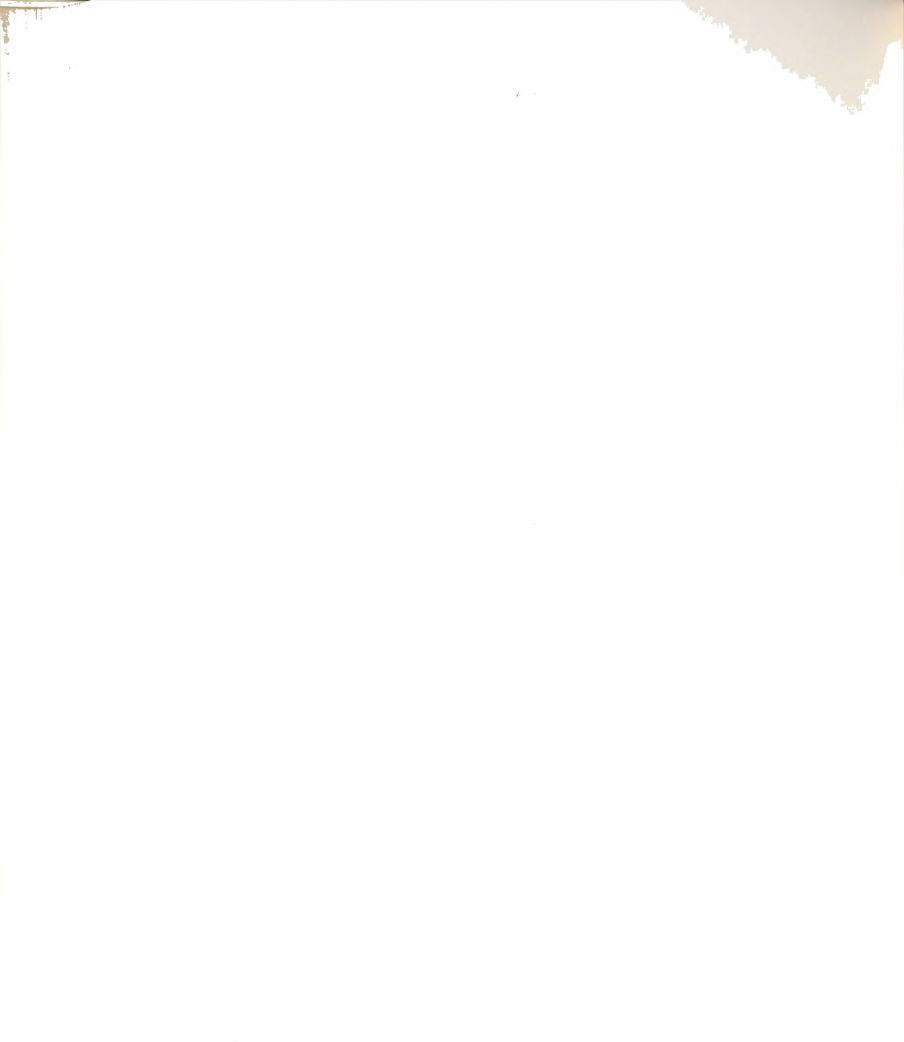
SUMMARY AND CONCLUSIONS

Introduction

In this chapter a summary of the purpose, the design of the study and the findings will be discussed. Conclusions based on the analysis and focusing on the degree to which the study credits or discredits the test-formula matching practice will be presented. Implications for (a) practitioners and (b) further research will be discussed and recommendations for further research will be given.

Summary

The purpose of this study was to investigate how effectively the Reading Test grade equivalency scores from the California Achievement Tests as a measure of student reading achievement, and Fry Readability Graph (1968) scores, as a measure of passage difficulty, would predict the degree of difficulty a given group of students would encounter when reading orally from material of varying Fry determined readabilities. To accomplish this, the subjects selected for the study were all third grade students with Reading test scores, from the California Achievement Tests, falling within a six month range from three months above to three months below their grade level at the time of testing, while the passages selected for them to read had Fry determined readabilities ranging from first to fifth grades.



Thus, the study was designed to hold the reading achievement of the subjects, as indicated by their Reading test scores, relatively stable, while the readability scores of the passages were allowed to vary. If the test scores and the readability data provide an effective means of matching readers with materials of suitable difficulty, we would expect to find very little variation in reading performance from student to student and considerable variation in performance from paragraph to paragraph. Generally speaking this was not the case.

When only quantity of miscues was considered, performance on all paragraphs tended to be very similar. There were not statistically significant differences in the means of the word recognition accuracy scores for each paragraph, and similar percentages of students tended to fall in each of the reading levels categories on all paragraphs. In each case, a small percentage of students, roughly 5%, were reading at their independent level. A slightly larger percentage, between 10% and 20%, were reading at their frustration level. At the instructional level there was a slight differentiation between the first three paragraphs and the last two, with approximately 60% of the readers able to read paragraphs 1, 2 and 3 at their instructional level and about 45% able to read paragraphs 4 and 5 at this level. Roughly, between 80% to 90% of all readers were able to read all paragraphs with 90% word recognition accuracy or better.

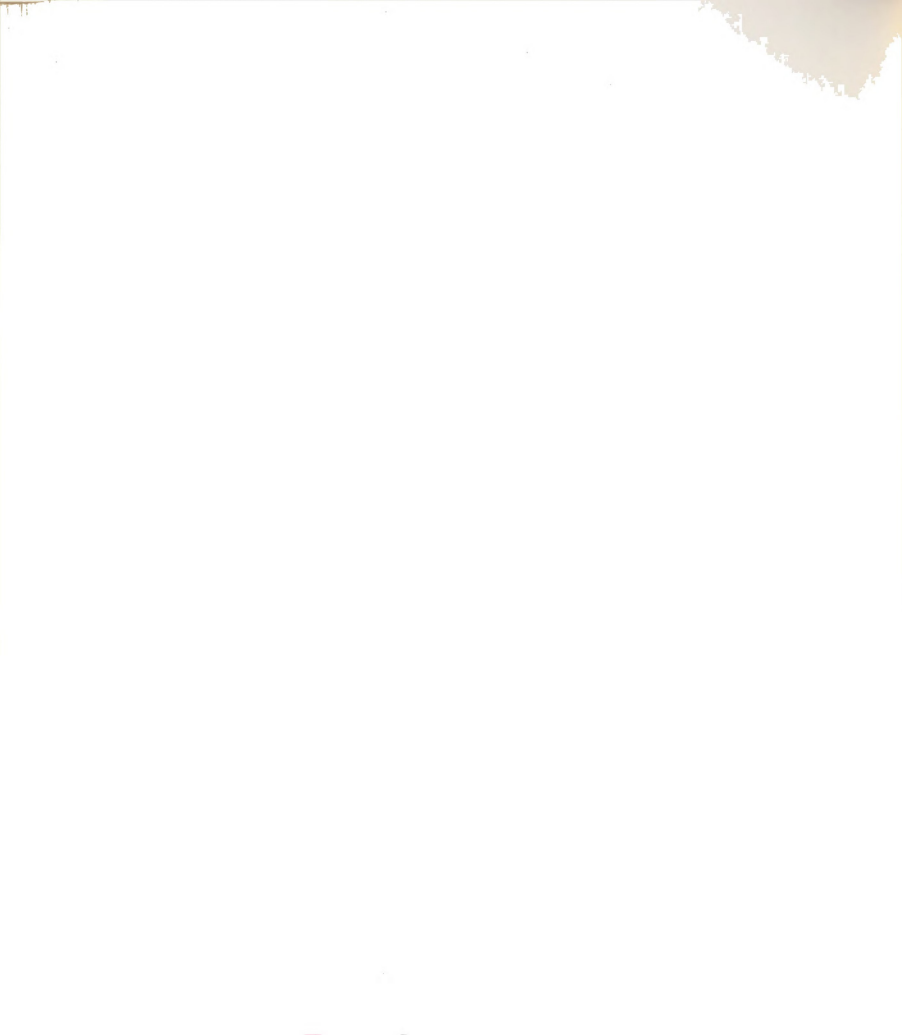
In terms of rate, unacceptable miscues and general impressions of fluency, there were indications that the second grade paragraph was the easiest to read, while the fifth grade paragraph was the most difficult. Differences were found between paragraphs 2 and 5 on all three of these measures. In addition, there were significant differences between paragraph 5 and all other paragraphs when comparing the means of word recognition accuracy scores when only unacceptable miscues were considered. There were significant differences between the means of the reading rate scores for paragraph 2 and paragraphs 1, 3, 4, and 5, with paragraph 2 being read faster in each case. Also, more readers were at their independent reading level when reading paragraph 2.

The difficulty of the other three paragraphs, 1, 3, and 4, appeared to fall somewhere between paragraphs 2 and 5, but there was little to suggest any distinction of difficulty between paragraphs 1, 3, and 4.

Conclusions

In analyzing the findings in this study, the following conclusions were reached.

1. The readability graph seemed to identify material within the reader's general range of ability, but did not seem able to discriminate difficulty as precisely as one or two grade levels.



2. In terms of miscues, readers seemed to encounter similar amounts of difficulty (quantity of miscues) on all paragraphs. However, there was a decided shift in the type of difficulty (quality of miscues) they were experiencing, especially when they reached the fifth grade paragraph.

3. On paragraphs with lower readabilities, the type of difficulty readers seemed to be experiencing appeared to be due to factors in the text which interfered with their prediction strategies. This resulted in many miscues, but they were generally miscues of an acceptable nature, that is miscues that had negligible effect on meaning or miscues for which the reader had the capacity to recover and thereby correct.

4. On paragraphs of higher readability, when readers began to encounter difficult words with which they were unfamiliar and for which they lacked adequate decoding strategies, they did not simply add these miscues to the types of miscues they had previously been making. Instead, the quantity of the miscues tended to stay the same but the quality of the miscues changed, resulting in a lower proportion of acceptable miscues and a greater proportion of requests for aid, gross mispronunciations and other miscues of an unacceptable nature.

