

1991

A Study of Qualitative Miscue Analysis Scoring Systems for Identification of Instructional Reading Levels

Sylvia Estelle Dean
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds



Part of the [Educational Assessment, Evaluation, and Research Commons](#), and the [Educational Methods Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Dean, Sylvia Estelle, "A Study of Qualitative Miscue Analysis Scoring Systems for Identification of Instructional Reading Levels" (1991). *Dissertations and Theses*. Paper 1388.
<https://doi.org/10.15760/etd.1387>

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

A STUDY OF QUALITATIVE MISCUE ANALYSIS
SCORING SYSTEMS FOR IDENTIFICATION OF
INSTRUCTIONAL READING LEVELS

by

SYLVIA ESTELLE DEAN

A dissertation submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF EDUCATION
in
EDUCATIONAL LEADERSHIP

Portland State University
1991

TO THE OFFICE OF GRADUATE STUDIES:

The members of the Committee approve the dissertation
of Sylvia Estelle Dean presented July 8, 1991.

[REDACTED]

Amy Driscoll, Chair

[REDACTED]

Brad Eliot aka Frank B. May

[REDACTED]

Eric Kimmel

[REDACTED]

Mary K. Kinnick

[REDACTED]

Cathleen Smith

APPROVED:

[REDACTED]

Robert B. Everhart, Dean, School of Education


[REDACTED]


C. William Savery, Vice Provost for Graduate Studies and
Research

AN ABSTRACT OF THE DISSERTATION OF Sylvia Estelle Dean for the Doctor of Education in Educational Leadership presented July 8, 1991.

Title: A Study of Qualitative Miscue Analysis Scoring Systems for Identification of Instructional Reading Levels.

APPROVED BY THE MEMBERS OF THE DISSERTATION COMMITTEE:


Amy Driscoll, Chairperson


Brad Eliot aka Frank B. May


Eric Kimmel


Mary K. Kinnick


Cathleen Smith

The purpose of this descriptive study was to determine the accuracy and practicality of May's Poor Reader (PR) scoring system for the informal reading inventory (IRI), an individual assessment device designed to determine a

student's instructional reading level. The PR is a qualitative scoring system developed by Frank May that examines only two miscues (defaults and meaning-denigrating substitutions) in arriving at an estimate of instructional reading level. The predictor variable, PR, was compared for accuracy and practicality with five other predictor variables consisting of four traditional quantitative scoring systems and an additional qualitative system of May's; PR was also compared with four criterion variables: (a) a scoring system created by Frank May on the basis of research concerning miscues and informal reading inventories, a system that requires the use of a context scale and a graphophonic scale, (b) the judgments of tape recordings made by an experienced and knowledgeable reading coordinator, (c) the judgments of ten reading teachers of the students under their tutelage, (d) and a silent reading score on Form B of the same IRI. The comparisons were made through the use of chi-square tests of significance in which each of the six predictor variables was compared with each of the four criterion variables as to accuracy of agreement with the criterion variables.

The study answers the following research questions:

- 1) Is the qualitative PR (poor reader) scoring system more practical and accurate than four traditional scoring systems in placing students in instructional materials?

- 2) Is the PR scoring system as practical and accurate as May's CG (context-graphophonic) scoring system based on miscue and IRI research findings?
- 3) Is the PR scoring system as accurate as an experienced and knowledgeable reading coordinator's estimates of the student's level of instruction?
- 4) How does the PR scoring system compare with two other variables, teacher estimates and silent reading scores, in placing students in instructional materials?

Examination of the results showed that there were no significant differences between the instructional estimates made by the PR scoring system and two of the four criterion variables, the research based scoring system and the experienced reading coordinator. This was also true for May's third qualitative scoring system called the CGQ. All other differences in the estimates of instructional level were highly significant--with the four traditional predictor variables and with two of the four criterion variables ($p < .01$).

The conclusions drawn from these results serve as answers to the four research questions and are limited to the population of the present study:

- (a) The PR (poor reader) scoring system was more practical and accurate than the four traditional

quantitative scoring systems in placing students in instructional materials ($p < .01$).

- (b) May's PR scoring system was about as accurate as, and also more practical than, May's two-scaled procedure called the CG (context-graphophonic) system in placing students in instructional materials.
- (c) The PR scoring system was about as accurate and practical as the estimates made by an experienced and knowledgeable reading coordinator in placing students in instructional materials.
- (d) The PR scoring system did as well as any of the other predictor variables in making estimates that agree either with the reading teachers or the silent reading scores in placing students in instructional materials.

The main implications drawn from this study were as follows:

1. It may be advisable for classroom teachers and reading teachers to make use of May's PR scoring system for the IRI as a quick and qualitative way of estimating students' instructional reading level.
2. Since the PR scoring system met the eight stringent criteria established prior to the study for a "qualitative IRI scoring system," researchers may

wish to employ this quick system as one of the independent variables used in studies related to the improvement of informal reading inventories.

ACKNOWLEDGMENTS

The completion of this dissertation conjures up within me feelings of exhilaration, and a sense of deep satisfaction. Yet I am well aware that I did not accomplish this goal alone. I would like to extend my deepest gratitude to those persons who have been supportive, helpful and concerned about me while I pursued this goal.

I wish first of all to acknowledge Amy Driscoll, my advisor, committee chairperson and friend, whose rigorous academic standards and unmitigated caring and belief in my ability enabled me to complete this study.

Additionally, my deepest gratitude is extended to Brad Eliot (aka Frank May), a member of my committee, a researcher par excellence, and professor of reading whose years of dedication to scholarly research, theorizing, and teaching provided the rationale and impetus for this study. His unwavering dedication to a qualitative approach to reading assessment is an important aspect of his contribution to the field of reading.

Both Dr. Driscoll and Dr. Eliot gave relentlessly of their time and talent to the critiquing of this document.

Gratitude is also extended to Drs. Cathleen Smith, Mary Kinnick, and Eric Kimmel, the other members of my committee,

and to Dr. Joel Arick, who taught the Doctoral Research sequence, for their commitment to academic excellence and for the significant insights which they contributed to this study.

Further support was lent by Jim Leffler and Virginia Anderson, doctoral students, and Mike Trevasain, a professional colleague. Their constant encouragement and words of wisdom inspired me when the "going got tough."

I extend my deepest gratitude to my children, Michael, Mary, Anna, Linda, Donna, Lisa, Paul, Paul and Terry. Their pride in what I was doing spurred me on and they continue to bring me an immeasurable amount of joy as they pursue their individual goals.

Most of all I wish to thank Paul Arnold Dean, my husband and best friend, for his unwavering faith and constant love. His persistent support and technical assistance have enabled me to complete this degree.

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	xi
 CHAPTER	
I INTRODUCTION	1
Problem Statement	4
Rationale	5
Current Practice	6
Potential for Change	8
Purpose of this Study	9
Scoring Systems	9
Traditional (T)	9
Traditional Minus Self-Corrections (TS)	10
Traditional Minus Repeats (TR)	11
Traditional Minus Self-Corrections and Repeats (TRS)	11
Context-Graphophonic (CG)	12
Context-Graphophonic Question (CGQ)	12
Poor Reader (PR)	12
Definition of Terms	13
Significance of the Study	16
Research Questions	17
Research Hypothesis	19
Hypothesis One	19
Hypothesis Two	19
Hypothesis Three	20
Hypothesis Four	20

CHAPTER	PAGE
Description of the Methodology	20
Assumptions	21
Limitations of the Study	22
Organization of the Document	24
II REVIEW OF THE LITERATURE	25
Introduction	25
Contents	26
Purpose	26
Definitions of Reading	27
Early Definitions of Reading	28
Current Definitions	28
Formal Reading Assessment	31
Background	31
Norm Referenced Tests	32
Rationale for Use	32
Current Use and Limitations	32
Criterion Referenced Tests	35
Current Use and Limitations	36
Summary of Discussion of Formal Reading Assessment	37
Informal Reading Assessment	38
General Purposes of Informal Reading Inventories	39
Examples of Informal Reading Assessment	39
The Informal Reading Inventory	40
Early Developments	41
Need to Improve Informal Assessment	41
Beginning of Research Support	42
Later Developments	43
The Classroom Reading Inventory	45

CHAPTER	PAGE
	Later Developments Related to IRIs . 46
	Current Development Related to IRIs . 46
	Miscue Analysis 48
	Early Developments 48
	Current Use 48
	Current Effect on Practice 49
	May's Qualitative Scoring Systems . . 51
	May's CG, CGQ, PR Systems 52
	Research Rationale 53
	The Future 54
	Summary 55
III	RESEARCH PROCEDURES 57
	Purpose 57
	Design 58
	Criterion and Predictor Variables 58
	Criterion Variables 58
	Predictor Variables 59
	Research Hypothesis 60
	Hypothesis One 60
	Hypothesis Two 61
	Hypothesis Three 61
	Hypothesis Four 61
	Setting 62
	Subjects 63
	Teacher Participants 63
	Subject Selection 64
	Participating Teacher Survey 64
	Survey 64
	Survey Validity 64
	Instruments and Scoring Systems 71
	The Classroom Reading Inventory . . . 71
	The Validity of the Classroom Reading Inventory 73
	Content Validity 75

CHAPTER

PAGE

	Comparison With Standardized Tests	77
	The Reliability of the Classroom Reading Inventory	79
	Alternate Forms Reliability of the Silvaroli	80
	Interrater Reliability	82
	Bett's Traditional Quantitative Scoring System	82
	Bett's Scoring Criteria	83
	Variations of Bett's Traditional Scale	84
	May's Qualitative Scoring Systems for the IRI	84
	The CG System	84
	The Context Scale for the CG System	85
	The Graphophonic Scale for the CG System	86
	Demonstration of the CG Scoring Procedure	87
	May's Criteria Based on CG Scores	87
	The CGQ System	88
	May's Criteria Based on CGQ Scores	89
	The PR System	89
	Data Collecting Procedures	90
	Student Selection	90
	Privacy and Informed Consent	91
	Oral Reading Sample Collecting	92
	Taping Procedures	92
	Coding Procedures	93
	Scoring Procedures	93
	Data Analysis Procedures	94
	Summary	95
IV	RESULTS	97
	Introduction	97
	Statistical Method	98
	Variables	99

CHAPTER	PAGE
Population Groupings	100
Research Hypotheses	100
Hypothesis One: CG as Expected	
Criterion	100
Hypothesis Two: AE as Expected	
Criterion	101
Hypothesis Three: TE as Expected	
Criterion	101
Hypothesis Four: SR as Expected	
Criterion	101
Data Presentation	102
Hypothesis 1A	102
Hypothesis 1B	105
Hypothesis 1C	107
Hypothesis 1D	110
Hypothesis 1E	112
Hypothesis 1F	115
Hypothesis 1G	117
Hypothesis 2A	120
Hypothesis 2B	123
Hypothesis 2C	125
Hypothesis 2D	127
Hypothesis 2E	130
Hypothesis 2F	133
Hypothesis 2G	135
Hypothesis 3A	138
Hypothesis 3B	140
Hypothesis 3C	142
Hypothesis 3D	145
Hypothesis 3E	147
Hypothesis 3F	149
Hypothesis 3G	151
Hypothesis 4A	153
Hypothesis 4B	155
Hypothesis 4C	158
Hypothesis 4D	160
Hypothesis 4E	162
Hypothesis 4F	164
Hypothesis 4G	166
V SUMMARY AND CONCLUSIONS	169
Implications and Recommendations	169

CHAPTER	PAGE
Introduction	169
Summary	169
Research Questions	170
Procedures	171
Statistical Hypothesis and Results	171
Comparison for All Seven Population	
Groups	179
Additional Results	180
Conclusions and Discussion	180
Conclusion One	181
Conclusion Two	181
Conclusion Three	182
Conclusion Four	182
Implications and Recommendations	186
Implication One	186
Implication Two	187
Closing Statement	192
REFERENCES	195
APPENDIX	205

LIST OF TABLES

TABLE		PAGE
I	Traditional Quantitative Scoring Systems Display and Discussion, 1988	10
II	Qualitative Scoring Systems Display and Discussion, 1988	11
III	Participating Teachers Training and Experience, 1989	66
IV	Current Methods Used by Participating Teachers to Select Chapter I Reading Students, 1989 . .	67
V	The Standardized Group Test is a Reasonably Accurate Method of Identifying a Student's Instructional Reading Level, 1989	67
VI	An Informal Reading Inventory is a Reasonably Accurate Method of Identifying a Student's Instructional Reading Level, 1989	68
VII	The Current Method Used to Select Students for Program is Reasonably Accurate, 1989	68
VIII	Composite Ranking of Teachers' Preferred Methods of Scoring IRIs, 1989	69
IX	Teachers' Perceptions of the Chapter I Students' Socio-Economic Status, 1989	69
X	Teachers' Perceptions of the Chapter I Students' Basic Family Structure, 1989	70
XI	Teachers' Perceptions of the Chapter I Students' Mobility, 1989	70
XII	Description of Chapter I Student Population, 1989	71
XIII	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable CG for the Total Population (N = 200), 1990	103

TABLE	PAGE
XIV	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable CG for the Population Girls (N = 100), 1990 106
XV	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable CG for the Population Boys (N = 100), 1990 108
XVI	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable CG for the Population 2nd Grade (N = 69), 1990 111
XVII	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable CG for the Population 5th Grade (N = 34), 1990 113
XVIII	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable CG for the Population 3rd Grade (N = 53), 1990 116
XIX	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable CG for the Population 4th Grade (N = 44), 1990 119
XX	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable AE for the Total Population (N = 200), 1990 121
XXI	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable AE for the Population Girls (N = 100), 1990 124
XXII	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable AE for the Population Boys (N = 100), 1990 126
XXIII	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable AE for the Population 2nd Grade (N = 69), 1990 129
XXIV	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable AE for the Population 5th Grade (N = 34), 1990 131
XXV	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable AE for the Population 3rd Grade (N = 53), 1990 134

TABLE	PAGE
XXVI	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable AE for the Population 4th Grade (N = 44), 1990 . . . 136
XXVII	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable TE for the Total Population (N = 200), 1990 139
XXVIII	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable TE for the Population Girls (N = 100), 1990 141
XXIX	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable TE for the Population Boys (N = 100), 1990 143
XXX	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable TE for the Population 2nd Grade (N = 69), 1990 . . . 146
XXXI	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable TE for the Population 5th Grade (N = 34), 1990 . . . 148
XXXII	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable TE for the Population 3rd Grade (N = 53), 1990 . . . 150
XXXIII	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable TE for the Population 4th Grade (N = 44), 1990 . . . 152
XXXIV	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable SR for the Total Population (N = 200), 1990 154
XXXV	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable SR for the Population Girls (N = 100), 1990 157
XXXVI	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable SR for the Population Boys (N = 100), 1990 159
XXXVII	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable SR for the Population 2nd Grade (N = 69), 1990 . . . 161

TABLE		PAGE
XXXVIII	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable SR for the Population 5th Grade (N = 34), 1990 . .	163
XXXIX	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable SR for the Population 3rd Grade (N = 53), 1990 . .	165
XXXX	Chi Square Analysis of Expected and Observed Frequencies for the Criterion Variable SR for the Population 4th Grade (N = 44), 1990 . .	167
XXXXI	Percentage of Agreement on Instructional Level Estimates Between Criterion Variables and the PR, CGQ, and the Most Qualitative Traditional Scoring System, 1990	178

CHAPTER I

INTRODUCTION

The use of formal assessment methods such as standardized, norm referenced tests to measure reading performance has been increasing steadily since the early 1950's (Resnick, 1981). The strength of such formal tests is in their ability to provide a quantitative score based on a fairly reliable, partial assessment in an inexpensive and efficient manner (Anderson, Hiebert, Scott & Wilkinson, 1985). The limit of these tests is their provision of little useful information to teachers to refine specific instructional strategies (Readence, 1987). Furthermore, the scores are often subject to misinterpretation when used by practitioners to place students in appropriate instructional materials (May, 1990).