5. Because reading levels were established using the traditional Betts' criteria which is based on the quantity of miscues, and not quality, and because the quantity of

miscues tended to remain the same but the quality of miscues changed, the students reading levels did not provide a complete picture of the difficulty they experienced.

6. Sentence length did not appear to be associated with the difficulty the subjects encountered when reading the passages in this study. Word difficulty, however, did seem to have an effect. Whether this was a function of factors measured by the readability graph, or a result of the vocabulary control used in developing the materials could not be determined.

7. The places within a passage where miscues occurred were highly predictable, with many readers miscuing at the same place in a sentence or on the same word, and frequently making the same or a similar response.

8. The factors in a sentence that seemed to be triggering a high number of miscues could be identified. These factors were not those traditionally associated with readability formulae, but they were virtually identical to factors reported by Laura Smith (1976) in previous miscue analysis studies. In fact when factors reported by Smith were not identified as a cause of difficulty for readers in this study, it was simply because the text chosen did not provide an opportunity to observe them. For instance, Smith found that direct quotations caused many readers difficulty. There were no direct quotations in the material used here, so therefore the readers' response to them could not be

observed. However, when there was an opportunity to observe a factor causing difficulty reported by Smith, the similarity of response by readers in this study was uncanny. This finding becomes even more significant when the ten year time gap between the Smith study and this study is considered.

Discussion

At least five major considerations seem to emerge from the conclusions of this study. First, most authorities in the field of reading, including the authors of readability formulae themselves, have repeatedly stressed that such devices should only be used as rough estimates of relative difficulty. The results of this investigation amplify the importance of such admonishments. Furthermore, many readability prediction methods only attempt to assign difficulty in a very general way, such as "below 4th grade" or in terms of "elementary", "high school" or "college" levels. The results of this study would suggest that it may not be possible to predict difficulty with much more precision than these methods have attempted. In addition, it must be noted that the type of difficulty readers experienced on the fifth grade passage in this study was highly associated with vocabulary load. Since the materials used in this study were specifically designed for instructional use, it cannot be determined if the difficulty was due to factors measured by the readability graph or a

function of the vocabulary control used in developing the materials. Only a replication of this study using selections from children's literature and trade books which do not use strict vocabulary controls, could ascertain if the readability graph even predicted general areas of difficulty.

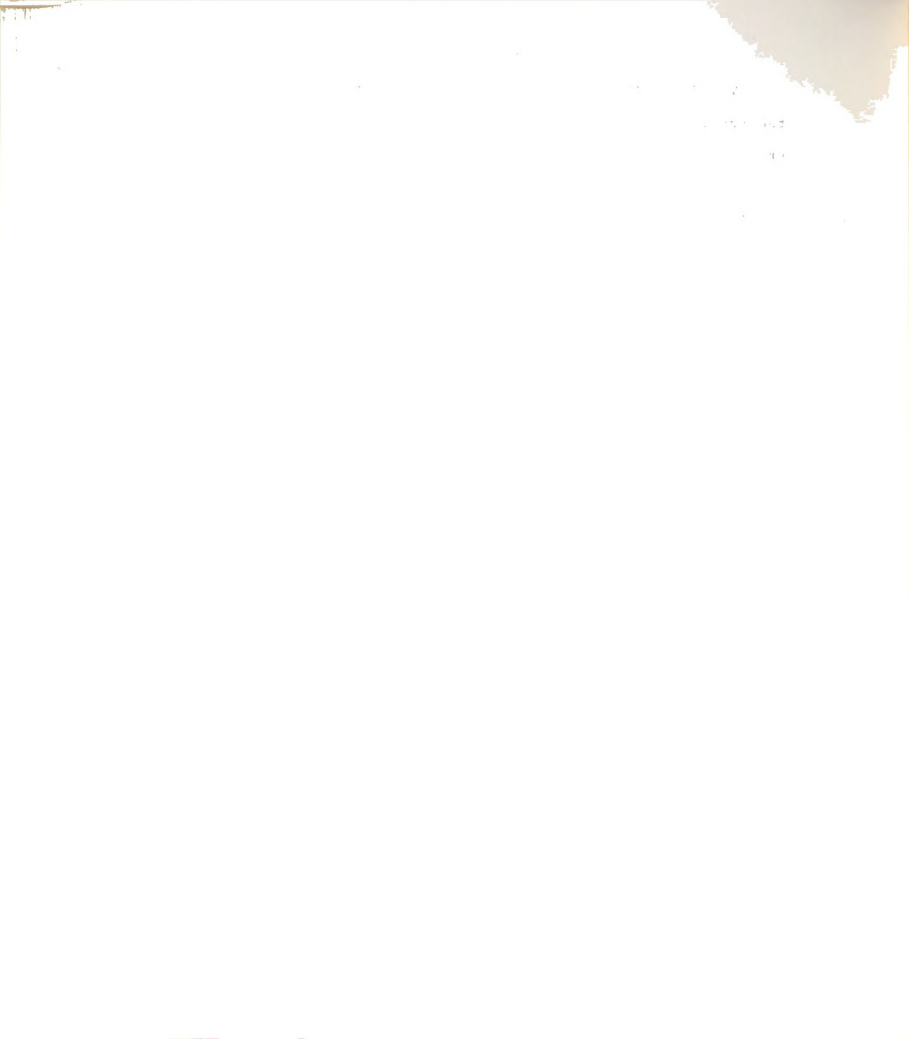
Secondly, the shift in quality of miscues, but not quantity of miscues, strongly suggests that readers changed their processing strategies when they encountered unknown words. When the words in the text were very familiar, the readers seemed to use top-down processing strategies, relying on language and meaning cues to direct their reading. Factors in the text or in the reader's ability to deal with those factors, however, seemed to repeatedly interfere with their prediction strategies, often causing many miscues, although the reader was usually able to reprocess the material and recover. When the readers began to encounter unknown words, however, they were forced to use word analysis methods and thus shift to bottom-up strategies. In doing so, they had to attend more closely to the grapho-phonetic cues in the writing. This would explain why there were fewer miscues on familiar words when requests for aid, gross mispronunciations and other unacceptable miscues, caused by unfamiliar words, increased. Such an explanation would be consistent with and supportive of an interactive-compensatory model of the reading process.



Third, this study raises serious questions concerning the traditional use of the Betts' criteria in establishing a student's functional reading levels. Such criteria assumes that, as material becomes more difficult, readers will simply make more miscues, and does not provide for a change in quality of miscue, rather than quantity. In this connection the effect of silent prereading on the quality and quantity of oral reading miscues needs to be examined further. It is possible that prereading allows the reader to work out miscues of an acceptable nature, so that only more serious miscues appear in the subsequent oral reading. Under these conditions quantity of miscues might then be more closely associated with the difficulty readers actually experienced. This might also be accomplished in unprepared oral reading, by classifying miscues and giving them weighted scores based on their seriousness.

Fourth, the uncanny and totally unanticipated similarity between miscue analysis findings in this study and those reported ten years earlier by Laura Smith (1976), suggests that there may be some universal miscue patterns characteristic at various stages of reading development. Knowledge of these patterns could be of great use to those writing for readers of various ages and grade levels.

Finally, because miscues did not occur randomly, and because it was possible to identify factors in the text which seemed to be causing many miscues, it would also seem that, if oral reading was used in developing the criterion



for the prediction method, it would be possible to develop a procedure that would better predict oral reading performance. The process of matching readers with material might be made even more reliable if oral reading were also used to measure the reading achievement of the student.

Implications

The results of this study should clearly demonstrate to reading practitioners as well as authors and publishers, that the usefulness of current readability prediction methods is probably very limited. The study largely discredits the notion that such devices can be used to place students in materials within one, two or even three grade levels of their reading achievement, at least for readers of this age. At the very best, it appears that such procedures may only be able to identify a general area of difficulty such as "primary", "elementary", "high school" or "college" level. Furthermore, since the discernable difficulty in this study appeared to be closely associated with vocabulary load, and since the materials involved were specifically written for classroom use, it cannot be determined if the difficulty readers experienced was due to something measured by the readability graph or a function of the vocabulary control used in developing the materials. Therefore, in the final analysis, the use of readability prediction devices, either to place students in material or to check passage difficulty when writing for a specific audience, appears to

remain a basically unsupported practice.

The results of this study do support previous research which has found identifiable and highly predictable factors in writing which seem to cause difficulty for young readers, although they are not factors generally measured by readability formulae. Knowledge of the effect of these factors, such as word order, unfamiliar word meanings and usage, and unfamiliar phrases, might aid practitioners and authors in the selection and writing of material for children. In this regard, there are many questions raised by the reader's performance in this study which have implications for further research. First of all, at what point do readers develop the abilities necessary to read material of the type used in this study, with speed and virtual perfection? Do readers at this stage of reading development, even have an independent reading level or are they still unable to read any material with the fluency the independent level implies? Are their miscues due to prior instructional practices or the inclusion of material in their basal reading series which does not expose them to the situations which caused them difficulty in this material, or are their miscues a natural part of any child's reading development? Research guided by these questions could provide valuable assistance to those selecting and writing materials for young readers, as well as providing further insights into the developmental stages involved in learning to read.

In addition, while the reading levels concept appears to be a useful one, the results of this study suggest that as reading material becomes more difficult, readers do not simply make more miscues, but instead make miscues of a more serious nature. Therefore, better criteria for determining functional reading levels, criteria which considers quality of miscues as well as quantity, needs to be developed.

Recommendations

As this investigation progressed many questions arose which suggest recommendations for further research. Such additional studies might answer questions which resulted from limitations in this study and might also extend the scope of this investigation further.

1. The study could be repeated using passages from children's literature or trade books which have not been developed specifically for instructional use. This might help to determine if the difficulty encountered on the fifth grade paragraph was due to factors measured by the readability graph, or a happenstance of the vocabulary control used in developing the materials.

2. The study could be repeated using new versions of the passages from this study, rewritten to eliminate the factors which appeared to be causing many miscues. This would help to determine if controlling these factors would, in fact, make the material easier for children of this age



to read.

3. Repeating the study with older and younger children could provide valuable information concerning the development of children's reading proficiency.

4. Repeating the study with children who have been given specific instruction and practice with material containing features which seemed to cause a high incident of miscues in this study, might help determine if these miscues are the result of previous instructional practices and experiences, or if they are a natural part of reading development.

5. Repeating the study, but giving the readers the opportunity to preread the passages silently before oral reading might provide valuable information concerning the effect of silent prereading on subsequent oral reading miscues and their relationship to passage difficulty.

6. Data from the present study could be reanalyzed using a classification and weighting system for miscues. This might help to determine if reading levels based on such a procedure would be more closely associated with the other indicators of passage difficulty (rate, unacceptable miscues and general fluency) observed in this investigation.

7. Finally, in a more general sense, it would appear that, since oral reading performance is frequently used as a measure of readability, more valid methods for predicting



readability, especially for young readers, could be developed if oral reading were used to rank the difficulty of the criterion passages. The use of oral reading in the development of new readability prediction methods, therefore, is worthy of research attention.

APPENDIX A

Letter from Principals to Parents

Parental Permission Slip



Bay City Public Schools

910 N. Walnut Street

Bay City, Michigan 48706

Dear Parent:

Currently one of our Chapter I Reading Teachers, Janet Dixon, is working on a study concerning readability formulas as part of her doctoral program at Michigan State University. These formulas claim to predict the difficulty of reading materials, however their usefulness is debatable. It is the purpose of Janet's study to listen to children read material which the formulas say vary in difficulty, and then see if the children will actually make more errors on the more difficult selections.

Your child has been selected as a possible subject in this study. In order to participate your permission will be necessary. Hopefully, the following information will reassure you and make you feel more comfortable about giving that permission.

If your child participates, he will be asked to do two things. In the first session he will take the Reading Test of the California Achievement Tests. If he is involved in the second session, he will be asked to read aloud list of words and five paragraphs which will be tape recorded for later analysis. All participants will take the test but not all will read the paragraphs. It will take about 45 minutes to complete the test and about 15 minutes to read the paragraphs.

As a subject, your child will be given a code name. Only the researcher (Janet) will have a list of the code names and this list will be destroyed once data collection is completed. You may know your own child's code name, but you must ask for it at the time of data collection. Once the list is destroyed there will be no way for anyone, even the researcher, to identify your child in the study.

Your child will not be used in the study unless he is a willing participant. The task involved is not a difficult one and should not cause any undue distress. Your child will be given continuous support and encouragement by the researcher throughout the project and may discontinue at any time if he, or the researcher, feels the situation is too stressful. Such situations will be handled carefully to make sure the child feels positive about the experience even if he decides to decline or discontinue, and there will be no penalty for such a decision.

In order to help Janet complete the list of subjects, please return the attached permission slip to your child's teacher as soon as possible. Return the slip even if you decide not to have your child participate. This will make a follow-up letter unnecessary.

If you have any further questions, Janet or I will be more than happy to discuss them with you. You may contact us at the following numbers:

Janet Dixon

Home: _____
_____ Elementary Center: _____

_____, Principal
_____ Elementary School: _____

If we are not in, please leave your name with the secretary and we will return your call.

Thank you for taking the time to read this letter, for giving this matter your consideration and for returning the permission slip. Of course, the most important thing in this study will be the children who participate. We are hoping your child will be among them.

Sincerely,

_____, Principal
_____ Elementary School



To Whom It May Concern:

My child _____
has my permission to participate in the study being conducted by Janet
Dixon concerning formulas used to predict the difficulty of reading
materials.

Parent's

Signature _____

Do you wish to know your child's code name? Yes ___ No ___

If you do not wish to have your child participate, please check here: ___
(A signature is not necessary in this case)

Please return this entire sheet to your child's teacher.

APPENDIX B

The Research Passages

HOW IS THE AIR TODAY?

Something is all around you at all times. You cannot see it. But sometimes you can feel it. Without it nothing can live. Do you know what it is? It is air. You know it is around you when it blows hard. You can feel it then. Sometimes you can feel air from a balloon. Blow up a balloon. Then let go of its mouth. Can you feel the air as it comes out?

Many things use air. Tires, windmills and footballs use air. Sailboats are pushed by air. Kites are kept up by air. Air helps to keep airplanes in the sky too.



A ROCK FENCE

Long ago many large rocks lay all over the ground. There was a farmer who wanted to grow things on the land. But nothing would grow where the rocks were. So he started picking up the rocks. He carried them to the sides of the field. He made a fence of the rocks. Then all the other farmers could see where his field was. Flowers grew along the rock fence. At first the rocks had been in the way. But soon they helped the farmer. And the farmer's rock fence made the field more beautiful.



WANT TO TRADE A TIGER?

Let's pretend you're running a zoo. In your zoo you have four tigers but only one polar bear. You're lucky to have the tigers. Very few tigers are born in zoos but two were born in your zoo a year ago. But you're unlucky to have only one polar bear. It isn't much fun for people who come to your zoo to watch one lonely polar bear. If you know how to run your zoo, you'll look around for a zoo that wants a tiger. Maybe you can trade for a polar bear. Like every zoo man you have trading in your blood.



FUN WITH MAGNETS

Maybe you already know how magnets work. If you were to hold a magnet near a paper clip on your desk, you'd know what to expect. When the magnet got close, the paper clip would suddenly flip over and stick to the magnet. This happens because paper clips are made of iron. And anything made of iron sticks to magnets. What would happen if you used a piece of paper instead of a clip? Have you ever tried that? Maybe not, but you know that there is no iron in paper. And because of this you feel sure that paper will not stick to the magnet.



BANKS ARE INTERESTING PLACES

The next time you are going into a bank stop a minute before you push open the heavy door. Look at the building. If it is an old bank there will be only a few windows facing the street or maybe none at all. The front of the bank will seem a solid stone wall. Through the door you may see a bank guard whose uniform includes a gun. If your bank is in a new building, there may be huge windows. Through them you can see the bank's workers and customers as well as a uniformed guard who is not armed.



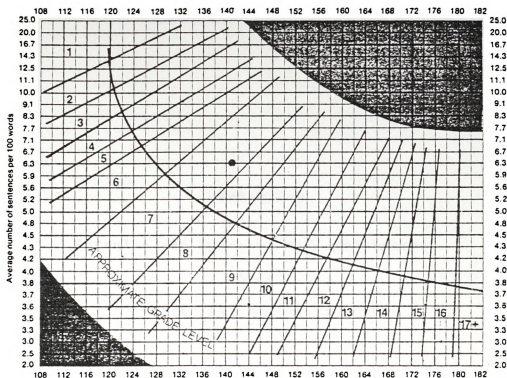
APPENDIX C

The Fry Readability Graph



GRAPH FOR ESTIMATING READABILITY — EXTENDED

Average number of syllables per 100 words



DIRECTIONS: Randomly select 3 one hundred word passages from a book or an article. Plot average number of syllables and average number of sentences per 100 words on graph to determine the grade level of the material. Choose more passages per book if great variability is observed and conclude that the book has uneven readability. Few books will fall in gray area but when they do grade level scores are invalid.

Count proper nouns, numerals and initializations as words. Count a syllable for each symbol. For example, "1945" is 1 word and 4 syllables and "IRA" is 1 word and 3 syllables.

EXAMPLE:

	SYLLABLES	SENTENCES
1st Hundred Words	124	6.6
2nd Hundred Words	141	5.5
3rd Hundred Words	158	6.8
AVERAGE	141	6.3

READABILITY 7th GRADE (see dot plotted on graph)

For further information and validity data, see Edward Fry, "Fry's Readability Graph: Clarifications, Validity, and Extension to Level 17," *Journal of Reading* (December 1977).