The criterion-referenced test is another formal reading assessment which is being increasingly used due to the continued focus on mastery learning, outcome based education and minimum competencies (Robinson & Craver, 1989). Criterion-referenced tests should allow reading assessment to be individualized or group oriented as demanded by circumstance or need. Yet, the reality is that when school

districts try to develop their own tests, they are difficult, time consuming and expensive to construct. Consequently, districts often use the preconstructed commercial criterion-referenced tests, even though these do not necessarily measure what is being taught.

Furthermore, criterion-referenced tests fragment the reading process into discrete skills as if each was important in its own right. These tests do not allow teachers to assess how, when, or why, students alter their approaches to reading. Therefore, this formal approach to assessment provides little information congruent with the current view of reading, which defines it as a holistic, constructive process rather than as a set of specific discrete skills.

Informal measures are also used in addition to formal measures, to assess reading performance. The informal reading inventory (IRI) has been used for over forty years as the preferred informal method of placing students in instructional materials (Bader & Wiesendanger, 1989; Johns & Lund, 1982; Pikulski & Shanahan, 1982). Most versions of the informal reading inventory (IRI) encourage teachers to look primarily at the quantity of a reader's errors rather than the quality of errors; but again, because this approach is efficient and practical, it is commonly used. Such a traditional approach often leads teachers to inaccurately

estimate a student's reading level (Farr & Carey, 1986; Goodman, Watson & Burke, 1987; May, 1990; Weaver, 1988).

What is needed is an approach to reading assessment that is accurate and practical. The approach should be authentic, that is, close to the nature of the reading act, and specifically reflective of a research-based definition of reading. It should also provide the practitioner with information necessary to appropriately place students in curriculum or instructional materials and to make sound instructional decisions.

It is believed (May, 1990) that a qualitative approach to the scoring of informal reading inventories would provide such information. Such a qualitative approach should adhere to the following criteria:

- * would reflect research findings on the definition of oral reading errors (miscues),
- * would be based on oral reading samples that provide miscue analysis evidence of a reader's concepts and strategies with easy, average and difficult reading,
- * would not penalize good readers in order to get estimates of instructional levels low enough for poor readers,
- * would be accurate and easily used by classroom teachers,

- * would be usable for a wide range of grade levels, at least from grades 2 through 5 (Aaronson & Farr, 1988; Burke, 1976; Dunkeld, 1970; Goodman, 1976; Goodman, Watson, & Burke, 1987; May, 1990; Rumelhart, 1985; Smith, 1988),
- * would place students in instructional materials as accurately as May's time-consuming but research based CG scoring system,
- * would place students as accurately as a knowledgeable and experienced reading coordinator,
- * would produce highly similar estimates in an inter-rater reliability study.

There have been some attempts to develop such qualitative approaches to reading assessment, (Allen, 1976; Cunningham, 1984; Sadowski & Lee, 1986) yet the results have not been sufficiently accurate nor practical (Christie, 1979; Goodman et al., 1987; Groff, 1980; Hoffman, 1980; Hood, 1976; Potter, 1980). What is needed is a qualitative approach to the scoring of informal reading inventories that is easy to use and which produces accurate results.

PROBLEM STATEMENT

Although norm referenced and criterion referenced tests continue to be used as formal assessments of reading performance, they provide the teacher little useful information for placement and instructional decision making.

The informal reading inventory (IRI) has been the preferred informal method of placing students in instructional materials because of its practicality and validity, yet such use often leads teachers to inaccurately estimate a student's reading level because of the currently used traditional method of scoring. Recently developed qualitative approaches to the scoring of informal reading inventories (IRIs) have been time consuming and tedious for practitioners to use. What is needed is a qualitative approach to the scoring of IRIs which is practical and accurate and which adheres to current conceptions of the reading process. The rationale for such an approach is worthy of consideration.

RATIONALE

Research of the past thirty years has fundamentally reshaped conceptions of the reading process, yet these changes have not been effectively applied to reading assessment methods (Rumelhart, 1985). There is little correspondence between contemporary theories of the reading process and assumptions implicit in these assessments (Bussis & Chittendon, 1987). This observation is supported by the continued use by classroom teachers and reading specialists of quantitative assessment techniques such as "standardized norm referenced, and criterion referenced tests," (Anderson et al, 1985, p. 98) and traditionally

administered, quantitatively scored informal reading inventories (Baumann, 1988; Betts, 1946; Harker, 1990; Johnston, 1984; Powell, 1978).

Current Practice

Three significant factors may account for the continuation of these ineffective practices. The first factor may be the strong historical preference for quantitative assessment instruments. Standardized norm referenced tests are examples of quantitative assessment instruments consistently used by professionals, even though these tests are now sometimes viewed as social artifacts, that is, something the profession uses as a customary rather than valid practice.

Researchers and theorists have continued such usage, while studies demonstrate the need to examine this practice (Aaronson & Farr, 1988; Coleman, 1982; Guzak, 1970; Sipay, 1964; Shanahan, 1983; Worthen & Spandel, 1991). Practitioners, in turn, rely on these traditional methods to collect data for instructional decision-making, even though these assessment approaches may be poor predictors of student performance (Johnston, 1984; May, 1990; Pikulski & Shanahan, 1982).

The use of such tests is discussed by Farr & Carey (1986):

There is no question that the administration and use or misuse of reading tests has increased significantly over the past fifteen years.

Another important change has been the direction of reading research which focused on reading as a cognitive process...this process revealing work has much to recommend to those who develop and use tests...

Little attention has been paid to the development of informal techniques which could be used to supplement or supplant the use of standardized tests. In fact, the search for alternatives has not produced much that is new (pp. 6, 14).

A second factor which probably fosters the continued use of traditional assessment methods is availability.

Basal reader programs have made "informal reading inventories" even more accessible to the teacher. Yet the methodology of penalizing a student for all reading errors and determining comprehension on the basis of question responses alone, leads to serious concerns regarding the accuracy of their results as well (D'Angelo & Mahalios, 1983; Gillis & Olson, 1987; Goodman & Burke, 1968; May, 1990).

Although qualitative scoring systems have been developed as a method of placing students in instructional materials, their ease of use has not been established (Goodman & Burke, 1987; May, 1986, 1990; Weaver, 1988). Goodman and Burke's (1987) Qualitative Approach To Miscue Analysis attempts to meet the need for a qualitative approach to the scoring of IRIs, but their methodology is impractical, time consuming, and difficult to consistently apply (Allen, 1976; Argyle, 1989; Goodman & Burke, 1987; Johns, 1982). Weaver (1988) also supports the qualitative

approach to miscue analysis, yet falls short of developing a methodology which is useful for the practitioner.

Potential For Change

In contrast to "current practices", May (1986, 1990) has heuristically developed qualitative scoring systems that show potential for practicality and accuracy. This qualitative approach may reduce the preference for formal and traditional reading assessment methods. The advantages of May's systems appear to be: 1) a reflection of the research on the meaning of miscues; 2) use of oral reading samples in order to provide miscue evidence of a reader's concept and strategies with easy, average and difficult reading; 3) lack of penalties for good readers in order to get estimates of instructional levels for poor readers; 4) promise of efficient, accurate and easy use by classroom teachers; 5) application to a wide range of grade levels.

In summary, the use of May's qualitative scoring systems may result in more practical and accurate placement of students in curriculum and instructional materials by classroom teachers and reading specialists (Beebe, 1980; Long, 1987; Marek, Goodman & Babcock, 1985; May, 1990). Their use may provide more information for instructional decision-making.

PURPOSE OF THIS STUDY

The purpose of this descriptive quantitative study was to determine the accuracy of May's Poor Reader (PR) scoring system for the informal reading inventory (IRI), an individual assessment device designed to determine a student's instructional reading level. The PR was compared for practicality and accuracy with four traditional and largely quantitative scoring systems (T, TS, TR, TSR) and a qualitative system, the CGQ. These results were further compared to four criterion scores (CG, AE, TE, SR). For this study the T, TS, TR, TSR, CGQ and PR are the predictor variables, and CG, AE, TE and SR are the criterion variables. The following definitions and descriptions should increase the reader's understanding of the nature of each approach to scoring.

SCORING SYSTEMS

The four traditional "quantitative" scoring systems are summarized by the following display and discussion.

Traditional (T)

The traditional scoring system (T) counts all errors made by the student during oral reading including insertions (word added to original text by student during oral reading), omissions (word omitted during a student's oral reading), repeats (the resaying of a word from a text during

oral reading), defaults (teacher pronounces an unknown word for a student during oral reading after waiting at least five seconds), substitutions (words, non-words, or mispronunciations used by the student during oral reading to replace the author's intended text), and any type of self-correction (where the student goes back and rereads a word correctly).

TABLE I
TRADITIONAL QUANTITATIVE SCORING SYSTEMS
DISPLAY AND DISCUSSION
1988

T	Traditional	Counts all errors made by the student during oral reading including insertions, omissions, repeats, defaults, substitutions and self corrections.
TS	Traditional Minus Self Corrections	Same as traditional with the exception that self corrections are not counted.
TR	Traditional Minus Repeats	Same as traditional with the exception that repeats are not counted.
TSR	Traditional Minus Repeats and Self Corrections	Same as traditional with the exception that repeats as well as self corrections are not counted.

Traditional Minus Self-Corrections (TS)

The (TS) scoring system is the same as the traditional, but self-corrections are not counted.

Traditional Minus Repeats (TR)

The (TR) scoring system is the same as the traditional but repeats are not counted.

Traditional Minus Self-Corrections and Repeats (TRS)

With the (TRS) system, neither repeats nor self-corrections are counted; all other errors are the same as the traditional scoring system.

The three qualitative scoring systems developed by May (1986, 1990) are summarized by the following display and discussion.

TABLE II

QUALITATIVE SCORING SYSTEMS DISPLAY AND DISCUSSION
1988

CG	Context	A research based qualitative scoring system which weighs self corrected substitutions as positive miscues, meaning-denigrating substitutions and defaults as poor reader errors.
CGQ	Context Graphophonic Question	A scoring system which is the same as the CG system except it also has a questioning and answering component.
PR	Poor Reader	A system which counts only defaults and "non-contextual" substitutions, non-word substitutions, and mispronunciation substitutions.

Context-Graphophonic (CG)

The (CG) provides, through two qualitative scales, a measure of a student's context sensitivity (C), and graphophonic sensitivity (G). The Context scale measures the students' combined use of three cues: semantic (meaning), syntactic (order and type of words expected), and schematic (mini-theories about things, people, language, places and other phenomena in our background of experiences). The Graphophonic scale measures the students' ability to notice letters. The CG scoring system weighs synonymous substitutions and self-corrected substitutions as "positive" miscues; it weighs meaning denigrating substitutions (non-contextual substitutions, nonword substitutions, and mispronunciation substitutions) and defaults as poor reader errors.

Context Graphophonic Question (CGQ)

The (CGQ) is the second of May's qualitative scoring systems and has the same elements as the (CG) scoring system with the addition of a question-answering component. This question scale measures the student's ability to respond to passage-related inferential and literal questions.

Poor Reader (PR)

This is the third of May's qualitative scoring systems and counts only "poor reader" errors as a portion of the

total words. A "poor reader" error consists of defaults and meaning-denigrating substitutions (May, 1986, 1990).

DEFINITION OF TERMS

This section provides the remaining definitions which relate directly to the study.

Accuracy. The quality attributed to an IRI scoring system that allows the teacher to place students in instructional materials with no more than 10% error. In a classroom setting, this would mean no more than three students in a class of 30.

Administrator's Estimate (AE). The IRI administrator's decision to place a student in a specific level of reading instruction materials based on her knowledge of the research on miscues and IRIs, on observation of the student's behavior during the administration of the IRI, and on 30 years of experience and professional development in the teaching of reading. In this study, the AE refers to the person who obtained and scored the 200 reading samples.

Informal Reading Inventory (IRI). An informal reading inventory is an individually administered reading test composed of a series of graded passages which the student reads in front of the teacher, who notes oral reading miscues such as mispronunciations, omissions, repetitions, and substitutions. After the oral reading, the student may be asked to respond to questions related to the passage.

Levels of Reading Competence. As defined by Betts (1946), the levels of reading competence refer to the three levels which describe a student's reading competence: independent, instructional, and frustrational.

Independent. Level at which a student can read material easily and independently, without assistance from others.

Instructional. Level at which the reader makes some errors and requires instructional aid to benefit from the reading.

Frustrational. Level at which the student understands little of what is read and makes many miscues.

Miscue. This term, as defined by Goodman (1968), refers to an actual observed response in oral reading which does not match the expected response.

Miscue Analysis. As defined by Goodman, this term refers to a system for comparing expected oral reading responses on an informal reading inventory with observed oral reading responses.

Practicality. The quality attributed to an IRI scoring system that (1) permits the teacher to compute a score as a simple percentage and (2) reduces the usual time needed for administering and scoring the IRI (due to the fact that not all errors are counted).

Reading. Reading is the process of constructing meaning through the dynamic interaction (transaction) among

the reader's existing knowledge, the information suggested by the written language, and the context of the reading situation (Weaver, 1988).

Reading Assessment. This term refers to the formal and informal acquisition of information about the reading ability of an individual or group and the evaluation of that information in order to make educational decisions.

Silent Reading Score (SR). A question answering scale which measures the students ability to silently read specific passages of an IRI, and to respond to passages related inferential and literal questions.

Silvaroli Informal Reading Inventory. This IRI, also known as the Classroom Reading Inventory, was developed by Silvaroli (cited in Johnson, 1978). This individually administered IRI consists of a series of graded passages, pre-primer through level eight, each of which is followed by passage related inferential and factual questions. The student completes the oral reading of specific passages and then the teacher estimates the student's comprehension by asking for responses to the passage related questions. The results of this inventory help the teacher determine the student's instructional reading level. Alternate parallel forms are available to assess silent reading.

Teacher Estimate (TE). A reading teacher's decision to place a student in instructional material based on criteria each teacher considers important, such as assessment scores,

prior reading history, or even body language observed while the student reads. In this study, TE refers to the teacher estimates made by 9 of the 10 Chapter I reading teachers who were voluntary participants in the study.

SIGNIFICANCE OF THE STUDY

One measure of the significance of educational research findings is their ability to make a contribution to the improvement of current practice. This study should improve practices for assessing reading currently used by practitioners in the field of reading by providing an accurate and practical research based scoring system for use with IRIs.

Many practitioners currently question the purposes and uses of traditional quantitative assessment methods whose results often penalize students and place them in programs inappropriately. These practitioners voice the need for a process rather than product approach to reading assessment. Such an approach would support the "holistic" view of the reading act and examine "miscues" as data sources for making instructional decisions rather than solely as error counts for scoring or grading purposes. Such a process approach would empower the teacher to become a facilitator of the student's learning rather than a "controller" of their destiny.

The data from this study should support such an approach. Additionally, data from this study should prove useful to future studies which focus on qualitative approaches to assessing instructional reading levels with informal reading inventories (IRIs). Research on reading assessment in general, and informal reading inventories (IRIs) in particular, has supported the use of a qualitative approach (Baumann, 1988; Beebe, 1980; Clay, 1979; D'Angelo & Mahlios, 1983; Woodley, 1988). What is needed is the use of a practical and accurate method of qualitatively scoring informal reading inventories (IRIs). Data from this study could support the use of May's (PR) poor reader scoring method by classroom teachers. These data may also support the use of May's more detailed scoring method, the CG method, for school district reading specialists. By providing such data on two qualitative scoring systems, appropriate for use by practitioners and reading specialists, this study could make a significant contribution to the field of reading.

RESEARCH QUESTIONS

Four research questions give direction to this study. The questions focus on the ability of various IRI (informal reading inventory) scoring systems to predict instructional reading levels. Specifically, answers were sought to the following questions:

1. Is the qualitative PR (poor reader) scoring system more practical and accurate than T, TS, TR, TSR (the four traditional scoring systems based on Betts, 1946) in placing students in instructional materials?
2. Is the PR scoring system as practical and accurate as May's (1988) CG (context, graphophonic) scoring system based on miscue analysis and IRI research findings?
3. Is the PR scoring system as accurate as the IRI administrator's estimates of the students' level of instruction (AE)?
4. How does the PR scoring system compare with two other criterion variables, TE (teacher estimate) and SR (silent reading) in placing students in instructional materials?

The variables of AE (administrator's estimate), TE (teacher's estimate), and SR (silent reading) were considered in the comparisons because they represent the reality of current practice. The AE represents thirty years of experiences in reading assessment, training and expertise in reading assessment, and training as a reading specialist. The SR score is the most commonly accepted criteria, but it primarily reflects student recall and neglects reading ability. Teacher ratings are also used with frequency, but they are often subjective and limited by teacher ability.

In responding to the research questions, this study attempts to identify an accurate and practical method of placing students in instructional reading materials.

RESEARCH HYPOTHESES

From the research questions, the following research hypotheses were developed for this study:

Hypothesis One

CG as Criterion: There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of the six predictor variables (T, TS, TR, TSR, CGQ and PR) and each of the seven population groupings (total, boys, girls, 2nd grade, 5th grade, 3rd grade and 4th grade), when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the CG estimates.

Hypothesis Two

AE as Criterion: There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables and each of the seven population groupings, when the expected frequencies are 5%, 90%, and 5% respectively and the

observed frequencies are the number of agreements between the predictor estimates and the AE estimates.

Hypothesis Three

TE as Criterion: There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables and each of the seven population groupings, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the TE estimates.

Hypothesis Four

SR as Criterion: There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables and each of the seven population groupings, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the SR estimates.

DESCRIPTION OF THE METHODOLOGY

The purpose of this study was to determine the accuracy and "practicality" of May's (1988) PR (poor reader) scoring system for the informal reading inventory (IRI), an

individual assessment device designed to identify a student's instructional reading level. This purpose was realized through a descriptive quantitative study.

Ten Chapter I remedial reading teachers from the same number of Northeast Portland elementary schools who volunteered to participate in the study each selected twenty Chapter I reading students per school (n=200) from grades two through five. These students were tape recorded by an experienced reading coordinator (the researcher in this study) reading aloud from an informal reading inventory (IRI) and their miscues scored, using four traditional scoring systems (T, TS, TR, TRS), and May's three qualitative systems (PR, CGQ, CG). Each system's accuracy in placing each child at the appropriate instructional level was compared, through Chi square statistics, with the placement recommended by the four previously mentioned criterion variables (CG, AE, TE, and SR).

ASSUMPTIONS

The following assumptions were made in this study:

1. Students would give valid reading performances when asked to participate in the study.
2. The Silvaroli Informal Reading Inventory, also called the Classroom Reading Inventory, represents currently used informal reading inventories.

3. The sampling procedures used to identify participants were appropriate to this study.

LIMITATIONS OF THE STUDY

The following limitations were considered:

1. Threats to Internal Validity. Threats to internal validity such as maturation of the students were not evident in this study. The experienced and knowledgeable reading coordinator, also known in this study as AE, collected oral reading samples over a period of only four months, therefore there was no chance for maturation to effect the results. In addition, there was no evidence of prior exposure by students to the selected measuring instrument, the Silvaroli Informal Reading Inventory.
2. Testing Effects. There was no pretest given. There was an alternate form of the Silvaroli IRI used to collect the silent reading scores. It is believed that these procedures limited the possibility of testing effects.
3. Potential Sample Bias. A volunteer population of ten Chapter I Reading Specialists from ten Portland public elementary schools randomly selected 200 of their Chapter I reading students to participate in the study. The fact that the teachers represent a volunteer rather than a random population, may present a sample bias.

The AE (administrator's estimates) of students' reading levels, one of the four criteria against which the PR was compared was obtained by the researcher in this study. The researcher collected the 200 oral reading samples via audio tapes. During the taping sessions, the researcher estimated the students' instructional reading level based on research related to miscues, training and practice in the use of the PR scoring system, and 30 years of reading assessment and instruction experience. These estimates are the AE criterion for this study. After all of the cases were collected, the researcher scored the results using the seven scoring systems (T, TS, TR, TSR, CGQ, PR, CG). These results were compared and analyzed using Chi square. There is a possibility of bias on the part of the researcher, although the scoring took place after the AE was obtained. The structure of the scoring systems limits the variability of the results. Additionally, the inter-rater reliability results support the accuracy of the obtained results.

4. Generalization of Findings. A sample of 200 second through fifth grade students was selected from one geographic area of the city. This sampling may not allow the study to be replicated to rural or other urban settings, nor with students of other

socio-economic strata. Therefore, caution must be exercised regarding the generalization of the findings.

ORGANIZATION OF THE DOCUMENT

The organization of this dissertation is as follows:

Chapter I provided an introduction, problem statement, rationale, purpose, definition of terms, significance, research questions, assumptions, limitations, and an organizational plan for the document.

Chapter II includes selected definitions of reading; a discussion of formal reading assessment and its limitations; the development of informal assessment with a focus on weaknesses discovered in informal reading inventory assessment; and miscue analysis and scoring systems used with informal reading inventories.

Chapter III offers a description of the purpose, design, variables, research hypotheses, setting, subjects, data collecting and scoring instruments, procedures, and statistical methods for analyzing the data.

Chapter IV presents the tables and the findings related to the statistical analyses of the data collected.

Chapter V summarizes the study, presents the conclusions and implications, makes recommendations, and offers suggestions for further areas of study and research.

CHAPTER II

REVIEW OF THE LITERATURE

INTRODUCTION

After years of dormancy, the measurement of reading has once again become a central concern for researchers, theorists and practitioners. The reason for this occurrence is that the economic, societal, political and personal implications of reading performance have increased in importance as society has become more complex (Anderson et al., 1985; Harris & Sipay, 1985; Johnston, 1976; Rumelhart, 1985; Worthen & Spandel, 1991).

According to Glazer & Searfoss (1988), reading assessment warrants re-examination in view of these changes, "in definitions of the reading process and purposes for instruction, in classroom and clinical pedagogical procedures, in the population of students, and in expectations of student literacy" (p. 1). In order for assessment to be a relevant tool of the reading practitioner, these changes must be reflected in current assessment purposes and methodology (Farr & Carey, 1986; Pearson & Valencia, 1987; Pikulski, 1990; Smith, 1987; Weaver, 1988).

CONTENTS

The body of literature on reading assessment is vast and must be narrowed to the scope of this study. Therefore, this review includes: 1) selected definitions of reading related to this study; 2) the formal assessment of reading and its limitations; 3) the informal assessment of reading and its limitations; 4) informal reading inventories and, 5) miscue analysis and scoring systems used with informal reading inventories.

PURPOSE

This review will establish the need for a more appropriate methodology than the traditional quantitative approaches used for placing students in instructional materials. Specifically, this review will support the development of an alternative approach to the traditional quantitative scoring of the informal reading inventory, one that:

- * is qualitative rather than quantitative
- * reflects research findings on the nature of oral reading errors (miscues)
- * is based on oral reading samples that provide miscue analysis evidence of a reader's concepts and strategies with easy, average and difficult reading

- * does not penalize good readers in order to get estimates low enough for poor readers
- * is accurate and practical for classroom teachers to use to place students in instructional materials
- * is usable for a wide range of grade levels, at least 2 through 5.

This review of the literature will begin with definitions of reading reviewed from a historical perspective. The progression toward current views of reading will establish the need for matching progression in assessment procedures. Among the many definitions presented, one has been selected to best reflect current perceptions of the reading process. This definition provides the rationale for a qualitative approach to informal reading assessment.

DEFINITIONS OF READING

Approaches to reading assessment can be linked to the very definition of reading itself, and to the changes in definition. To understand the current definition of reading, it is appropriate to review, from a historical perspective, the development of several representative definitions.

Early Definitions of Reading

Early definitions of reading as presented in Harris and Hodges (1981) portray it as a scientific, product-oriented act, or in Plato's words "an exact science distinguishing the separate letters both by the eye and the ear" (p. 264).

By the 1900's definitions had broadened and reading was beginning to be appreciated for its complexity. Thorndike (1917) described reading as, "The understanding of the printed word...a very elaborate procedure" (p. 323).

Until recently narrow, mechanical descriptions have been consistently maintained in contrast to the emerging process description. In 1955, Flesch viewed reading as, "a precise process which involves the exact detailed sequential perception and identification of letters and words" (p. 206).

In 1968, however, an extreme opposite description was posed by Huey when he defined reading as, "a wonderful process, by which our thoughts and thought-wanderings to the finest shades of detail...are reflected from us to another soul who reads us through our words" (p. 6).

These early definitions are precursors of a more process oriented approach to the reading act. Such an approach is reflected in the more current definitions.

Current Definitions

More recent definitions of reading emphasize an interactive process between the reader and the printed word

(Collins, Brown & Larkin, 1980, cited in Glazer & Searfoss, 1988; Tierney & Pearson, 1983).

Goodman (1976), who has consistently influenced reading theory and practice, particularly as a researcher and supporter of the "Whole Language" movement, offers this definition, "an interaction between the reader and written language through which the reader attempts to reconstruct a message from the writer" (p. 130).

Smith (1979) whose writings significantly influence theory and practice in reading, furthers this interactive view, "Reading involves a selective process of asking questions of the printed text. Whereby one deliberately seeks just the information that we need" (p. 105).

The definition offered by Valencia and Pearson (1987) extends the definition of reading to one of a strategic process, within a specific situation. "Every act of reading requires the orchestration of many resources including the text, the reader's prior knowledge, other learners, and the constraints of the situation" (p. 728).

May (1990), who also specifies the reading strategy as a meaning-seeking process, states that it is also something more. "Reading is a social process. That is, it involves the willingness on the part of readers and authors to communicate with each other" (p. 19).

The complexity of reading is well-established when these definitions are considered. These definitions have

evolved from a focus on the application of specific precise skills to a complex process of interacting conceptually with the printed word to obtain the author's intended meaning.

The current definitions of reading are represented well in a position paper from the Michigan Department of Education (1988). This definition, as cited in Weaver (1988) states, "Reading is the process of constructing meaning through the dynamic interaction [transaction] among the reader's existing knowledge, the information suggested by the written language, and the context of the reading situation" (p. 5).

With this definition of reading considered, there is cause for discontent with current reading assessment practices and support for a new approach to assessment. The discontent prompts questions of whether currently used tests do, or even can, measure the reading process, especially when the current model describes reading as an interactive process and recommends that assessment practices be holistic and authentic, rather than skill-specific and linguistically artificial.