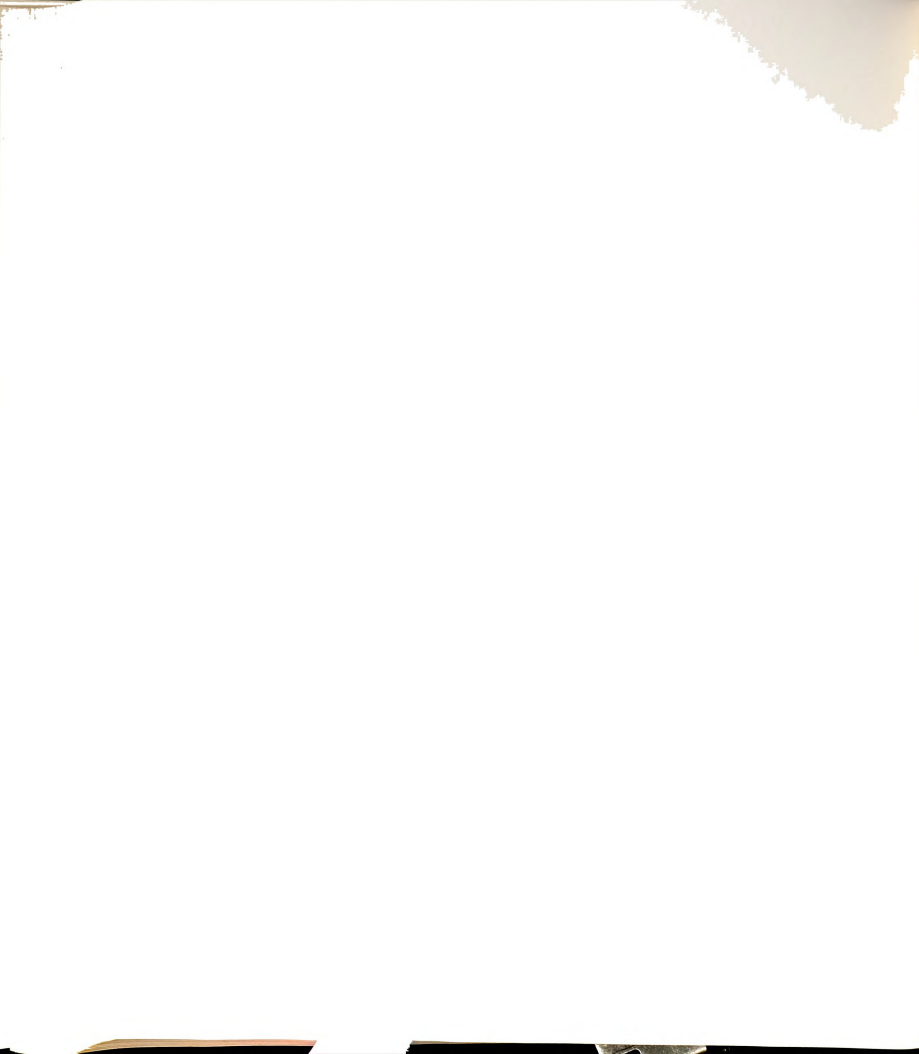
Computational Procedures

APPENDIX D

Formulae and Computational Procedures

ANOVA

Scheffe Post-Hoc Comparisons



Formula and Computational Procedures

ANOVA

The following computational procedure for Analysis of Variance for Single Factor Experiments with Repeated Measurements of the Same Elements was used in this study to determine if mean differences did exist. The procedure was taken from Winer (1971, p. 268).

K=number of treatments X=an individual score
 n=number of subjects in a treatment group P=the sum of scores of all treatments for one subject
 T=the sum of all scores for one treatment G=the sum of all scores for all treatments
 subscript j=all treatment groups (1 to 5) subscript i=all subjects (1 to 50)

$$I = G^2 / Kn \quad II = \sum X^2 \quad III = (\sum T)^2 / n \quad IV = (\sum P)^2 / K$$

Source of Variation	SS (Sum of Squares)	df (Degrees of Freedom)
Between People	SSB = IV - I	n-1
Within People	SSW = II - IV	n(K-1)
Treatments	SST = III - I	K-1
Residual	SSR = II - III - IV + I	(n-1)(K-1)
Total	SSTO = II - I	Kn-1

$$MST = SST / df = SST / (K-1)$$

$$MSR = SSR / df = SSR / ((n-1)(K-1))$$

$$F = \frac{MST}{MSR}$$

The critical value for the F ratio is taken from the tables for the F Distribution for K-1 and (n-1)(K-1) degrees of freedom. A significance level of .05 was used in this study. If the computed F value exceeded the critical F value, the null hypothesis was rejected and it was assumed that there were differences in the means.



Formula and Computational Procedures

Scheffe Test for Post-Hoc Comparisons

When analysis of variance indicated mean differences did exist, the Scheffe Test for Post-hoc comparisons was used to determine where differences occurred. The formula and computational procedures used were taken from Hinkle, Wiersma and Jurs (1979, p. 276-280). When used with the ANOVA for repeated measures, MSR takes the place of MSW (Winer, 1971, p. 270).

The formula used for each set of contrasts was

$$F = \frac{(M1 - M2)^2}{MSR (1/n1 + 1/n2)}$$

where M1 = the means of the first contrast, n1 = the number of scores in that mean, M2 = the mean of the second contrast and n2 = the number of scores in the second contrast.

The critical value for F used in the Scheffe is the critical value used in the ANOVA multiplied by K-1 where K is the number of groups. Therefore the critical value for F used in the Scheffe tests in this study was (2.45)(4) = 9.8.



APPENDIX E

Summary of Computations

Word Recognition Accuracy Scores
Based on Total Number of Miscues

Reading Rate Scores

Scheffe Post-Hoc Comparisons
Reading Rate Scores

Word Recognition Accuracy Scores
When Only Unacceptable Miscues Were Counted

Scheffe Post-Hoc Comparisons
Word Recognition Accuracy Scores
When Only Unacceptable Miscues Were Counted

General Impression of Fluency Scores

Scheffe Post-Hoc Comparisons
General Impression of Fluency Scores



Summary of Computation
for
Word Recognition Accuracy Scores
Based on Total Number of Miscues

Totals	T1	T2	T3	T4	T5
	4696	4720	4718	4690	4646
Means	93.92	94.4	94.36	93.8	92.92

$$G = 23470 \quad \Sigma \Sigma X^2 = 2207554 \quad \Sigma P^2 = 11026292 \quad K=5 \quad n=50$$

$$I = G^2 / Kn = (23470)^2 / 250 = 2203363.44$$

$$II = \Sigma \Sigma X^2 = 2207554$$

$$III = (\Sigma T)^2 / n = 110171755 / 50 = 2203435.1$$

$$IV = (\Sigma P)^2 / K = 11026292 / 5 = 2205258.4$$

Source of Variation	SS (Sum of Squares)	df (Degrees of Freedom)
SSB (people)	= IV-I = 1894.96	n-1 = 49
SSW (people)	= II-IV = 2295.6	n(K-1) = 200
SST (treatment)	= III-I = 71.66	K-1 = 4
SSR (residual)	= II-III-IV+I = 2223.94	(n-1)(K-1) = 196
SSTO (total)	= II-I = 4190.56	(Kn)-1 = 249

$$MST = SST / K - 1 = 71.66 / 4 = 17.915$$

$$MSR = SSR / (n-1)(K-1) = 2223.94 / 196 = 11.3466326$$

$$F = \frac{MST}{MSR} = \frac{17.915}{11.3455326} = 1.57888252$$

$$\text{Critical } .05 F (4, 196) = 2.45$$

F < Critical F; Therefore accept null.
Assume there are no differences.

Summary of Computation
for
Reading Rate Scores

Totals	T1	T2	T3	T4	T5
	4648.8	5546.5	4999.9	4908.6	4717.9
Means	93.976	110.93	99.998	98.172	94.358

$$G = 24871.7 \quad \sum \sum X^2 = 2588508.22 \quad \sum P^2 = 12808179.5 \quad K=5 \quad n=50$$

$$I = G / Kn = (24871.7) / 250 = 474405.82$$

$$II = \sum \sum X^2 = 2588508.22$$

$$III = (\sum T^2) / n = 124194314 / 50 = 2483886.28$$

$$IV = (\sum P^2) / K = 12808179.5 / 5 = 2561635.91$$

Source of Variation	SS (Sum of Squares)	df (Degrees of Freedom)
SSB (people)	= IV-I = 87230.09	n-1 = 49
SSW (people)	= II-IV = 26872.31	n(K-1) = 200
SST (treatment)	= III-I = 9480.46	K-1 = 4
SSR (residual)	= II-III-IV+I = 17391.85	(n-1)(K-1) = 196
SSTO (total)	= II-I = 114102.4	(Kn)-1 = 249

$$MST = SST / K - 1 = 9480.46 / 4 = 2370.115$$

$$MSR = SSR / (n-1)(K-1) = 17391.85 / 196 = 88.7339285$$

$$F = \frac{MST}{MSR} = \frac{2370.115}{88.7339285} = 26.710358$$

$$\text{Critical } .05 F (4, 196) = 2.45$$

F > Critical F; Therefore reject null.
Assume there are differences.



Summary of Computation
for
Scheffe Post-Hoc Comparisons
Reading Rate Scores

$n_1=n_2=n_3=n_4=n_5=50$ $MSR = 88.734$ $K=5$

M_1 =First mean to be contrasted
 M_2 =Second mean to be contrasted

Critical F (from ANOVA) = 2.45
Critical F for Scheffe Test = $2.45(K-1) = 9.8$

$$F = \frac{(M_1 - M_2)^2}{MSR (1/50 + 1/50)}$$

	Contrasts	Means	Computed F	Significance
Paragraph	1 vs. 2	*93.976-110.93	80.991	*yes
	1 vs. 3	*93.976-99.998	10.249	*yes
	1 vs. 4	*93.976-98.172	4.961	no
	1 vs. 5	*93.976-94.358	.041	no
	2 vs. 3	101.93-99.998	33.674	yes
	2 vs. 4	101.93-98.172	45.862	yes
	2 vs. 5	101.93-94.358	77.383	yes
	3 vs. 4	99.998-98.172	.939	no
	3 vs. 5	99.998-94.358	8.963	no
	4 vs. 5	98.172-94.358	4.0986	no

*indicates contrasts in which the first element is smaller than the second. If significance is found it suggests that differences existed between the means but in a direction opposite of that which would be expected.



Summary of Computation
for
Word Recognition Accuracy Scores
When Only Unacceptable Miscues Were Counted

Totals	T1	T2	T3	T4	T5
	4964	4979	4951	4957	4868
Means	99.28	99.58	99.02	99.14	97.36

$$G = 24719 \quad \sum \sum X^2 = 244795 \quad \sum P^2 = 12221529 \quad K=5 \quad n=50$$

$$I = G^2 / Kn = (24719)^2 / 250 = 2444115.74$$

$$II = \sum \sum X = 2444795$$

$$III = (\sum T)^2 / n = 122213408 / 50 = 2444268.16$$

$$IV = (\sum P)^2 / K = 12221529 / 5 = 2444305.8$$

Source of Variation	SS (Sum of Squares)	df (Degrees of Freedom)
SSB (people)	= IV-I = 190.06	n-1 = 49
SSW (people)	= II-IV = 489.2	n(K-1) = 200
SST (treatment)	= III-I = 152.42	K-1 = 4
SSR (residual)	= II-III-IV+I = 336.78	(n-1)(K-1) = 196
SSTO (total)	= II-I = 679.26	(Kn)-1 = 249

$$MST = SST / K - 1 = 152.42 / 4 = 38.105$$

$$MSR = SSR / (n-1)(K-1) = 336.78 / 196 = 1.7182653$$

$$F = \frac{MST}{MSR} = \frac{38.105}{1.7182653} = 22.1764357$$

$$\text{Critical } .05 F (4, 196) = 2.45$$

F > Critical F; Therefore reject null.
Assume there are differences.