A discussion of current assessment practices is in order at this time, beginning with the topic of formal reading assessments, and proceeding to informal reading assessment. The discussion of these assessment practices will continue to provide a rationale for the present study.

FORMAL READING ASSESSMENT

Background

Contributions to the field of formal reading assessment were made by such early psychologists as Binet, Huey, and Thorndike. Binet (cited in Johnston, 1984), developed a precursor of a reading test in his search for an instrument which would measure intelligence. Huey (1908, 1968) discussed at length the very process of reading itself, and questioned what should be measured. But it is Thorndike's (1917) notable contributions of a silent reading test and the development of the first readability formula which still influence current reading assessment practices.

The development of test scoring machines and computer summaries speed the impact of the obtained results, yet the format and structure of current formal tests are similar to Thorndike's time. It is important to note that such reading tests were developed for the sake of convenience in construction, administration and scoring, and such tests are too often poor predictors of individual students' performance (Worthen & Spandel, 1991).

These tests were a forerunner of the standardized norm referenced tests, the Gates MacGinitie and Stanford, which are used today. It is appropriate to discuss norm referenced tests in the light of their current use and limitations.

NORM REFERENCED TESTS

Anastasi (cited in Robinson & Craves, 1989) states that norm referenced tests are "measures that compare an individual or group of students to a specific distribution of scores obtained from a similar group of individuals" (p. 65).

Norm referenced tests measure children's performance in relation to other children in the same grade. Test publishers establish norms based on the performance of representative, nationwide samples of children in each grade.

Rationale for Use

The influence of the scientific movement, which valued the search for quantifiable results, encouraged the use of norm referenced tests. Such tests were used as a selection method during the Industrial Revolution to counteract the practice of using favoritism and power to place people in positions of influence. Ironically, these norm referenced tests are presently viewed as posing barriers to the same opportunities which they historically sought to provide (Farr & Carey, 1987).

Current Use and Limitations

Ironically, norm referenced tests are currently the most common formal assessment measures used to determine an individual student's reading instructional level. This is

primarily because of their convenience, and yet their results are often mis-interpreted by teachers. May (1986) explains why this occurs:

If the norm group is representative of the total population, a teacher should expect half of the class to score at or below "grade level" and half to score at or above grade level. It is expecting too much to have every student at or above grade level. In addition, the test publishers are translating raw scores into standard scores based on a great deal of mathematical predicting, thus saving the time and money it would take to test a norm group at each grade level and for each month of the year. (p. 337)

In addition to the misinterpretation of norm referenced tests, May (1986) discusses other limitations of such standardized reading tests:

- * because of their length, they are administered to groups, depriving the teacher of observing each child in action;
- * because they are timed, they penalize those who read well but slowly;
- * they require recognition of correct answers to multiple choice questions, while reading is a far more complex act;
- * they frequently over-rate a poor reader's ability, causing the teacher to inappropriately place a child in instructional materials;
- * some students are good readers but poor test takers, therefore the test is an inaccurate view of their true ability;
- * the length of passages in the test is not representative of the length of passages the child will be required to read normally, therefore the test results are not a true picture of the child's ability to perform in class;
- * the norm group is not necessarily representative of the group being tested;

- * the text of standardized tests includes twice as much informational content as fictional content, therefore it is not representative of widely used instructional materials. (p. 341)

Furthermore, Becoming a Nation of Readers: The Report of the Commission on Reading, Anderson et al. (1985) cites this criticism of norm referenced tests: "Performance on such tests depends not only on a student's ability, but also on their prior knowledge of the passage content of the test items" (p. 98). This report continues with the following observation:

Most tests do not permit skilled readers to use strategies that are important to normal reading. For example, good readers use the structure an author has provided to organize, learn, and ultimately remember information. A norm referenced test, as presently constituted, does not allow such strategies to come into play and therefore gives an impoverished picture of reading competence. (p. 98)

Carroll (1987) discusses the continued use of formal reading tests by administrators, school boards, parents and teachers to describe the performance of groups such as classes, schools and districts, without such information being used to improve reading ability. Carroll concurs with Anderson (1985) that the findings from such tests are often misread and inaccurately reported.

Some standardized norm referenced tests are marketed as providing diagnostic information, yet a survey indicates that teachers do not use the results for such purposes (Salmon & Cox, 1981). This use of tests merely to describe

performance does nothing to improve the student's learning opportunity.

The Report of the Commission on Reading in Becoming a Nation of Readers (1985) summarizes what is needed to more accurately assess reading performance:

A more valid assessment of basic reading proficiency than that provided by standardized tests could be obtained by ascertaining whether:

- * students read aloud familiar but grade-appropriate material with acceptable fluency
- * To support this assessment, teachers should tape record the oral reading of each child three times a year and keep the tapes on file for diagnosing and reporting. (p. 99)

This recommended approach is part of the methodology used in May's (1986, 1990) informal qualitative scoring systems. These systems are suggested as appropriate alternatives to current formal assessment practices such as the norm referenced tests previously discussed. This review now turns to a discussion of the criterion referenced test, another formal approach to reading assessment currently used by practitioners.

CRITERION REFERENCED TESTS

Baumann (1988) states that criterion referenced tests are, "Measures which focus on the mastery of defined skills or content as determined by comparing a student's performance (a test score) to some criterion or standard" (p. 6).

Criterion reference tests are generally regarded as assessments related to specific instructional decision making. The developer of the criterion referenced test, according to Farr & Carey (1986), is not interested in how much of a specific behavior an examinee possesses, but rather, if he can demonstrate proficiency on the specific objectives included in the test. As such, this approach is the one most often used by those who define reading as a conglomeration of subskills rather than a process of obtaining meaning from the author's text. An absolute level of performances is set as the criterion against which an individual is evaluated (Anderson et al., 1985).

Current Use and Limitations

Criterion referenced tests are used heavily in schools that employ "skills management systems" for reading instruction because they are skill specific tools. The Commission on Reading (1985) cites several weaknesses in the use of this assessment approach:

- * insufficient attention is given to helping children integrate all of the small subskills into the overall skill of reading. Many children pass the tests, yet do not read well.
- * neither research nor conventional wisdom furnishes an agreed upon division of reading into a sequence of subskills.
- * learning to read appears to involve close knitting of reading skills that complement and support one another, rather than learning one skill, adding a second. (p. 97)

In summary, the use of criterion referenced tests must be viewed with caution, for they can inhibit rather than enhance sound instruction and appropriate placement decisions by classroom teachers. For instance, students receive instruction on discrete skills based on criterion test results. These specific lessons are often focused on skills which do not strengthen true reading ability and in fact they often adversely effect students' attitudes since they take time away from readings for pleasure opportunities in the classroom.

Often the discrete skills dictate grouping practices which result in an artificial lack-step criterion as opposed to the child's own natural development, a more appropriate criterion.

Summary of Discussion of Formal Reading Assessment

A review of the literature has revealed that formal reading assessment methods, rooted in the mental measurement movement and influenced by early psychologists such as Binet, Huey, and Thorndike, persist because of their convenience and accessibility. Thorndike's silent reading test, developed for convenience in construction, administration and scoring, was the forerunner of such currently used standardized norm referenced tests as the Gates MacGinitie and the Stanford-Binet (Farr & Carey, 1986; Johnston, 1984; Worthen & Spandel, 1991).

Despite their many limitations, such as misinterpretation of results by teachers, group administration practices, time limitations, multiple choice questions, poor length of passages, and non-representative norming groups, these tests are still used because of their convenience.

In addition, the criterion referenced test is being increasingly used because of its perceived ability to measure specific skills -- data needed to support the current outcome based education movement. Yet commercially published criterion referenced tests do not necessarily measure what is being taught in specific schools.

The use of formal reading assessment is viewed as artificial and inappropriate to the current view of how to measure reading. What is needed is an approach which supports the current view of reading while providing information to the teacher for instruction and placement decisions. Informal approaches to reading assessment are more authentic and dynamic, and therefore more supportive of the current view of reading. Such approaches can provide more useful information to the practitioner for placement and instruction decisions. It is to these informal assessment approaches that this review now turns.

INFORMAL READING ASSESSMENT

Pikulski and Shanahan (1982) define informal reading assessment as, "the process of using specific strategies to assess the student's reading performance, generally in an instructional setting" (p. 2).

Informal reading assessment is important because practitioners in the field of reading instruction often disagree with results obtained on an individual or group formal test. Yet, these same practitioners perceive themselves as powerless to challenge the results of such assessment practices because they lack viable alternatives. Current literature in the field of reading assessment is supportive of the use of informal qualitative approaches to assessment (Pearson & Valencia, 1987; Pikulski & Shanahan, 1982; Valencia, 1990).

General Purposes of Informal Reading Inventories

Jerry Johns (cited in Pikulski & Shanahan, 1982) notes that although informal tests and measures of reading performance vary in their scope and sophistication, there are generally four major purposes for their use:

1. Studying, evaluating, or diagnosing reading behavior.
2. Monitoring student progress.
3. Supplementing and confirming or negating information gained from standardized norm-referenced and criterion referenced tests.

4. Obtaining information not readily available from other sources. (p. 1)

Examples of Informal Reading Assessment

Examples of informal assessment methods are: conferences, teacher observation, attitude surveys, cloze procedure, interest inventories, workbooks and worksheets, informal reading inventories and miscue analysis (Pikulski & Shanahan, 1982). The informal reading inventory, miscue analysis, and various scoring systems used with informal reading inventories will be discussed in depth because they are related to the methodology of the present study.

THE INFORMAL READING INVENTORY

Pikulski and Shanahan (1982) describe the informal reading inventory as, "A sequential series of reading selections, graded in difficulty, which students read and answer questions about, and a set of procedures for analyzing the student's reading behavior in an instructional situation" (p. 94).

The informal reading inventory is the most widely used informal approach to assessment used by classroom teacher and reading specialist alike. At the beginning of the twentieth century, there was no informal reading inventory (Johns & Lunn, 1984).

Early Developments

Waldo (cited in Johns & Lunn, 1983) first promoted the informal assessment of oral and silent reading ability to help teachers and administrators ultimately improve instruction. Waldo's work is important because it is one of the earliest attempts to assess silent reading rate and comprehension, although his method was to assess the whole class at once. Waldo kept records of pupils' oral reading ability and compared teaching methods for their effectiveness. He emphasized the study of silent reading for "it is of much greater importance than oral reading not only in school, but in future life" (p. 9).

In 1917, Gray's tests (cited in Johns & Lunn, 1983), although standardized, also influenced the development of the informal reading inventory (IRI) by using a format similar to many of today's informal reading inventories. These similarities of format include: paragraphs increasing in difficulty; individual testing by a trained person; an error marking system; assessment of silent reading comprehension through questions; and a criterion for evaluation (p. 10).

Need to Improve Informal Assessment

Professional publications of the 1920's (cited in Johns & Lunn, 1983) suggested the need to improve the informal assessment of reading. Suggestions for such improvements included making provisions for individual differences in

reading ability, use of student texts for informal classroom testing, and the need to develop a uniform process and criteria for informal assessment.

Writers during this period discussed at length the issues of process and criteria related to informal assessment. "Criteria" were important for they established the number of errors used to indicate the student's instructional, frustration, and independent reading levels.

Thorndike's writings (1934) supported the need for researched criteria for informal assessment while Gates (1935) recommended a uniform process but did not address the issue of uniform criteria. Betts (1936) offered pupil behaviors which might be observed during test administration, but did not refer to criteria nor reading levels. Durrell (1937) suggested the need for a basic criterion for word recognition, but none of these writers offered empirical evidence to support these suggestions.

Beginning of Research Support

The 1940's can be viewed as the decade in which researchers' efforts culminated in the development of standard criteria, primarily due to the contributions of Betts and Killgallon (1942). Betts, using his "subjective reading inventory" consisting of three levels: independent, instruction, and frustration, assigned criteria for each of these levels. The empirical evidence for the establishment of such criteria, and the size of his sample, raised

questions about the effectiveness of this study. Betts himself stated, "While research has validated most of the items included in the criteria for appraising reading performance by means of an informal reading inventory, total criteria for this purpose are in need of further study" (p. 439).

It appears, when reviewing his discussion of formal and informal assessment, that Betts (1946) regarded informal assessment as a more accurate indicator of a pupil's reading ability, but recognized the limitations of his research support (pp. 269, 273).

The 1950's continued the examination of the criteria for the independent, instructional and frustration reading levels. Cooper (1952) appears to be the first to present other criteria after Betts' effort. His experimental study has been virtually ignored (Kender, 1970), possibly because of methodological problems. Even with questionable findings, though, it served as a stimulus for further research in subsequent years (Dunkeld, 1970; Johns & Magliari, 1989; Powell, 1960).

Later Developments

The 1960's produced a continuance of research on criteria for instructional level, and a number of studies comparing standardized tests to IRIs (Brown, 1964, Coleman & Harmer, 1982; Johnston, 1984; Powell & Dunkeld, 1972). With the commercial publication of informal reading inventories

(Botel, 1966; McCracken, 1964; Silvaroli, cited in Johns & Lunn, 1983), came efforts to establish the validity and reliability of the inventories.

In 1961, McCracken conducted a study using an IRI for the expressed purpose of recording the oral reading performances of second grade readers in materials below, at, and above grade level and comparing the performance of good, average, and poor readers. The analysis included looking at errors both quantitatively and qualitatively (pp. 113-116).

McCracken's (1964) later study remains one of the most thorough and ambitious efforts to establish the validity and reliability of an informal reading inventory. His study sought to determine if the data from two IRIs and one standardized achievement test would support the concept of three reading levels; independent, instruction and frustration. The data affirmed that something labeled "instructional level" could be measured validly (pp. 366-369).

In addition to the significant contributions of McCracken to the research of the informal reading inventory, the development of the Classroom Reading Inventory by Silvaroli (1969) during this period affects current practice because of its continued use. Since it is the instrument used in this study it is the topic to which this review turns.

The Classroom Reading Inventory

The Classroom Reading Inventory (1967), also known as the Silvaroli, is an informal reading inventory developed by Silvaroli, and is the IRI used for this study because of its history of use by researchers, classroom teachers and reading specialists. The Mental Measurement Yearbook (Buros, 1976) provides the following observations regarding its descriptive information and effectiveness:

- * provides a formula for determining independent, instruction and frustration levels.
- * passages appear to possess appropriate level to level difficulty and satisfactory interest.
- * the questions appear to be good, though some are not passage dependent, that is they could be answered independent of the selections.
- * compares favorably with other published inventories. The selections are deliberately brief to control testing time.
- * the motivational statement provides a set for the reading passage.
- * is a useful tool in the hands of a knowledgeable teacher.
- * the pictures provide a context for the content presented (pp. 748-750).

Although there are some problems in construction, it is the most consistently used IRI by classroom teachers. It is useful for its stated purpose, to support the identification of appropriate instructional placement level. The issue of its validity and reliability will be discussed in Chapter III.

Later Developments Related to IRIs

Commercially published IRIs continued to grow in number during the 1970's (Johns, 1978; Sucher & Allred, 1973). They appeared in various reviews and in an annotated bibliography (Johns, Garton, Schoenfielder, Skriba, 1977). A study conducted by Dunkeld (1970) sought to validate the use of the IRI as a technique for designating a student's instructional reading level. His study concluded that valid instructional levels can be usefully and precisely defined by a child's performance on an IRI (p. 6274-A). A concern for criteria related to errors made, was still in evidence (Dunkeld & Powell, 1972; Pikulski, 1974). Research was also conducted on a teacher's ability to use IRIs to determine a student's instructional level (Windell, 1975).

Current Development Related to IRIs

The search for an appropriate scoring criteria, begun with Betts (1946), whose guidelines are still widely used, continues today. Cramer (1980) and Johns and Lunn (1983) also contributed to the understanding of IRIs with descriptions of currently available commercially developed forms.

In a study conducted by Masztal and Smith (1984) results revealed that most elementary grade teachers who responded to a questionnaire are familiar with IRIs and know how to administer them. In addition, most of the respondents actually utilized the informal reading

inventory. An important additional finding places emphasis on the appropriate interpretation and use of the information obtained by teachers.

McKenna (1990) cites passage readability, passage content, choice of questions, passage dependency of questions, scoring criteria and allowable miscues as some potential problems in the use of IRIs. Yet McKenna further states that these concerns are not sufficient reason to abandon the use of IRIs but rather, he recommends such concerns be mitigated by further research (p. 676).

Areas of concern regarding IRIs have been focused primarily on common administration, scoring and interpretation procedures (Anderson & Joels, 1985). Bader and Wisendanger (1989) also indicated that IRIs afford the possibility of evaluating reading behavior in depth and have potential for training prospective teachers about reading behavior unequalled by other types of learning opportunities.

Perhaps the greatest influence on the thinking concerning informal reading inventories came about as a result of insights from psycholinguists. These insights, primarily influenced by Goodman (1967), led to the term "miscues" to replace the term "errors" to describe misreadings (Goodman, 1967; Goodman & Burke, 1976).

"Miscue analysis", as defined by Goodman (1976), is not a fixed or prescribed appraisal of an informal reading

inventory performance. Rather, it is intended to provide a "window on the reading process" (p. 5). The contributions by Kenneth and Yetta Goodman have led to a large body of research focused on miscue analysis. It is to this subject that this review now turns.

MISCUE ANALYSIS

Early Developments

Goodman (1968) further describes "miscue analysis" as a system for comparing expected oral reading responses on an informal reading inventory with observed oral reading responses (p. 5). The analysis of such reading miscues began in 1962 as a technique for studying closely what children do when they read. Goodman's purpose was to develop a way to compare expected oral reading responses with observed oral responses. The difference between the responses became the key to a new approach in teaching reading.

Using "miscue analysis", Goodman and Burke (1965-74) studied a group of six youngsters as they learned to read. The result was the development of the RMI, a device for applying the analysis used by teachers during observations of students' oral reading.

Current Use

Miscue analysis is now a diagnostic procedure rooted in the newer psycholinguistic view of reading which is that

through miscues, readers show their strengths, as well as their weaknesses. No longer are deviations from the author's text considered errors, but rather miscues are defined as "an actual observed response in oral reading which does not match the expected response" (Goodman, 1981, p. 5).

Current Effect on Practice

The user of miscue analysis tries to answer the question "why" for each miscue examined. He or she examines the observed behavior of oral readers as an interaction between language and thought, as a process of constructing meaning from a graphic display. And thus the reader's use of all four cueing systems -- graphophonic, syntactic, semantic and schematic -- is considered. Long range implications of miscue research are complex. The Goodmans (1981) admonish the practitioner to be ever aware that oral reading observations are but surface indications of the reading process; that reading must be treated as a search for information by the reader, rather than that of making sounds we expect; and that the practitioner's perception of and application of this principle is needed to appropriately analyze "miscues", no matter the level of experience (cited in Allen & Watson, 1976). The publication of the RMI Reading Miscue Inventory (Goodman & Burke, 1972) could have brought miscue analysis within the reach of the classroom teacher, but it was far too complex.

Goodman (1969) had developed miscue analysis as a research technique to get at readers underlying strategies by examining their oral reading performances (cited in Long, 1986). Yet, the RMI developed for classroom teachers' use has proven to be undeniably complex, and effective training in its use is both exacting and time consuming (p. 1).

Over fifteen published modifications of the RMI have been located and examined by Long (1986). The majority of them attempt to simplify the use of the RMI.

The significance of this output is that it demonstrates the need for a qualitative approach to scoring of IRIs using miscue analysis research. Long's study (1986) also indicates a need for evidence of validation through field testing and further research of any such approach (p. 35).

In addition to the contributions of the Goodmans and Burke, Weaver (1988) has written extensively in defense of an approach to observe oral reading, yet her work falls short of a workable system. Her analysis of current work in the field leaves the reader frustrated with the complexity of her approach. She does not see such a methodology as a practical placement tool for teachers, she admits, but rather as a tool "for helping them gain insight into the reading process" (p. 329). Her discussion leaves the reader overwhelmed with the process and in need of a great deal of background information to apply her particular methodology (Weaver, 1988).

The significant contribution of the theory of miscue analysis to the field of reading continues. In an annotated bibliography (Marek, Goodman, & Babcock, 1985) over four-hundred references to the use of miscue analysis with informal reading inventories are cited.

Another significant contribution to the refining of this approach is the work of May (1986, 1990). His continued work to develop and refine a qualitative approach to the scoring of miscues provides a potential tool for the practitioner which is accurate and practical in its use with informal reading inventories. It is to this work that this review now turns.

May's Qualitative Scoring Systems

May (1982, 1986, 1990) heuristically developed systems of qualitatively analyzing miscues in order to assess background reading levels and provide instructional information. His research spans ten years of data collecting of individual case studies of students' performance on IRIs, from which he developed the CG, CGQ, and PR scoring systems, systems which help determine which miscues should be considered as errors and which as merely the reader's application of learned strategies to get at the author's meaning. May's research provides the theoretical basis for this study in which his CG, CGQ, and PR qualitative scoring systems are compared with other

quantitative traditional approaches to the scoring of informal reading inventories.

May's CG, CGO, PR Systems

May's (CG) Context, Graphophonic System provides, through two qualitative scales, a measure of context (C), and graphophonic (G) sensitivity. The CG system rates three miscues only: self-correction of substitutions, defaults, and meaning denigrating substitutions, using a 6 point scale for the use of context, and another scale for the number of letters in the reader's word that match the author's word; all other miscues are ignored.

May's (CGO) Context, Graphophonic, Questioning System is the same as CG but with the additional measurement of the student's effectiveness in answering inferential and literal questions developed by the IRI author.

In May's (PR) Poor Reader System, defaults, meaning denigrating substitutions (noncontextual substitutions -- horse for house, nonword substitutions -- bod for bad, and mispronunciation substitutions -- grāvel for gravel) are the only miscues examined. All others are ignored.

May discovered that while teachers were using IRIs more frequently, they used the commercially prepared quantitative scoring method rather than a qualitative approach. This approach resulted in penalizing good reading. This quantitative method of counting miscues, according to May, often produces the following results:

- * it often underestimates a child's reading level
- * its use often results in the prescribing of inappropriate skill lessons
- * it does not distinguish between good and poor reader errors.

Research Rationale

A brief review of the research on good and poor reader errors will help bring to the surface those problems related to a quantitative scoring method.

1. With quantitative procedures, insertions and omissions are considered negative features of reading and are counted against a reader. Yet, good readers not only substitute in a meaningful way, they also tend to omit and insert words in a meaningful way (Goodman, 1976; Goodman and Gollasch, 1980-81). In fact, D'Angelo and Mahlios (1983) have shown that "insertion and omission miscues made by either good or poor readers...cause very little syntactic and semantic distortion (p. 781).

Consequently, they recommended elimination of the time spent on coding and interpreting them.

2. In contrast to good readers, poor readers generally do much less self-correcting of their miscues (D'Angelo & Mahlios, 1982, p. 780).
3. Teachers are usually instructed to count repetitions against a student's score on the IRI. Yet the research conducted by May (1986) demonstrates that the repetitions tend to provide very ambiguous data to the teacher, that is they may reflect either good or poor reading behavior. By counting these repetitions, the teacher may be penalizing good reading (Goodman, 1968, p. 18; Guzak, 1970, p. 668).

It appears that by having teachers count all errors against a student's score, publishers cause teachers to

place students in the wrong instructional materials. Based on his own previously cited research, May determined that there are only a few miscues that consistently provide useful data for determining a student's placement in instructional materials:

- * self-corrections of substitutions (usually demonstrating that the student is letting two or more cueing systems interact, syntactic, semantic, graphophonic and schematic).
- * substitutions either with real words or nonsense words often labeled "mispronunciations" (can demonstrate message comprehension, lack of comprehension, and inappropriate use of graphophonic clues).
- * defaults (usually demonstrating that the student is afraid of errors, is doing very little predicting, or is unfamiliar with specific high-frequency words, phonograms, vowel patterns, or morphemes and needs to have the teacher pronounce the word).

THE FUTURE

After nearly a half century of inquiry into the value of IRIs, the following observations appear to be appropriate:

- * IRIs have evolved as a valuable tool to assess reading behavior and they are likely to be used in future years.
- * One of the most difficult issues to be resolved concerns appropriate criteria for the instructional level. Although there is considerable research attempting to validate criteria for independent, instructional and frustrational levels, none has been universally accepted.
- * With the growing use of miscue analysis, the question of which miscues should be significant also has become a subject of debate.
- * There is the need for a practical and accurate scoring system to be used with informal reading inventories that is research based.

It is believed that this study has the potential to address some of these specific needs related to the use of the informal reading inventory.

SUMMARY

The purpose of this literature review has been to discuss the various types of current reading assessment, the why and how of their usage, their appropriateness and in-appropriateness in the light of current research. Both formal and informal reading assessment practices are

currently used to evaluate potential and achievement. Often, the purposes of these assessments become confused and the results do nothing to contribute to the student's success as a learner. A qualitative miscue analysis scoring system to be used with IRIs has been demonstrated as an approach to assessment which can help foster student success.

This qualitative approach to assessment can provide information to the classroom teacher and reading specialists which can be used to support placement and instruction decisions. It is believed that this study will support such a methodology.

As a result of this surfaced need, the question is asked: can the existence of a qualitative scoring system be established by a descriptive study which compares such a system to existing traditional quantitative scoring systems and to other selected criteria.

This study attempts to identify a qualitative scoring system to be used with informal reading inventories. Such an approach to assessment should provide reading specialists and classroom teachers with appropriate placement and instructional information which will enable them to support the reading development of their students.

CHAPTER III

RESEARCH PROCEDURES

This chapter, which describes the study's research procedures, includes the following sections: (a) purpose, (b) design, (c) predictor and criterion variables, (d) research hypotheses, (e) setting, (f) subjects, (g) instruments and scoring systems, (h) data collecting procedures, and (i) data analysis procedures.

PURPOSE

The purpose of this descriptive quantitative study was to determine the accuracy and "practicality" of May's PR (poor reader) scoring system for the (IRI) informal reading inventory, an individual assessment device designed to identify a student's instructional reading level. This purpose was realized through a descriptive quantitative study using the Chi square test of significance. The PR was compared for practicality and accuracy with four traditional and largely quantitative scoring systems (T, TS, TR, TSR) and a qualitative system, the CGQ. These results were further compared to four criterion scores (CG, AE, TE, SR).

DESIGN

This descriptive study included the following procedures. Elementary school students were tape recorded reading aloud from an informal reading inventory (IRI) and their miscues scored, using four traditional scoring systems (T, TS, TR, TRS), and May's three qualitative systems (PR, CGQ). Each system's accuracy in placing each student at the appropriate instructional level was compared, through Chi square statistics, with the placement recommended by four independent criterion variables: teacher estimate (TE), IRI administrator estimate (AE), a silent reading score (SR), and a context graphophonic score (CG) based on the miscue and informal reading inventory research findings.

CRITERION AND PREDICTOR VARIABLES

The following describes the criterion and predictor variables in this study.

Criterion Variables

The criterion variables include: CG, AE, TE, and SR.

CG = Context-Graphophonic Scales: Rates three miscues only (self-corrections of substitutions, defaults, and substitutions), using a six point scale for the use of context, and another scale for the number of letters in the reader's word that match the author's word; all other miscues are ignored.

AE = IRI Administrator's Estimate: The IRI administrator's decision to place a student in a specific level of reading instruction materials, this decision based on knowledge of the recent research on miscues, observation of the student's behavior during the administration of the IRI, and thirty years experience and professional development in the teaching of reading. In this case, AE refers to the researcher in this study who gathered reading samples and analyzed the results.

TE = Teacher Estimate: A reading teacher's decision to place a student in instructional material based on criteria each teacher considers important, such as assessment scores, prior reading history, or even body language observed while a student reads. In this study, TE refers to the participating Chapter I Reading Teachers.

SR = Silent Reading: an instructional level estimate based on scored responses to passage-related inferential and literal questions developed by the IRI author.

Predictor Variables

The predictor variables include: T, TS, TR, TSR, CGQ, and PR:

T = Traditional: counts all errors made by the student during oral reading, including insertions, omissions, repeats, defaults, substitutions and self-corrections.

TS = Traditional minus Self-corrections: same as the traditional, but self-corrections are not counted.

TR = Traditional minus Repeats: same as the traditional, but repeats are not counted.