Summary of Computation
for
Scheffe Post-Hoc Comparisons
Word Recognition Accuracy Scores
When Only Unacceptable Miscues Were Counted

$n_1=n_2=n_3=n_4=n_5=50$ $MSR = 1.718$ $K=5$

M_1 =First mean to be contrasted

M_2 =Second mean to be contrasted

Critical F (from ANOVA) = 2.45

Critical F for Scheffe Test = $2.45(K-1) = 9.8$

$$F = \frac{(M_1 - M_2)^2}{MSR (1/50 + 1/50)}$$

	Contrasts	Means	Computed F	Significance
Paragraph 1	vs. 2	*99.28-99.58	1.304	no
	1 vs. 3	99.28-99.02	.9797	no
	1 vs. 4	99.28-99.14	.284	no
	1 vs. 5	99.28-97.36	53.426	yes
	2 vs. 3	99.58-99.02	4.545	no
	2 vs. 4	99.58-99.14	2.8057	no
	2 vs. 5	99.58-97.36	71.426	yes
	3 vs. 4	*99.02-99.14	.2087	no
	3 vs. 5	99.02-97.36	39.936	yes
	4 vs. 5	98.14-97.36	45.918	yes

*indicates contrasts in which the first element is smaller than the second. If significance is found it suggests that differences existed between the means but in a direction opposite of that which would be expected.

Summary of Computation
for
General Impression of Fluency Scores

Totals	T1	T2	T3	T4	T5
	116	104.5	114	117	130
Means	2.32	2.09	2.28	2.34	2.6

$$G = 581.5 \quad \sum X^2 = 1485.249 \quad \sum P^2 = 7209.24987 \quad K=5 \quad n=50$$

$$I = \frac{G^2}{Kn} = 1352.56699$$

$$II = \frac{\sum X^2}{n} = 1485.2499$$

$$III = \frac{(\sum T)^2}{n} = 1359.22499$$

$$IV = \frac{(\sum P)^2}{K} = 1441.84997$$

Source of Variation	SS (Sum of Squares)	df (Degrees of Freedom)
SSB (people)	= IV-I = 89.282977	n-1 = 49
SSW (people)	= II-IV = 43.399926	n(K-1) = 200
SST (treatment)	= III-I = 6.657997	K-1 = 4
SSR (residual)	= II-III-IV+I = 36.741929	(n-1)(K-1) = 196
SSTO (total)	= II-I = 132.682903	(Kn)-1 = 249

$$MST = SST/K-1 = 6.657997/4 = 1.66449925$$

$$MSR = SSR/(n-1)(K-1) = 36.741929/196 = .187458821$$

$$F = \frac{MST}{MSR} = \frac{1.66449925}{.187458821} = 8.87927939$$

$$\text{Critical } .05 F (4, 196) = 2.45$$

F > Critical F; Therefore reject null.
Assume there are differences.

Summary of Computation
for
Scheffe Post-Hoc Comparisons
General Impression of Fluency Scores

$n_1=n_2=n_3=n_4=n_5=50$ $MSR = .18746$ $K=5$

M_1 =First mean to be contrasted
 M_2 =Second mean to be contrasted

Critical F (from ANOVA) = 2.45
Critical F for Scheffe Test = $2.45(K-1) = 9.8$

$$F = \frac{(M_1 - M_2)^2}{MSR (1/50 + 1/50)}$$

	Contrasts	Means	Computed F	Significance
Paragraph 1	vs. 2	*2.32-2.09	7.0721925	no
1	vs. 3	*2.32-2.28	.2139	no
1	vs. 4	2.32-2.34	.0535	no
1	vs. 5	2.32-2.6	10.481283	yes
2	vs. 3	2.09-2.28	4.826	no
2	vs. 4	2.09-2.34	8.3556	no
2	vs. 5	2.09-2.6	34.7727	yes
3	vs. 4	2.28-2.34	.48128	no
3	vs. 5	2.28-2.6	13.689	yes
4	vs. 5	2.34-2.6	9.037	no

*indicates contrasts in which the first element is larger than the second. If significance is found it suggests that differences existed between the means but in a direction opposite of that which would be expected.



REFERENCES



REFERENCES

- Allen, P.D. (1976). The Miscue Research Studies. In P.D. Allen & D.J. Watson (Eds.), Findings of research in miscue analysis: Classroom implications. Urbana, IL: National Council of Teachers of English, ERIC Clearinghouse on Reading and Communication Skills.
- Allington, R.L. (1984). Oral Reading. In D.P. Pearson (Ed.), Handbook of reading research. New York: Longman.
- Anderson, R.C., & Faust, G.W. (1973). Educational psychology: The science of instruction and learning. New York: Dodd, Mead & Co.
- Ausubel, D.R. (1968). Educational psychology: A cognitive view. New York: Holt, Rinehart & Winston.
- Bader, L.A. (1980). Reading diagnosis and remediation in classroom and clinic. New York: Macmillan.
- Baker, R.G. (1942). Success and failure in the classroom. Progressive Education, 19, 221-224.
- Baker, R.G., Dembo, T., & Lewin, K. (1941). Frustration and regression: An experiment with young children (University of Iowa Studies in Child Welfare, No. 1.). Ames: University of Iowa.
- Berliner, D.C. (1981). Academic learning time and reading achievement. In J.T. Guthrie (Ed.), Comprehension and teaching: Research reviews. Newark, DE: International Reading Association.
- Bernard, H.W. (1965). Psychology of learning and teaching. New York: McGraw Hill.
- Bernard, J. (1966). Achievement test norms and time of year of testing. Psychology in the Schools, 3, 273-275.
- Betts, E.A. (1940). Reading problems at the intermediate grade level. Elementary School Journal, 15, 737-746.
- Betts, E.A. (1946). Foundations of reading instruction. New York: American Book Co.



- Block, J.H., & Anderson, L.W. (1975). Mastery learning in classroom instruction. New York: Macmillan.
- Block, J.H. (Ed.). (1971). Mastery learning: Theory and practice. New York: Holt, Rinehart & Winston.
- Bloom, B.S. (1976). Human characteristics and school learning. New York: McGraw-Hill.
- Bloom, B.S. (Ed.). (1956). Taxonomy of educational objectives, handbook I: Cognitive domain. New York: Longman.
- Bloomer, R.H. (1959). Level of abstraction as a function of modifier load. Journal of Educational Research, 52, 269-272.
- Borg, W.R., & Gall, M.D. (1979). Educational research: An introduction, (3rd ed.) New York: Longman.
- Bormuth, J.R. (1969). Development of readability analyses (Final Report, Project No. 7-0052, Contract No. 1 OEC-3-7-070052-0326). Washington, DC: USOE Bureau of Research, HEW.
- Bormuth, R.C. (1966). Readability: A new approach. Reading Research Quarterly, 1 (3), 79-132.
- Bormuth, J.R. (1975). The cloze procedure: Literacy in the classroom. In W.D. Page (Ed.), Help for the reading teacher: New directions in research. Urbana, IL: National Conference on Research in English, ERIC Clearinghouse on Reading and Communication Skills, National Institute of Education.
- Bradley, J.M., & Ames, W.S. (1976). The influence of intrabook readability variation on oral reading performance. Journal of Educational Research, 70, 101-105.
- Bradley, J.M., & Ames, W.S. (1977). Readability parameters of basal readers. Journal of Reading Behavior, 9, 195-183.
- Brecht, R.D. (1977). Testing format and instructional level with the informal reading inventory. Reading Teacher, 31, 57-59.
- Britton, G., & Lumpkin, M. (1977). A consumer's guide on readability: Ginn and Company, Ginn Reading 720. Corvallis, OR: G. Britton & Associates.

- Britton, J.E., & Danielson, W.A. (1958). A factor analysis of language elements affecting readability. Journalism Quarterly, 35, 420-426.
- California achievement tests: Norm tables. (1977). Monterey, CA: McGraw Hill.
- Carlson, R. (1980, April). Reading level difficulty. Creative Computing, 60-61.
- Carroll, J.B. (1963). A model of school learning. Teachers College Record, 64, 723-733.
- Carver, R.P. (1975-1976). Measuring prose difficulty using the rauding scale. Reading Research Quarterly, 11, 660-685.
- Carver, R.P. (1974). Improving reading comprehension: Measuring readability (Final Report, Contract No. N00014-72-C0240, Office of Naval Research). Silver Spring, MD: American Institute for Research.
- Caylor, J.S., Sticht, T.G., Fox, L.C., & Ford, J.P. (1973). Methodologies for determining reading requirements of military occupational specialties (Tech. Rep. No. 73-75, HumRRO Western Division). Presidio of Monterey, CA: Human Resources Research Organization.
- Chall, J.S. (1958). Readability: An appraisal of research and application. Columbus: The Bureau of Educational Research, Ohio State University.
- Christie, J.F., & Alonso, P.A. (1980). Effects of passage difficulty on primary-grade children's oral reading error patterns. Educational Research Quarterly, 5, 41-49.
- Christie, J.F. (1981). The effects of grade level and reading ability on children's miscue patterns. Journal of Educational Research, 74, 419-423.
- Coke, E.U. (1974). The effects of readability on oral and silent reading rates. Journal of Educational Psychology, 66, 406-409.
- Coleman, E.B. (1965). On understanding prose: Some determiners of its complexity (NSF Final Report GB-2604). Washington, DC: National Science Foundation.
- Cooper, J.L. (1952). The effect of adjustment of basal reading materials on achievement. Unpublished doctoral dissertation, Boston University, Boston, MA.