TSR = Traditional minus Self-corrections and Repeats: same as traditional but self-corrections and repeats are not counted.

CGQ = Context, Graphophonic and Questions: same as CG (see previous page) but with an addition of the student's score related to answering inferential and literal questions developed by the IRI author.

PR = Poor Reader: defaults and meaning denigrating substitutions (noncontextual substitutions, nonword substitutions, and mispronunciation substitutions) are the only miscues examined. All others are ignored.

RESEARCH HYPOTHESES

Chi square analysis was used to test the following research hypotheses.

Hypothesis One

CG as Criterion: There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of the six predictor variables (T, TS, TR, TSM, CGQ, PR) and each of the seven population groupings (Total, Girls, Boys, 2nd grade, 5th grade, 3rd grade, 4th grade), when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are

the number of agreements between the predictor estimates and the CG estimates.

Hypothesis Two

AE as Criterion: There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables and each of the seven population groupings, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the AE estimates.

Hypothesis Three

TE as Criterion: There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables and each of the seven population groupings, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the TE estimates.

Hypothesis Four

SR as Criterion: There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables and each

of the seven population groupings, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the SR estimates.

SETTING

Portland Public Schools, the site of this study, is the largest urban school district in Oregon. The total enrollment of this district is 53,500, of which approximately 39,000 are elementary students. Of the 60 elementary schools, 45 are designated Chapter I schools (Portland Public Schools Demographic Report, 1990). The population in these Chapter I schools is below the district mean socio-economic level as determined by free and reduced lunch applications.

The racial composition of the Portland Public Schools is most comparable to moderate sized mid-western urban areas such as Omaha, Nebraska and Minneapolis, Minnesota. Blacks comprise the largest minority group and Asians are the second largest non-white group. The population in Chapter I schools however, tends to have a larger minority composition than the other Portland schools (Johnson, 1990).

Ten Chapter I elementary schools in the north east section of the city were selected as the sites for this study. The selection was not random but based on practicality for the teachers and administrators and also on

accessibility to the researcher. The ten schools provided the identified population to be studied, 200 second through fifth grade Chapter I reading students. A more extensive description of the teachers and their subjects follows.

SUBJECTS

Teacher Participants

Prior to the beginning of the 1988-89 school year, a letter was mailed to each of the reading specialists in the north east area Chapter I schools, requesting their participation. Chapter I schools are schools which receive additional intervention funding from the federal and state government based on the number of free and reduced student lunch applications.

These ten Chapter I schools were selected as the sites for the data collecting because of their geographic accessibility to the researcher. The teachers were chosen because they were the reading specialists assigned to the Chapter I programs and agreed to participate in the study.

The IRI administrator scheduled a visit to each school during which the details of the plan were shared. During the visits, all ten teachers agreed to participate in the study. A survey focused on teacher and student demographics and Chapter I selection methods was distributed at this time. Nine of the ten teachers completed the survey.

Subject Selection

The ten Chapter I reading teachers each selected twenty Chapter I reading students per school (200n) from grades two through five. This resulted randomly in 100 of each sex. Table XII describes the sample population.

Participating Teachers Survey

A participating teacher survey collected demographic information about teacher training, experience and certification. The survey identified and ranked current Chapter I student - selection methods. Finally, teacher perceptions of the sample populations' socio-economic status, basic family structure and population mobility were gathered.

Survey

The teacher survey information was collected because of its potential influence on the student reading level estimates provided by the participating teachers. The teachers' estimates (TE) were used as one of the criterion variables. This information will be further discussed in the results section.

Survey Validity

Two procedures were used to establish the validity of the teacher survey. They were: (a) experts' reviews of the questions followed by recommended revisions, and (b) field testing.

Expert Review. An expert review of the survey was done by Dr. Brad Eliot aka Dr. Frank May and Dr. Jim Leffler. Dr. Eliot is a Professor of Education at Portland State University. His area of expertise is reading. Dr. Jim Leffler, past Chapter I program monitor, is currently a coordinator of assessment for a local school district. Recommendations by Dr. Eliot and Dr. Leffler resulted in several question revisions on the teacher survey.

Field Testing. The second procedure used to establish validity of the survey was field testing. Two reading specialists from a local school district completed the survey. As a result of this field testing, several questions underwent revisions based on verbal feedback and notes from participants. Tables III through XII summarize the obtained results.

TABLE III
 PARTICIPATING TEACHERS TRAINING AND EXPERIENCE
 1989

Question	Total	Lowest	Highest	Average
Reading Courses Taken	66	0	15	7
Years Teaching Experience	13	28	22	15
Certifications/ Degrees Obtained	17	1	3	1.8
Basic Certificate	9			
Masters Degree	3			
Reading Endorsement	5			

Note. N = 9 of 10 Reading Teachers Reporting

TABLE IV
CURRENT METHODS USED BY PARTICIPATING TEACHERS
TO SELECT CHAPTER I READING STUDENTS
1989

Method	Number of Teachers
Portland Area Levels Test	9
Standardized group test	8
Teacher-made assessment	7
Classroom reading record	6
Prior Chapter I service	5
Classroom teacher referral	4
Basal placement	4
Informal Reading Inventory	1
Teacher observation	1
Other: Parent Referral	1

Note. N = 9

TABLE V
THE STANDARDIZED GROUP TEST IS A REASONABLY ACCURATE
METHOD OF IDENTIFYING A STUDENT'S INSTRUCTIONAL
READING LEVEL
1989

Rating	Number of Teachers	Percent
Strongly agree	1	11%
Agree	2	22%
Mildly disagree	2	22%
Disagree	0	0%
Strongly disagree	1	11%

Note. N = 9

TABLE VI

AN INFORMAL READING INVENTORY IS A REASONABLY
ACCURATE METHOD OF IDENTIFYING A
STUDENT'S INSTRUCTIONAL READING LEVEL
1989

Rating	Number of Teachers	Percent
Strongly Agree	1	11%
Agree	1	11%
Mildly agree	2	22%
Mildly disagree	3	33%
Disagree	1	11%
Strongly disagree	1	11%

Note. N = 9

TABLE VII

THE CURRENT METHOD USED TO SELECT STUDENTS
FOR PROGRAM IS REASONABLY ACCURATE
1989

Rating	Number of Teachers	Percent
Agree	2	22%
Mildly Agree	3	33%
Mildly Disagree	1	11%
Disagree	0	0%
Strongly Disagree	0	0%

Note. N = 9

TABLE VIII
COMPOSITE RANKING OF TEACHERS' PREFERRED
METHODS OF SCORING IRIS
1989

Scoring Description	Ranking	Total Score
(Range: 1 = First Choice - 7 = Last Choice)		
Count all errors	1	14
Rely primarily on "comp." questions	2	17
Count all errors but self-corrections	3	19
Count all errors but repeats	4	23
Count all errors using CG scales	5	24
Count all errors but self-corrections and repeats	6	29
Count only "poor reader" errors	7	32

Note. N = 9 teachers reporting (6 complete responses, 2 partial responses, 1 no response to this question).

TABLE IX
TEACHERS' PERCEPTIONS OF THE CHAPTER I STUDENTS'
SOCIO-ECONOMIC STATUS
1989

Condition	Number of Teachers	Percent
Lower	6	66%
Low Middle	2	22%
Middle	1	11%

Note. N = 9

TABLE X
TEACHERS' PERCEPTIONS OF THE
CHAPTER I STUDENTS' BASIC FAMILY STRUCTURE
1989

Family Structure	Number of Teachers	Percent
Single Parent	6	66%
Two parent non-biological	2	22%
Two parent biological	1	11%

Note. N = 9

TABLE XI
TEACHERS' PERCEPTIONS OF THE
CHAPTER I STUDENTS' MOBILITY
1989

Mobility	Number of Teachers	Percent
High	4	44%
Medium	2	22%
Moderately Stable	3	33%

Note. N = 9

TABLE XII
DESCRIPTION OF CHAPTER I STUDENT POPULATION
1989

Grade	Age Range	Boys	Girls	Total
2	7-9	34	35	69
3	8-10	27	26	53
4	8-11	20	24	44
5	10-12	19	15	34

Note. N = 200

INSTRUMENTS AND SCORING SYSTEMS

The instruments used to collect and score the data from the study included the following:

- * the Silvaroli Informal Reading Inventory also known as the Classroom Reading Inventory
- * the traditional quantitative scoring system developed by Betts (with three scoring variations)
- * three qualitative scoring systems developed by May.

The Classroom Reading Inventory

The Silvaroli Classroom Reading Inventory was used to collect miscue data. This testing instrument was selected on the basis of its long tradition of use by classroom

teachers and reading specialists (Powell & Dunkeld, 1971; Silvaroli, 1967-85). Furthermore, the author's use of pictures related to the passages adds to the authenticity of this assessment tool.

The Silvaroli Classroom Reading Inventory (1967-85) is designed to provide the teacher with information concerning the student's independent, instructional and frustration reading levels. In addition, it attempts to assess the student's comprehension ability through questions related to each passage. By using it also as a diagnostic tool, the teacher can determine the reading student's specific strengths and weaknesses through an analysis of oral miscues.

This individually administered IRI consists of a series of graded passages pre-primer through level eight, each of which is followed by passage related inferential and factual questions. The student completes the oral reading of specific passages (and optionally the silent reading of similar passages) and then the teacher estimates the student's comprehension by asking for responses to the passage related questions.

Johnson (1978) reviews Silvaroli Classroom Reading Inventory with the following statements (paraphrased by the researcher):

- * provides a formula for determining independent, instruction and frustration levels
- * passages appear to possess appropriate level to level difficulty and satisfactory interest
- * questions are well-formed and passage dependent
- * compares favorably with other published inventories
- * passages are deliberately brief to control test time
- * a motivational statement provides a context for the reading passage
- * it is a useful tool in the hands of a knowledgeable teacher
- * pictures add authenticity to this assessment tool (p. 1234).

The Validity of the Classroom
Reading Inventory

Validity, in general, is the degree to which a test measures what it purports to measure. A test is valid only for a particular purpose and for a particular group. Tittle (1989) states that, "It is both the test maker's and the test user's expertise that contribute to a test's validity" (p. 4). Tittle further suggests that, "Although the test maker's perspective on what a test's scores mean and whether

they are useful is a necessary one, this perspective must be expanded to include the teacher's and student's support to validate a test's usefulness" (p. 5).

The review of the literature in Chapter II provides support for the statement that the IRI is the most frequently used means to determine reading levels and diagnose reading behaviors (Dechant, 1981; Farr & Carey, 1985; Pikulski & Shanahan, 1982). More specifically, the Silvaroli Classroom Reading Inventory has a long history of wide use by classroom teachers and reading specialists (Cramer, 1980; Johns & Lunn, 1984; May, 1986; Powell & Dunkeld, 1971). It was developed in 1967, has been revised periodically, and used extensively to the present time. Although research on specific elements of the Silvaroli Classroom Reading Inventory is limited, the CRI has often been included in studies of informal reading inventories in general (Harris & Niles, 1982; Johns & Lunn, 1982; Johns & Magliari, 1988; McKenna, 1983; Powell & Dunkeld, 1971; Pikulski, 1982). In a study by Maztal and Smith (1984) the overwhelming majority of elementary teachers questioned indicated they were familiar with and knew how to administer IRIs. The Silvaroli Classroom Reading Inventory was one of the IRIs specified in the study. Using Tittle's rationale regarding teacher support, this frequent use by classroom teachers and reading specialists probably contributes to the validity of the Silvaroli Classroom Reading Inventory.

Content Validity

It is important to seek evidence to support the validity of a specific test. Both content validity and concurrent validity will be discussed in relation to the Silvaroli Classroom Reading Inventory, in order to provide rationale for its use in this study.

Content validity, as defined by Gay (1987) "is the degree to which a test measures an intended area" of knowledge or skill acquisition (p. 129). He further posits that "content validity is determined by expert judgment; there is no formula by which it can be computed and there is no statistical way to express it" (Gay, 1987, p. 130).

The test administrator in this study assumed the role of "expert" and determined that the Silvaroli Classroom Reading Inventory was content-valid as a reading test for this study on the basis of the following criteria:

- * utility
- * well-established readability levels
- * passage-related questions
- * use of text-related pictures.

These criteria which are used to support the content-validity of the Silvaroli Classroom Reading Inventory will now be discussed.

Utility. Utility is often a better indicator of validity than traditional validity (Tittle, 1989). The importance of the utility of a test is further supported by

Ebel (1967). Besides the meaningfulness of the test scores, Ebel cites two "very important elements for judging the quality of a test: One is the importance (usefulness as a basis for effective or satisfying behavior) of the knowledge or abilities required by the test. The other is the convenience of the test in use" (Ebel, 1967, p. 230). The CRI has demonstrated that it provides the teacher with useful information in an easily accessible manner.

Well Established Readability Levels of Passages. The grade levels of the passages were established using the following respected readability measures:

Dale & Chall, 1948; Flesch, 1943; Fry, 1972; Spache, 1960; (cited in Silvaroli, 1983, p. 4).

Passage Related Questions. Passage related questions are essential to a good IRI, for they lessen the effect of prior knowledge on the question answering task. In the third edition of the Silvaroli Classroom Reading Inventory, Silvaroli (1982) states that he reviewed and edited all the questions to insure that they are all passage dependent. Although such an attempt is laudable, the experience of this IRI administrator finds that a student's prior knowledge and schemata use are factors which consistently effect questioning results.

Use of Text Related Pictures. The pictures normally add to the quality of this IRI, because they increase schemata background and the student's interest in the

passage. In addition, the pictures increase the validity of this IRI by making the text similar to currently used instructional materials (literature, basals, content area texts).

Concurrent Validity. Concurrent validity is the degree to which one test ranks students in the same way that another already established test or other valid criterion does. There have been several concurrent validity studies comparing the results of several IRIs with other well-established tests (Betts, 1940; Killgallon, 1942; Powell & Dunkeld, 1971).