- Criscoe, B.L., & Gee, T.C. (1984). Content reading: A diagnostic prescriptive approach. Englewood Cliffs, NJ: Prentice-Hall.
- Cunningham, P. (1976). ARRF: A book that fits!. The Reading Teacher, 30, 206-207.
- Cunningham, P., Arthur, S., & Cunningham, J. (1977). Classroom reading instruction K-5: Alternative approaches. Lexington, MA: D.C. Heath & Co.
- Dale, E., & Chall, J.S. (1948). A formula for predicting readability. Educational Research Bulletin, 27, 11-20, 37-54.
- Dale, E., & Tyler, R.W. (1934). A study of the factors influencing the difficulty of reading materials for adults of limited reading ability. Library Quarterly, 4, 384-412.
- Danielson, W.A., & Bryan, S.D. (1963). Computer automation of two readability formulas. Journalism Quarterly, 39, 201-206.
- Daw, S.E. (1938). The persistence of errors in oral reading in grades four and five. Journal of Educational Research, 32, 81-90.
- DeBeaugrande, R. (1981). Design criteria for process models of reading. Reading Research Quarterly, 16, 261-315.
- DeCecco, J.P. (1968). The psychology of learning and instruction: Educational psychology. Englewood Cliffs: NJ: Prentice-Hall.
- Duffy, G.B., & Durrell, D.D. (1935). Third grade difficulties in oral reading. Education, 56, 37-40.
- Dunkeld, C.G. (1970). The validity of the informal reading inventory for the designation of instructional reading levels: A study of the relationship between children's gains in reading achievement and the difficulty of instructional materials. Unpublished doctoral dissertation, University of Illinois, Champaign.
- Dunlap, C.C. (1954). Readability measurement: A review and comparison. Unpublished doctoral dissertation, University of Maryland, College Park.
- Durrell, D.D. (1937, revised 1955). Durrell analysis of reading difficulty. New York: Harcourt, Brace & World.



- Eberwein, L.D. (1979). The variability of readability of basal reader textbooks and how much teachers know about it. Reading World, 18, 259-272.
- Ekwall, E.E. (1974). Should repetitions be counted as errors?. The Reading Teacher, 27, 365-367.
- Ekwall, E.E., & English, J. (1971). Use of the polygraph to determine elementary school students' frustration level. (Final Report, Project No. OG078). Washington, DC: United States Department of Health, Education & Welfare.
- Ekwall, E.E. (1976). Diagnosis and remediation of the disabled reader. Boston, MA: Allyn & Bacon.
- Ekwall, E.E. (1979). Ekwall reading inventory. Boston: Allyn & Bacon.
- Ekwall, E.E., Solis, J., & Solis, E. (1973). Investigating informal reading inventory scoring criteria. Elementary English, 50, 271-274.
- Entin, E.B., & Klare, G.R. (1978). Factor analyses of three correlation matrices of readability variables. Journal of Reading Behavior, 10, 279-290.
- Fairbanks, G. (1937). The relation between eye-movements and voice in the oral reading of good and poor silent readers. Psychological Monographs, 48, 78-107.
- Farr, R. (1969). Reading: What can be measured? Newark, DE: International Reading Association.
- Farr, R. (Ed.). (1970). Measurement and evaluation of reading. New York: Harcourt, Brace & World.
- Farr, R., & Carey, R.F. (1986). Reading: What can be measured? (2nd Edition). Newark, DE: International Reading Association.
- Flesch, R.F. (1948). A new readability yardstick. Journal of Applied Psychology, 32, 221-233.
- Flesch, R.F. (1958). A new way to better English. New York: Harper & Brothers.
- Flesch, R.F. (1954). How to make sense. New York: Harper & Brothers.
- Flesch, R.F. (1949). The art of readable writing. New York: Harper & Row.



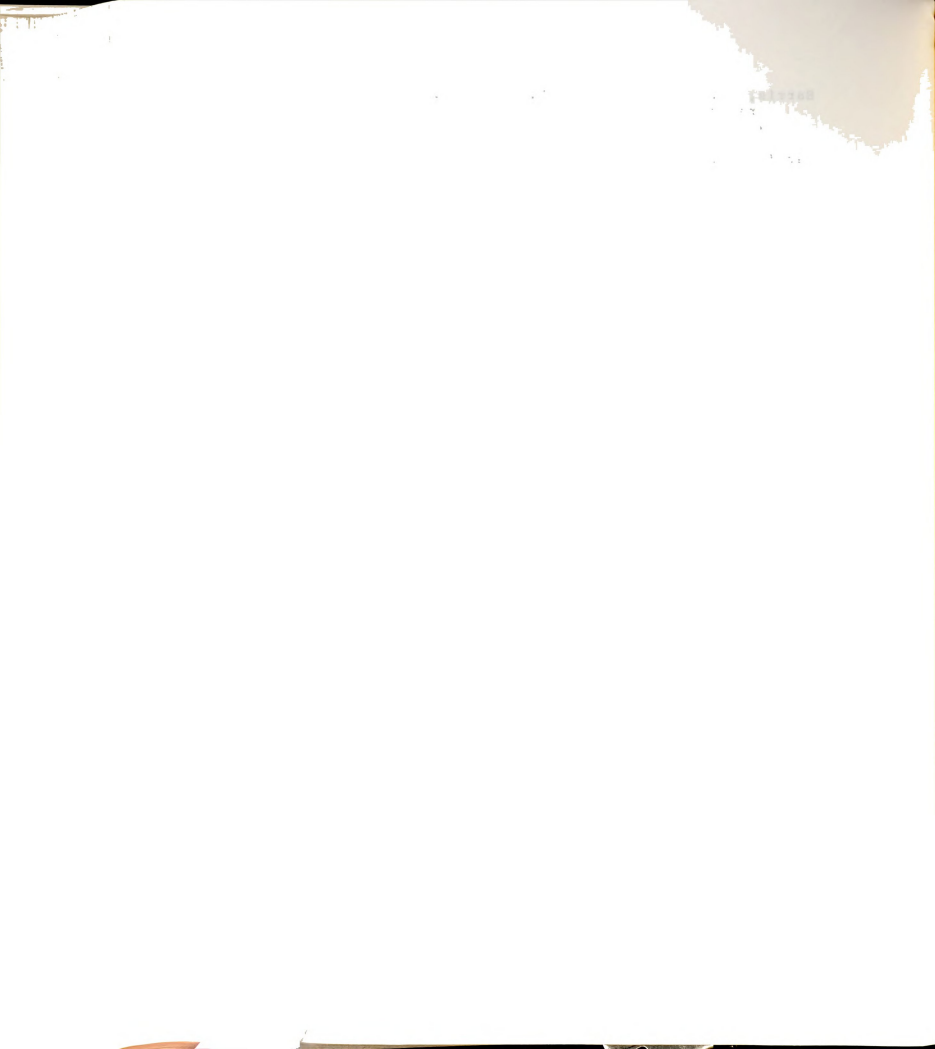
- Forbes, T.W., & Cottle, W.C. (1953). A new method for determining readability of standardized tests. Journal of Applied Psychology, 37, 185-190.
- Ford, P.L. (Ed.). (1962). The New England primer: A history of its origin and development. Teachers College, Columbia University.
- Fox, A.C. (1979). Foxies comparative chart (A study of readability results on stories in 32 basal series). Coeur d'Alene, ID: Fox Reading Research Company.
- Fry, E.B. (1980). Comments on the preceding Harris and Jacobson comparison of the Fry, Spache, and Harris-Jacobson readability formulas. The Reading Teacher, 33, 924-926.
- Fry, E.B. (1969). The readability graph validated at primary levels. The Reading Teacher, 22, 534-538.
- Fry, E. (1968). A readability formula that saves time. Journal of Reading, 11, 513-516; 575-578.
- Gage, N.L., & Berliner, D.C. (1984). Educational psychology (3rd Edition). Boston, MA: Houghton Mifflin.
- Gagne, R.M. (1968). Learning hierarchies. Educational Psychologist, 6, 1-9.
- Gagne, R.M. (1969). The acquisition of knowledge. Psychological Review, 69, 355-365.
- Gagne, R.M. (1965). The analysis of instructional objectives for the design of instruction. In R. Glaser (Ed.), Teaching machines and programmed learning II: Data and directions. Washington, DC: Department of Audio-Visual Instruction, National Education Association.
- Gambrell, L.B., Wilson, R.M., & Gantt, W.N. (1981). Classroom observations of task attending behaviors of good and poor readers. Journal of Educational Research, 74, 400-404.
- Geoffrion, L.D., & Geoffrion, O.P. (1978). Computers and reading instruction. Reading, MA: Addison Wesley.
- Gerbens, A. (1978). Read any good books lately? Kilobaud, 14, 104-106.
- Gill, D., Polin, R.M., Vinsonhaler, J.F., & VanRoekel, J. (1980). The impact of training on diagnostic consistency. E. Lansing, MI: The Institute for Research on Teaching, Michigan State University.



- Gillmore, J.V. (1974). The relation between certain oral reading habits and oral and silent reading comprehension. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.
- Glaser, N.A. (1964). A comparison of specific reading skills of advanced and retarded readers of fifth grade reading achievement. Unpublished doctoral dissertation, University of Oregon, Eugene.
- Glaser, R. (Ed.). (1965). Teaching machines and programmed learning II: Data and directions. Washington, DC: Department of Audiovisual Instruction, National Educational Association.
- Glasser, W. (1969). Schools without failure. New York: Harper & Row.
- Goodman, D., & Schwab, S. (1980, April). Computerized testing for readability. Creative Computing, 46-51.
- Goodman, K.S., & Fleming, J.T. (Eds.). (1969). Psycholinguistic and the teaching of reading. Newark DE: International Reading Association.
- Goodman, K.S. (1967). Reading: A psycholinguistic guessing game. Journal of the Reading Specialist, 4, 126-135.
- Goodman, Y.M., & Burke, C.L. (1972). Reading miscue inventory manual: Procedure for diagnosis and evaluation. New York: Macmillan.
- Gray, L., & Reese, D. (1957). Teaching children to read. New York: Ronald Press.
- Gray, W.S., & Leary, B.E. (1935). What makes a book readable? Chicago: University of Chicago Press.
- Gray, W.S. (1915). Standardized oral reading paragraphs. Bloomington, IL: Public School Publishing Co.
- Guthrie, J. (1974). The maze technique to assess, monitor reading comprehension. The Reading Teacher, 28, 161-168.
- Guzak, F.J. (1970). Dilemmas in informal reading assessments. Elementary English, 47, 666-670.
- Harris, A.J., & Jacobson, M.D. (1980). Comparison of the Fry, Spache, and Harris-Jacobson readability formulas for primary grades. Reading Teacher, 33, 920-924.