Comparison With Standardized Tests

Compared to norm referenced tests, the ability of IRIs to determine functional reading levels appears to hold up consistently (Betts, cited in Johns & Magliari, 1989; Killgallon, cited in Shanahan, 1983). In another study by Coleman and Harmer (1982), the Word Identification and Passage Comprehension subtests of the Woodcock (a standardized test), and IRI instructional level placement were found to compare favorably, 2.7 - 2.9. (p. 370)

McCracken's (1964) study is still one of the most extensive studies of the validity of an informal reading inventory, although he found that the standardized test tends to estimate higher instructional levels than the IRI. Farr and Carey (1985), however, state:

Such comparisons between IRIs and standardized tests are entirely dependent on the particular standardized test used, the materials used to construct the IRI, the criteria used to judge performance on the IRI, and the ability and skill of the examiner in recording errors and judging performance on the IRI. (p. 168)

Bristow, Pikulski, and Pelosi (1983) compared the estimates of reading levels for five different assessment approaches. The researchers found that a teacher made IRI, a commercial IRI, and the MAT (Metropolitan Achievement Test) results produced reading level estimates that were within one grade level of one another. The MAT is a published reading test that provides criterion referenced scores which can be used to assign students to instructional reading levels.

It should be pointed out, however, that it is not likely that researchers will ever get large correlations between IRI estimates and standardized estimates. The reasons for this are these: 1) standardized reading test scores are low on validity and high on reliability. Such lack of validity on the part of standardized reading scores was discussed in Chapter Two. 2) IRI scores, on the other hand are high on validity and low on reliability. This problem will be addressed next.

Farr & Carey (1986) offer this summary statement regarding the general validity of IRIs:

IRIs seem to provide some of the best information for planning reading instruction. This information must be accepted by school administrators and the lay public as important to

the student's reading development. Such informal testing should not be thought of as "second class" information that is useful only when more formal information is not available. Indeed, such testing holds the greatest promise for increasing the validity and reliability of information used for instructional planning. (p. 175)

The Reliability of The Classroom Reading Inventory

Reliability is the degree to which a test consistently measures whatever it measures. Alternate forms reliability is a commonly used estimate of reliability for tests used in research.

Helgren (1984) conducted an analysis of alternate form reliability of three commercially prepared informal reading inventories including the Silvaroli Classroom Reading Inventory. Results indicated that the coefficients ranged between .61 and .78. this researcher declared that little error could be directly attributed to the alternate forms, as the largest ariane component was attributed to the subjects themselves. The researcher concluded that while the results did not reveal perfect reliability, critics could not condemn the IRIs for being unreliable. Reading researchers, rather, according to the researcher, need to "address the question of what an acceptable level of reliability would be for informal reading inventories" (Dissertation Abstracts Service No. ADB84-16328, p. 1082).

Alternate Forms Reliability of the Silvaroli

Alternate forms of a test are designed to provide a researcher and the practitioner choices for such purposes as pre and post tests and test-retest situations, with assurance of similar results. McCracken (1964) studied the reliability between alternate forms of an IRI and suggested that they could, from a practical standpoint, be used interchangeably. Others (Pikulski and Shanahan, 1984) disagree.

Silvaroli has developed parallel forms A, B, C, and D for the Classroom Reading Inventory. In fact, he recommends the use of Form A to measure oral reading errors and Form B to measure silent reading comprehension, which is what the present researcher did. Silvaroli (1987) supports the reliability of these parallel forms with the following discussion:

Form B is similar in design but not content to form A. Thus form A and form B can be interchanged. The passages in form B can be used in the following ways:

- * As an additional set of oral passages for post-testing
- * As a set of silent paragraphs for students who might reject oral reading
- * As a set of silent paragraphs to enable the teacher to give an oral paragraph and a corresponding silent paragraph for a more complete assessment of the student's overall reading achievement
- * Forms A and B can be used interchangeably (p. 4).

The readability studies conducted by Silvaroli and discussed in the previous section contribute somewhat to the possibility that forms A and B are reliable parallel forms. That is, the selections used for the parallel forms have the same readability levels, as determined by a readability formula.

In order to examine such a potential weakness in this study, that is, the lack of data supporting the reliability of the parallel forms, a simple Pearson r correlation comparing the oral question results (Form A) and the (Form B) silent question results on levels 1, 2, and 3 for 30 students was obtained. The result was a reliability coefficient of only 0.34. For formal standardized tests a .90 level is generally accepted as demonstrating the degree of relationship needed to support their reliability.

Both schemata influence on comprehension and the effect of oral versus silent reading on obtained results need to be considered when examining these results, as well as the previously mentioned issue of the "acceptable level" of reliability for IRI parallel forms.

It should be further noted that, as for reliability of the IRI method, the traditional split-half and parallel-form estimates don't often apply well to this problem. By using parallel forms of the IRI, for instance, one introduces the traditional problem of comparing two texts that are supposed to have the exact same level of readability. And yet,

because of schemata differences in the two selections, the texts may not really be that parallel at all. A more relevant and practical form of reliability for the IRI may be an interjudge type, which estimates how well teachers agree on a child's placement when using the same scoring system. This form of reliability estimate will now be discussed.

Interrater Reliability

This form of reliability was estimated by having the researcher and a teacher trained by the researcher independently assess the instructional levels of the same thirty students. The audio taped oral reading of thirty randomly selected students from the original sample of 200 were scored using the sex scoring methods (T, TS, TR, TSR, CGQ, PR). The interjudge reliability coefficient for the results was .99 as determined by a simple Pearson r correlation. This significant result supports the scoring methodology and will be discussed further in Chapter V.

BETT'S TRADITIONAL QUANTITATIVE SCORING SYSTEM

The traditional quantitative scoring system developed by Betts (1946) was one method selected to score the collected miscues for this study. It was Betts (1941) who established the three reading levels for the IRI: independent, instructional, and frustration, and assigned a criterion score range (shown later) for each of these

levels. In addition, this has been the method most often used by practitioners to score informal reading inventories since the 1940's. As discussed in the review of literature, there has been criticism of this methodology, but no other system has yet been accepted as its replacement (Ekwall, 1974; Hood, 1976; Johns & Magliari, 1988; Powell, 1978).

Bett's Scoring Criteria

The criteria used by Betts (1946) are based on a quantitative scoring (percentage) of words pronounced and questions correctly answered with respect to the oral reading of a specific passage. Betts rated each student's performance according to the following criteria:

- * Independent reading level: 98% words correct
 90% questions correct
- * Instructional reading level: 94% words correct
 60% questions correct
- * Frustration reading level: below 90% words
 correct
 below 50% questions
 correct.

According to this traditional method of scoring, all repeats (resayings), omissions (words omitted during the oral reading of a passage), substitutions (words, non-words, or mispronunciations used by the student during oral reading to replace the author's intended text), all self-corrections, defaults (teacher's pronounced word during student's oral reading after five seconds of waiting),

insertions (words added to original text by student during oral reading), are counted as errors.

Variations of Bett's Traditional Scale

The following variations of this traditional scale were used to score the obtained data for the present study:

(TS) traditional minus self-corrections (same as traditional but self-corrections were not counted); (TR) traditional minus repeats (same as traditional but repeats were not counted); (TSR) traditional minus self-corrections and repeats (neither self-corrections nor repeats were counted).

MAY'S QUALITATIVE SCORING SYSTEMS FOR THE IRI

The other methods used to score the subjects' miscues were May's three qualitative systems. One of these systems, the CG (context-graphophonic) system, serves as one of the criterion variables, since it was developed by May to complement as closely as possible the most recent research on miscues. An explanation of the CG system follows.

The CG System

The CG system requires that the teacher or researcher rate three kinds of miscues and ignore the rest. The three rated are substitutions, self-corrections of substitutions, and defaults. As mentioned in Chapter Two, omissions, insertions, and repeats have been shown by researchers to be seldom important with respect to comprehension. Teachers

and researchers accomplish their ratings of miscues through May's two CG rating scales, shown on the next page.

The Context Scale for the CG System

Use of the C (context) score provides an estimate of a student's relative ability to approximate the author's word meaning within the context of the surrounding text. For each substitution a ratio is determined by using May's context scale. The denominator for a context ratio is always six; the numerator varies, however, with the proximity of the substitution to the author's meaning. A context ratio of 3/6, for instance, would be given for a substitution of "smash the curtain" for "slash the curtain." The student would be rewarded only for coming up with a substitution of the same syntax class, a verb. A ratio of at least 4/6 would be given for a substitution of crash the car for smash the car. A ratio of 5/6 would be given for a substitution of house for home or mash for crush.

For each self-correction of a substitution the ratio is always 6/6. Again the CG system provides credit for self-monitoring behavior. (No other type of self-correction is considered and there is no such thing as a self-correction of a default.)

For each default the context ratio is nearly always 0/6. The only exception is when a child uses a substitution and asks if it is correct. The teacher or researcher counts this as a default since she has to provide him with the

correct word. In this case, though, the context ratio is the same as it would be for a substitution.

The context ratios for all of the appropriately rated miscues in the selection are then added up and changed to a percent. For example, ratios of 4/6, 3/6, 0/6, 6/6, and 5/6 combined would yield a ratio of 18/30 and a percentage of 60. Thus 60 % is the student's context score for this selection.

The Graphophonic Scale for the CG System

Use of the G (graphophonic) scale provides an estimate of a student's capacity to recognize and decode letters in sight words and graphophonic patterns. For each substitution, a ratio is computed for the number of letters a student decodes, divided by the total number of letters in the word. A student who substitutes gravy for gravel, for instance, receives a ratio of 4/6. If the student pronounces the substitution with more phonemes than the author used, one point is subtracted from the numerator of the ratio. For example, a student who says either gravels or graveling for gravel, receives a ratio of 5/6.

For each default, the teacher or researcher must first determine whether the student said nothing before the teacher provided the word (after a five-second pause), or whether the student said part of the word, then stopped trying. For instance, a student who said only /ma/ for the

word, materials, would receive a graphophonic ratio of 2/9. A student who said nothing would receive a ratio of 0/9.

For each self-correction of a substitution, the numeral and denominator of the graphophonic ratio must be the same--the exact number of letters in the word. For instance, a student who says manners for materials, then corrects his own miscue, receives a graphophonic ratio of 9/9. The CG system provides credit for self-monitoring behavior.

The graphophonic ratios for all of the appropriately rated miscues in the selection are then added up and changed to a percent. For example, ratios of 4/6, 5/8, 0/9, and 7/7 would yield a combined ratio of 16/30 and a percentage of 53. Thus, 53% is the student's graphophonic score for this selection.

DEMONSTRATION OF THE CG SCORING PROCEDURE

The Appendix displays the application of the CG scoring procedure. A paragraph is marked with the miscues made by the student during audio taping of oral reading. The system used to score the miscues, also displayed, is explained in detail.

May's Criteria Based on CG Scores

The following are the criteria established by May for determining whether a selection represents a student's instruction, frustration, or independent level of reading.

Instruction Level. A minimum score of 60% on the context and graphophonic score; and or an average score of 65%.

Frustration Level. Less than 60% on both the context and graphophonic score; less than 60% average score.

Independent Level. 80% or better on both scales.

The CGQ System

May's CGQ system differs from the CG system in two ways. First, in addition to a C (context) score and a G (graphophonic) score, a Q score is also computed to reflect the quality of the answers a student gives to questions asked by the IRI administrator after the oral reading has been completed. A rating of 0 to 2 is made for each answer.

In this study five questions were asked by the IRI administrator after each oral reading. Since the questions and answers were tape recorded, the researcher was able to listen to them later and rate them on the 0 to 2 scale. Since the maximum points on the questions were 10, a percentage score from 0 to 100 was easily computed. For example, a student with scores of 2, 1, 0, 2, and 2 would receive 7 points and a Q (question) score of 70%.

A separate C, G, and Q score was possible for each student, providing a more thorough profile than the CG system. Appendix A shows the criteria established by May for determining whether a selection represents a student's instruction, frustration, or independent level of reading.

Note, as a second difference, that May used lower requirements for C and G scores than he did for the CG score, thus allowing for more flexibility in the total profile.

May's Criteria Based on CGQ Scores

Instruction Level. 40% absolute minimum on the context score and on the graphophonic score; 50% absolute minimum on the question score; 60% absolute minimum on the average score for C,G, and Q scores.

Frustration Level. Less than 40% on either the C or G score; less than 60% on the average score for C, G, and Q scores.

Independent Level. 70% absolute minimum on the context score and on the graphophonic score; 80% absolute minimum on the average score for C, G, and Q scores.

The PR System

The most important predictor variable and the real focus of this study is the PR (Poor Reader) Scoring System because of its potential practicality and accuracy for use by classroom teachers in identifying the instructional reading level of students. Can it work as well as the CG scoring system? Is it better than the traditional systems? Here is how this score is determined.

The PR Scoring System, the third of May's qualitative scoring systems, determines the student's instructional

reading level by counting only defaults and those substitutions which demonstrate that the reader does not understand the schema that is the author's intended meaning. These two types of miscues seem to differentiate a poor reader from a good reader more than any other miscue. For example, a level five passage read by a student produced 12 traditional "errors". Of those errors, only three were considered "poor reader" errors, that is defaults, and substitutions which change the author's meaning. They were: "express" for "experts", "make" for "tell", and "lap" for "tap". Three errors (3/102) gives the student a poor reader score of 3%, well within the established criteria of 2-5% for instructional level. Above 5% is considered frustrational; below 2% is considered independent.

DATA COLLECTING PROCEDURES

Student Selection

Prior to the beginning of the 1988-89 school year, Chapter I teachers in the north east area of Portland Public Schools were invited to participate in the study. The ten teachers who agreed to participate selected twenty students from their second through fifth grade Chapter I population to be tape recorded reading the Silvaroli Informal Reading Inventory. Chapter I students are those students who participate in an extension of their regular instruction previously based on an identified need for strengthening of

reading performance. The teacher simply selected students from their classes who were accessible, keeping in mind the need to balance numbers in terms of grade levels and sex.

An additional element which contributed to the quality of the obtained oral reading samples was the fact that the students "volunteered" to participate. They were informed that they were helping the IRI administrator obtain information to become a better reading teacher. In addition, they were allowed to listen to their responses at the end of the taping session, a new experience for most of them. Some students refused to participate and were therefore eliminated from the study.