REVISED
JAN 1964
1000

- Harris, A.J., & Jacobson, M.D. (1976). Predicting twelfth graders' comprehension scores. Journal of Reading, 19, 43-47.
- Harris, A.J., & Jacobson, M.D. (1974, October). Revised Harris-Jacobson readability formula. Paper presented at the annual meeting of the College Reading Association, Bethesda, MD.
- Harris, A.J., & Sipay, E.R. (1980). How to increase reading ability (7th Ed.). New York: Longman.
- Hinkle, D.E., Wiersma, W., & Jurs, S.G. (1979). Applied statistics for the behavioral sciences. Boston, Houghton Mifflin.
- Huey, E.B. (1908, reprinted 1968). The psychology and pedagogy of reading. New York: Macmillan.
- Irving, S.L., & Arnold, W.B. (1979 September). Measuring readability of text. Personal Computing, 34-36.
- Irwin, J.W., & Davis, C.A. (1980). Assessing readability: The checklist approach. Journal of Reading, 23, 129-130.
- Jacobson, M.D., Kirkland, C.E., & Selden, R.W. (1978). An examination of the McCall-Grabbs standard test lessons in reading. Journal of Reading, 22, 224-230.
- Johnson, M.S., & Kress, R. (1965). Informal reading inventories. Newark, DE: International Reading Association.
- Johnston, P.H. (1984). Assessment in reading. In D.P. Pearson (Ed.), Handbook of reading research. New York: Longman.
- Jorgenson, G.W. (1977). Relationship of classroom behavior to the accuracy of the match between material difficulty and student ability. Journal of Educational Psychology, 69, 24-32.
- Judd, C.H., & Buswell, G.T. (1922). Silent reading: A study of the various types (Supplementary Educational Monographs, No. 23). Chicago: University of Chicago Press.
- Keller, P.T.G. (1982). Maryland micro: A prototype readability formula for small computers. Reading Teacher, 35, 778-782.



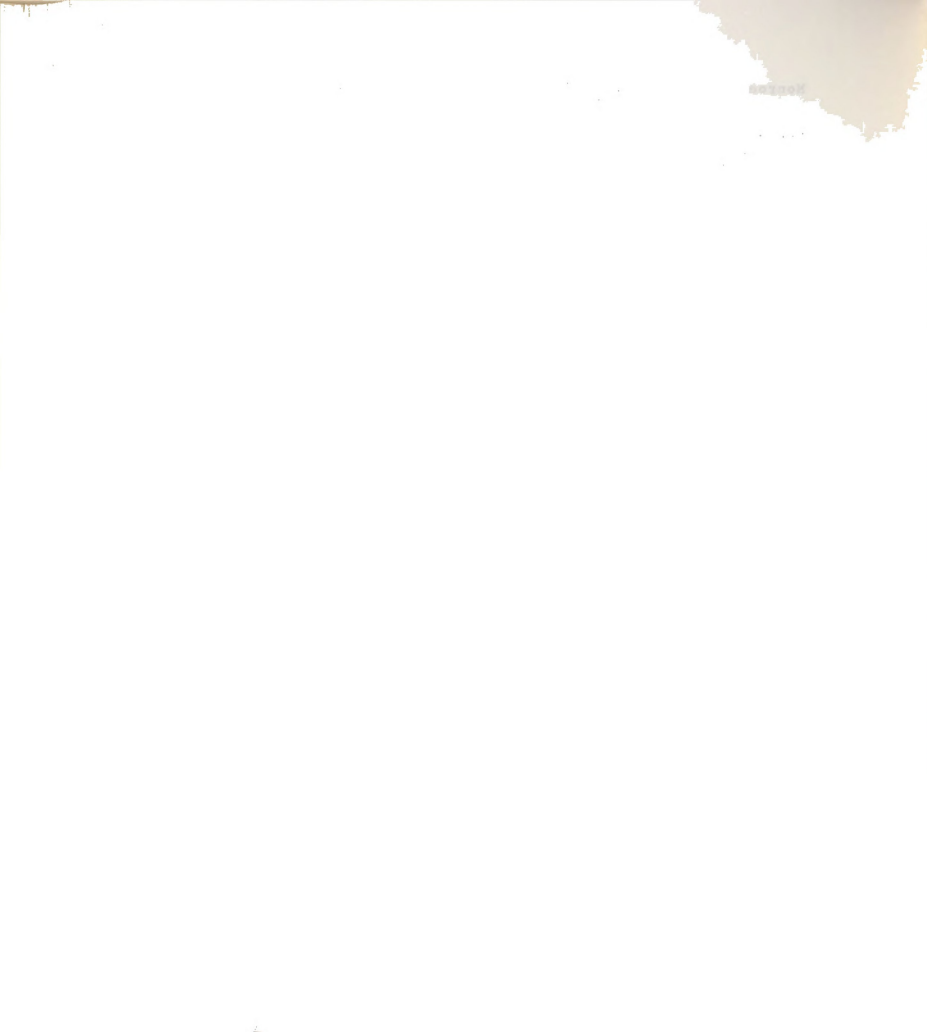
- Kibby, M.W. (1979). Passage readability affects the oral reading strategies of disabled readers. The Reading Teacher, 32, 390-396.
- Killgallon, P.A. (1942). A study of relationships among certain pupil adjustments in language situations. Unpublished doctoral dissertation, Pennsylvania State University, State College.
- Kincaid, J.P., Fishburne, R., Rogers, R.L., & Chissom, B.S. (1975). Derivation of new readability formulas (Automated readability index, fog count, and Flesch reading ease formula) for Navy enlisted personnel (Branch Report 8-75). Millington, TN: Chief of Naval Training.
- Klare, G.R. (1984). Readability. In P.D. Pearson (Ed.), Handbook of reading research. New York: Longman.
- Klare, G.R., & Buck, B. (1954). Know your reader: The scientific approach to readability. New York: Hermitage House.
- Klare, G.R. (1963). The measurement of readability. Ames: Iowa State University Press.
- Lantz, B. (1945). Some dynamic aspects of success and failure. Psychological Monographs, No. 271.
- Latimer, E.H. (1948, April). A comparative study of recent techniques for judging readability (Abstracts of Doctoral Dissertations). University of Pittsburgh Bulletin 44, 246-256.
- Leibert, R.E. (1965). An investigation of the differences in reading performance on two tests of reading. Unpublished doctoral dissertation, Syracuse University, Syracuse, NY.
- Lennon, R.T. (1951). The stability of achievement test results from grade to grade. Educational and Psychological Measurement, 11, 121-127.
- Leslie, L., & Osol, P. (1978). Changes in oral reading strategies as a function of quantities of miscues. Journal of Reading Behavior, 10, 442-445.
- Lewerenz, A.S. (1930). Vocabulary grade placement of typical newspaper content. Educational Research Bulletin, Los Angeles City Schools, 10, 4-6.
- Lively, B.A., & Pressey, S.L. (1923). A method for measuring the vocabulary burden of textbooks. Educational Administration and Supervision, 9, 389-398.



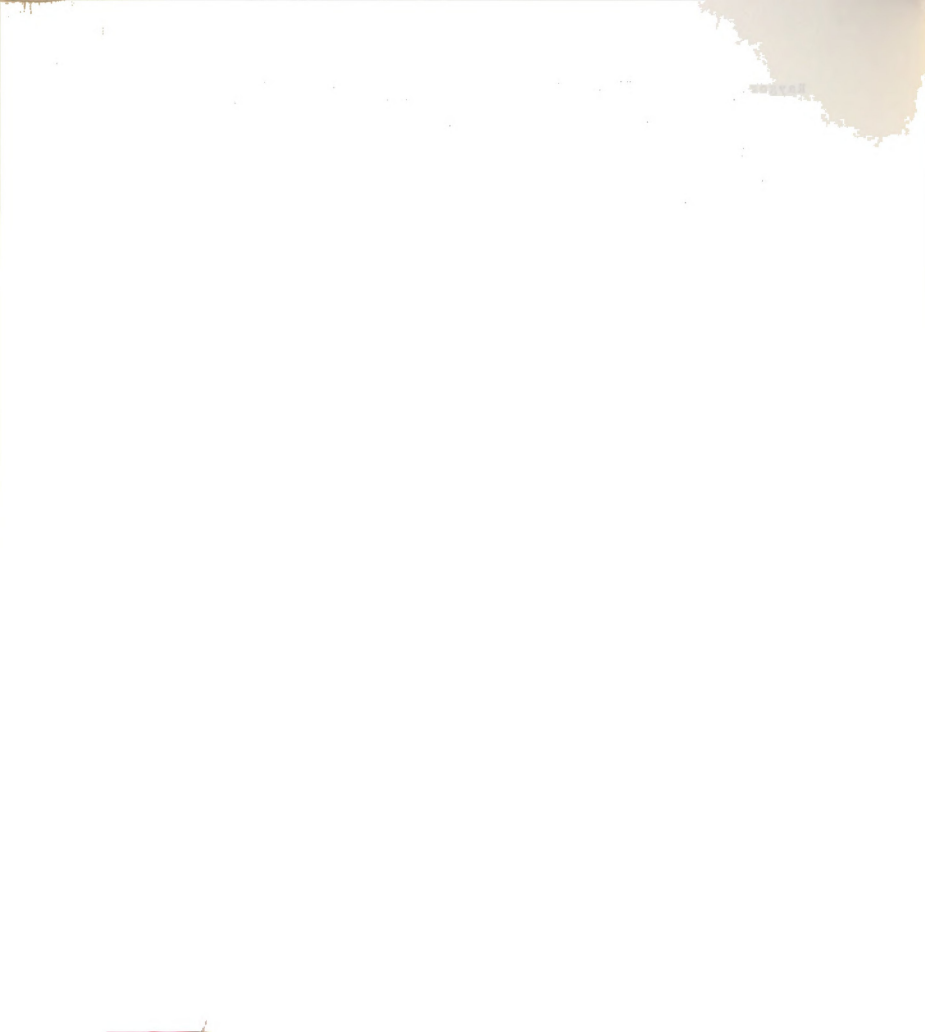
- Lorge, I. (1939). Predicting reading difficulty of selections for children. Elementary English Review, 16, 229-233.
- Lorge, I. (1949). Readability formulae-An evaluation. Elementary English, 26, 86-95.
- Lumsdaine, A.A. (1964). Educational technology, programmed learning, and instructional science. In E.R. Hilgard (Ed.), Theories of learning and instruction, Part I of the 63rd yrbk. of the National Society for the Study of Education. Chicago: University of Chicago Press.
- Lumsdaine, A.A., & Glaser, R. (Eds.). (1950). Teaching machines and programmed learning: A source book. Washington, DC: Department of Audiovisual Instruction, National Education Association.
- McCall, W.A., & Crabbs, L.M. (1925). Standard test lessons in reading: Teacher's manual for all books. New York: Bureau of Publications, Teachers College, Columbia University.
- McCracken, P. (1962, February). Standardized tests and informal reading inventories. Education, 366-369.
- McElroy, J. (1953, June). Fog count readability formula. In Guide for Air Force writing, Air Force manual 11-3. Maxwell, AL: Department of the Air Force, Maxwell Air Force Base, Air University.
- McLaughlin, G.H. (1969). SMOG grading: A new readability formula. Journal of Reading, 12, 639-646.
- MECC #749, School Utilities - Vol. 2. (1980). Minnesota Educational Computer Consortium.
- Miller, G.R., & Coleman, E.B. (1967). A set of thirty-six passages calibrated for complexity. Journal of Verbal Learning and Verbal Behavior, 6, 851-854.
- Miller, G.R., & Coleman, E.B. (1972). The measurement of reading speed and the obligation to generalize to a population of reading materials. Journal of Reading Behavior, 4, 48-56.
- Miller, R.B. (1962). Task description and analysis. In R.M. Gagne (Ed.), Psychological principles in system development. New York: Holt, Rinehart & Winston.
- Mills, R.E., & Richardson, J.R. (1963). What do publishers mean by grade level? The Reading Teacher, 16, 359-362.



- Monroe, M. (1932). Children who cannot read. Chicago: University of Chicago Press.
- Mosenthal, P. (1976-1977). Psycholinguistics properties of aural and visual comprehension as determined by children's abilities to comprehend syllogisms. Reading Research Quarterly, 12, 55-92.
- Mosenthal, P. (1978). The new and given in children's comprehension of presuppositive negatives in two modes of processing. Journal of Reading Behavior, 10, 267-278.
- Ojemann, R.H. (1934). The reading ability of parents and factors associated with reading difficulty of parent education materials. University of Iowa Studies in Child Welfare, 8, 11-32.
- Paolo, M.F. (1977). A comparison of readability graph scores and oral reading errors on trade books for beginning readers. Unpublished master's thesis, Rutgers: The State University of New Jersey, New Brunswick.
- Payne, C. (1930). The classification of errors in oral reading. Elementary School Journal, 31, 142-146.
- Peterson, M.J. (1956). Comparison of Flesch readability scores with a test of reading comprehension. Journal of Applied Psychology, 40, 35-36.
- Pikulski, J.J., & Shanahan, T. (1982). Informal reading inventories: A critical analysis. In J.J. Pikulski & T. Shanahan (Eds.), Approaches to the informal evaluation of reading. Newark, DE: International Reading Association.
- Pitner, R. (1913). Oral and silent reading of fourth grade pupils. Journal of Educational Psychology, 4, 330-337.
- Polin, R.M. (1981). A study of preceptor training of classroom teachers in reading diagnosis. (Reading Series No. 110). E. Lansing, MI: The Institute for Research on Teaching, Michigan State University.
- Powell, W.R. (1969). Reappraising the criteria for interpreting informal inventories. In D. DeBoer (Ed.), Reading diagnosis and evaluation: Proceedings of the thirteenth annual convention. Newark, DE: International Reading Association.
- Powell, W.R., & Dunkeld, C. (1971). Validity of the IRI reading levels. Elementary English, 48, 637-642.



- Raygor, A.L. (1977). The Raygor readability estimate: A quick and easy way to determine difficulty. In P.D. Pearson (Ed.), Reading: Theory, research and practice (26th Yearbook of the National Reading Conference). Clemson, SC: National Reading Conference.
- Rumelhart, D. (1977). Toward an interactive model of reading. In S. Dornic (Ed), Attention and performance VI. Hillsdale, NJ: Erlbaum.
- Samuels, J.S. (1979). The method of repeated readings. The Reading Teacher, 33, 403-408.
- Samuels, S.J., & Kamil, M.L. (1984). Models of the reading process. In P.D. Pearson (Ed.), Handbook of Reading Research. New York: Longman.
- Schlieper, A. (1977). Oral reading errors in relation to grade and level of skill. The Reading Teacher, 31, 283-287.
- Schuyler, M.R. (1982). A readability program for use on microcomputers. Journal of Reading, 25, 560-591.
- Sears, P.S. (1940). Level of aspiration in academically successful and unsuccessful children. Journal of Abnormal and Social Psychology, 35, 498-536.
- Sherman, G.B., Weinshank, A., & Brown, S. (1979). Training reading specialists in diagnosis (Research Series No. 31). E. Lansing, MI: The Institute for Research on Teaching, Michigan State University.
- Silvaroli, N.J. (1965). Classroom reading inventory. Dubuque, IA: Wm. C. Brown Co.
- Singer, H., & Dolan, D. (1980). Reading and learning from text. Boston: Little, Brown & Co.
- Singer, H. (1975). The SEER technique: A non-computational procedure for estimating readability level. Journal of Reading Behavior, 7, 255-267.
- Sipay, E.R. (1964). Comparison of standardized reading scores and functional reading levels. Reading Teacher, 17, 265-268.
- Smith, J.K. (1977). Perspectives on mastery learning and mastery testing. Princeton, NJ: Eric Clearinghouse on Tests, Measurement and Evaluation.



- Smith, L. (1976). Miscue research and readability. In P.D. Allen & D.J. Watson (Eds.), Findings of research in miscue analysis: Classroom implications. Urbana IL: National Council of Teachers of English, ERIC Clearinghouse on Reading and Communication Skills.
- Spache, G.D. (1953). A new readability formula for primary grade reading materials. Elementary School Journal, 53, 410-413.
- Spache, G.D., & Spache, E.B. (1977). Reading in the elementary school (4th ed.). Boston, MA: Allyn & Bacon.
- Spache, G.D. (1972). Diagnostic reading scales. Monterey, CA: California Test Bureau.
- Spache, G.D. (1974). Good reading for poor readers (Rev. 9th ed.). Champaign, IL: Garrard Publishing Co.
- Spiro, R.J., & Myers, A. (1984). Individual differences and underlying cognitive processes in reading. In P.D. Pearson (Ed.), Handbook of Reading Research. New York: Longman.
- Stadlander, E.L. (1936). A scale for evaluating the difficulty of reading materials for the intermediate grades (Abstract of Doctoral Dissertation). University of Pittsburgh Bulletin, 35, 347-352.
- Stanovich, K.E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. Reading Research Quarterly, 16, 32-71.
- Stolurow, L.M., & Newman, J.R. (1959). A factional analysis of objective features of printed language presumably related to reading difficulty. Journal of Educational Research, 52, 243-251.
- Stone, C. (1957). Measuring difficulty of primary reading material: A constructive criticism of Spache's measure. Elementary School Journal, 57, 36-41.
- Taylor, W.L. (1953). Cloze procedure: A new tool for measuring readability. Journalism Quarterly, 30, 415-433.
- Thorndike, E.L. (1921). Educational psychology, 1: The original nature of man, 2: The psychology of learning, 3: Work and fatigue, individual differences. New York: Bureau of Publications, Teachers College, Columbia University Press.



- Thorndike, E.L. (1917). Reading reasoning: A study of mistakes in paragraph reading. Journal of Educational Psychology, 6, 323-332.
- Thorndike, E.L. (1921). The teacher's word book. New York: Teacher's College, Columbia University.
- Traxler, A.E. (1950). Reading growth of secondary school pupils during a five year period. Educational Records Bulletin, 54, 98-107.
- Vacca, R.T. (1981). Content area reading. Boston: Little, Brown & Co.
- Veatch, J. (1978). Reading in the elementary school (2nd ed.). New York: John Wiley & Sons.
- Washburne, C., & Vogel, M. (1926). What books fit what children? School and Society, 23, 22-24.
- Washburne, C., & Vogel, M. (1928). An objective method of determining grade placement of children's reading materials. Elementary School Journal, 28, 373-381.
- Weiner, B. (1972). Theories of motivation: From mechanism to cognition. Chicago: Markham.
- Wienshank, A.B. (1980). Investigation of the diagnostic reliability of reading specialists, learning disabilities specialists, and classroom teachers: Results and implications. E. Lansing, MI: The Institute for Research on Teaching, Michigan State University.
- Wells, C.A. (1950). The value of an oral reading test for diagnosis of the reading difficulties of college freshmen of low academic performance. Psychological Monographs, 64, 1-35.
- Winer, B.J. (1971). Statistical principles in experimental design. New York: McGraw Hill.



MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03071 4806