Privacy and Informed Consent

A coding system consisting of the teacher's initials and a case number was used to identify the oral reading samples, assuring anonymity and adherence to human subjects procedures. In addition, no reading instruction took place as part of the study, no information regarding the child or his family was gathered, and no assessment was made regarding personality, attitudes or intelligence. No information was given by the test administrator to the teacher on the reading level obtained for individual children. The only information given to the teacher was the general results of the research, that is, what scoring systems worked best.

This procedure eliminated the need to obtain permission from parents in order for children to participate in the study. However, an informational letter was sent by each teacher informing parents of what was occurring and allowing parents the option to deny their child's participation. No parent declined permission.

Oral Reading Sample Collecting

Two hundred oral reading samples were collected by the test administrator on audio-cassettes from January, 1989 through March, 1989, using the Silvaroli Classroom Reading Inventory (IRI). The details describing the collection procedures follow.

Taping Procedures

Each student was escorted by the test administrator to a previously selected room, where the materials and equipment were available. These included a tape recorder, tapes, and both student and administrator's IRIs.

The student was made to feel at ease by an initial informal conversation. This conversation included sharing: that the procedure was to help the test administrator learn more about how to teach reading; that there would be no grade given; and that the student would have an opportunity to listen to a portion of the recording.

Many of the students had never been tape recorded before, therefore the test administrator discussed the

procedure, answered questions, and in some cases let the student just talk into the recorder and hear it before beginning the reading session. Once rapport was established, the appropriate level of the Silvaroli Classroom Reading Inventory was placed in front of the student and the student was asked to begin reading the passages. The sessions lasted approximately twenty-five minutes per child.

Each student began with a reading selection two levels below grade placement, and read as many passages as necessary, from pre-primer through grade eight, until a frustration level was obtained.

Coding Procedures

Following the administration of the IRI to each of the 200 students, the IRI administrator listened to each tape at least twice, coding the miscues on a copy of the original test pages. The coding used was that recommended by Frank May (1986). A sample of the coding is as follows:

the bulls *d were*
 Before a bull fights, some people wait in the streets.

Scoring Procedures

Once the miscues were coded for a selection, they were scored seven times, using the four traditional and three qualitative methods already described. The instructional level was determined for each of the 200 students with each

of these seven scoring systems. One system, for instance, might lead to an estimate of grade 5 for instruction; another might lead to an estimate of grade 4 for instruction; the other systems might lead to an estimate of grade 6.

The instructional level for each scoring system is based on the particular criterion used by each scoring system. For example, the traditional scoring system counts all errors (substitutions, omissions, repeats, defaults, insertions, self-corrections) and computes them against the total number of words in the passages. The percentage formula established by Betts is then applied to determine the instructional reading level results.

DATA ANALYSIS PROCEDURES

Chi square was used to test the differences between expected and observed frequencies of students being "accurately" placed with respect to each criterion. Leading up to the Chi square computations, the six predictor variables (T, TS, TR, TSR, CGQ, PR) were first compared with each of the four criterion variables (CG, AE, TE, SR) for each of the populations (Total, Girls, Boys, 2nd, 5th, 3rd, 4th grade). An example for the total population will demonstrate this procedure.

As a result of these comparisons the researcher found that the PR predictor score placed 181 students, or 90.5%,

at the same instructional level as the CG criterion score. Furthermore, the PR score placed 6 students, or 3%, below the level that the CG score placed them. It also placed 13 students, or 6.5%, above the level that the CG score placed them. These frequencies, then, 6 (3% below), 181 (90.5% the same), and 13 (6.5% above) were used as the observed frequencies in the Chi square calculations.

The expected frequencies for the 200 students were 10, 180, and 10, or 5%, 90%, and 5%. That is, 90% of the population was expected (required) to be placed by the predictor variable as accurately as the criterion variable placed them; 5% were expected (allowed) to be placed too low and 5% too high. This permitted a total of 10% error, and from a practical standpoint could yield a scoring system that would misplace no more than 3 students in a class of 30. (In this study the expected frequencies for the three-by-two cell Chi square procedure were always the same percentages of whatever portion of the population was being examined. That is, they were always 5%, 90%, and 5%, whether the total population or only the second grade population was being examined.)

SUMMARY

Chapter Three presented the purpose and the design of the study, as well as a description of the predictor and criterion variables. Also presented and discussed were the

research hypotheses, subjects, setting, and data collecting methods and procedures. Chapter Four presents the data analysis.

CHAPTER IV

RESULTS

INTRODUCTION

The purpose of this descriptive quantitative study was to determine the accuracy of May's poor reader (PR) scoring system for the informal reading inventory (IRI), an individual assessment device designed to determine a student's instructional reading level. The PR compared four traditional and largely quantitative scoring systems (T, TS, TR, TSR) and a qualitative system, (CGQ). The predictor variables were individually tested for "accuracy" against each of four criterion estimates of students' instructional levels (CG, AE, TE, SR). For this study, the T, TS, TR, TSR, CGQ, and PR are the predicted variables and CG, AE, TE, and SR are the criterion variables. The data presentation is organized to address the research hypotheses. The following sections are included in this chapter: 1) Statistical Method; 2) Variables; 3) Population Groupings; 4) Research Hypotheses; 5) Data Presentation; 6) Summary.

STATISTICAL METHOD

Chi square was used to test the differences between expected and observed frequencies of students being "accurately" placed with respect to each criterion. An example presented in the last chapter of how the Chi square data were obtained are repeated for the convenience of the reader.

Example: In examining the raw data for the total population, the researcher found that the PR predictor score placed 181 students, or 90.5% at the same instruction level as the CG criterion score. Furthermore, the PR score placed 6 students, or 3%, below the level that the CG score placed them. It also placed 13 students, or 6.5%, above the level that the CG score placed them. These frequencies, then, 6 (3% below), 181 (90.5% the same), and 13 (6.5% above) were used as the observed frequencies in the Chi square calculations.

The expected frequencies for the 200 students as previously explained (p. 96) were 10, 180, and 10, or 5%, 90%, and 5%. That is, 90% of the population was expected (required) to be placed by the predictor variable as accurately as the criterion variable placed them; 5% were expected (allowed) to be placed too low and 5% too high. This permitted a total of 10% error, and from a practical standpoint could yield a scoring system that would misplace no more than 3 students in a class of 30. (In this study

the expected frequencies for the three-by-two cell Chi square procedure were always the same percentages of whatever portion of the population was being examined.) That is, they were always 5%, 90%, and 5%, whether the total population or only the second grade population was being examined.

VARIABLES

Six predictor variables were individually compared with each of four criterion variables in order to check the relative accuracy of the predictor variables in predicting students' instructional reading levels. The two types of variables are as follows:

- (T) Traditional scoring
- (TS) T minus Self-corrections
- (TR) T minus Repeats
- (TSR) T minus Self-corrections and Repeats
- (CGQ) Context-Graphophonic-Question scales
- (PR) Poor Reader scoring Criterion Variables
- (CG) Context-Graphophonic scales
- (AE) Administrator Estimates
- (TE) Teacher Estimates
- (SR) Silent Reading score

POPULATION GROUPINGS

Seven population groupings of the 200 student sample were used to examine the data. Those groupings included:

- * total n = 200
- * girls n = 100
- * boys n = 100
- * 2nd grade n = 69
- * 5th grade n = 34
- * 3rd grade n = 53
- * 4th grade n = 44

RESEARCH HYPOTHESES

Chi square analysis was used to test the four Research Hypotheses.

Hypothesis One: CG as Expected Criterion

There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables and each of the seven population groupings, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the CG estimates.

Hypothesis Two: AE as Expected Criterion

There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables and each of the seven population groupings, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the AE estimates.

Hypothesis Three: TE as Expected Criterion

There will be no significant differences between observed and expected frequencies of students places below, at, and above their instructional reading level, for each of six predictor variables and each of the seven population groupings, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the TE estimates.

Hypothesis Four: SR as Expected Criterion

There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables and each of the seven population groupings, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of

agreements between the predictor estimates and the SR estimates.

DATA PRESENTATION

The remainder of this chapter presents the results of Chi square analysis for each hypothesis (Tables XIII through XXXX). These tables represent the data analysis related to each population group (Total, Girls, Boys, 2nd, 5th, 3rd and 4th grade) with respect to the expected and observed frequencies of the criterion and predictor variables. Frequencies, percents and Chi square values are presented. Amplification of these results and a discussion of the Participating Teacher Survey findings will be presented in Chapter V.

Hypothesis 1A

CG Criterion--Total Population. For the total population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the observed frequencies are the number of agreements between the predictor estimates and the CG (context-graphophonic scale) estimates.

Frequencies, percents, and Chi square values for the predictor variables with the CG criterion variable, total population, are presented in Table XIII.

TABLE XIII

CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
 FREQUENCIES FOR THE CRITERION VARIABLE CG
 FOR THE TOTAL POPULATION
 (N = 200)
 1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
10	5	180	90	10	5		
OBSERVED							
Predictor Variables	Below N	Below %	At N	At %	Above N	Above %	Chi Square
T	129	64.5	68	34.0	3	1.5	1490.7*
TS	124	62.0	72	36.0	4	2.0	1368.0*
TR	125	62.5	72	36.0	3	1.5	1284.2*
TSR	119	59.5	77	38.5	4	2.0	1147.6*
CGQ	2	1.0	197	98.5	1	0.5	16.1*
PR	14	7.0	181	90.5	5	2.5	4.1

Note. * Indicates significance demonstrated at .01 level p < .01 = 9.210

For the total population, significant differences were found at the p < .01 level between five of the six predictor variables T, TS, TR, TSR, CGQ, and the criterion variable CG, in estimating the students' instructional reading levels. For the same total population, nonsignificant differences were found between the predictor variable PR and

the criterion variable CG in estimating the students' instructional reading levels. Therefore, Hypothesis 1A was refuted with respect to the predictor variables T, TS, TR, TSR and CGQ, and supported with respect to the predictor variable PR.

Table XII shows the results for the total population of 200 children from grades two through five. Statistically significant differences were found between the instructional placements estimated by the four quantitative scoring systems and the CG criterion system. With the quantitative systems the instructional estimates coincided with those of the CG criterion only 35% to 38% of the time. From 60% to 64% of the time the quantitative systems estimated below the CG criterion estimate, and from 0.5% to 2.5% of the time they estimated above.

There was no statistically significant difference between the PR qualitative placements and the CG criterion placements. The PR system placed students the same as the CG system 90.5% of the time, with 7.0% below and 2.5% above. Since no more than 10% of students were "misplaced," the PR system met the "practicality" requirement of the study.

The CGQ placements were significantly different from the CG placements. However, this difference turned out to be a statistical anomaly, caused by the CGQ system's success in estimating as well as the CG system. Only 1.0% of the students were placed lower and only 0.5% of the students

were placed higher than the CG placement. The observed frequencies were 2 below, 197 the same, and 1 above while the expected were 10, 180, and 10 respectively. The observed yielded 34 cases different from the expected and thus the difference from a statistical standpoint. The CGQ system met the "practicality" requirement, with 98.5% of the placements the same as those of the CG criterion.

Hypothesis 1B

CG Criterion--Population Girls. For the population of girls, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the CG estimates. Frequencies, percents and Chi square values for the predictor variables with the CG criterion variable, population girls, are presented in Table XIV.

For the population of girls, significant differences were found at the $p < .01$ level between each of the four traditional predictor variables T, TS, TR, TSR and the CG criterion variable in estimating the students' instructional reading levels. Nonsignificant differences were found between the two qualitative predictor variables CGQ and PR, and the research-based criterion variable CG in estimating the students' instructional reading levels. Therefore,

Hypothesis 1B was refuted with respect to the traditional predictor variables T, TS, TR, and TSR, and supported with respect to the qualitative predictor variables CGQ and PR.

TABLE XIV
CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
FREQUENCIES FOR THE CRITERION VARIABLE CG
FOR THE POPULATION GIRLS
(N = 100)
1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
5	5	90	90	5	5		
OBSERVED							
Predictor Variables	Below N	Below %	At N	At %	Above N	Above %	Chi Square
T	64	64.0	34	34.0	2	2.0	732.8*
TS	59	59.0	38	39.0	3	3.0	614.0*
TR	61	61.0	37	37.0	2	2.0	607.2*
TSR	60	60.0	38	38.0	2	2.0	584.8*
CGQ	1	1.0	99	99.0	0	0.0	9.1*
PR	6	6.0	90	90.0	4	4.0	0.4

Note. * Indicates significance demonstrated at .01 level p <.01 = 9.210

It should be noted that the four traditional predictor variables T, TS, TR, TSR when compared to the CG research-based criterion variable, placed from 34% to 38% of

the population of girls at the same instructional level, from 59% to 64% below, and 2% to 3% above the instructional level designated by the researcher. Only a little over a third of the time were the traditional predictor variables "accurate" in the sense that they placed students at the level indicated by the criterion variable. These findings do not fit either the "practicality" or the accuracy required for a scoring system.

On the other hand, the CGQ and the PR predictor variables placed about 99% and 90% of the population of the girls, respectively, at the same level of instruction as the research-based CG criterion (90%), and above the amount required for an accurate and practical scoring system.

Furthermore, the CGQ placed 1% below and 0% above the level recommended by the CG research-based criterion. The PR presents slightly different results with 6% placed below and 4% placed above the instructional level recommended by the CG criterion. These findings fit the "practicality" requirement for a scoring system.

Hypothesis 1C

CG Criterion--Population Boys. There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of

agreements between the predictor estimates and the CG estimates. Frequencies, percents, and Chi square values for the predictor variables with the CG criterion variable, population boys are presented in Table XV.

TABLE XV
CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
FREQUENCIES FOR THE CRITERION VARIABLE CG
FOR THE POPULATION BOYS
(N = 100)
1990

Predictor Variables	EXPECTED						Chi Square	
	Below		At		Above			
	N	%	N	%	N	%		
	5	5	90	90	5	5		
OBSERVED								
Predictor Variables	N	%	N	At	%	Above	%	Chi Square
T	65	65.0	34	34.0	1	1.0		758.0*
TS	65	65.0	34	34.0	1	1.0		758.0*
TR	64	64.0	35	35.0	1	1.0		678.0*
TSR	59	59.0	39	39.0	2	2.0		562.9*
CGQ	1	1.0	98	98.0	1	1.0		7.1
PR	8	8.0	91	91.0	1	1.0		5.0

Note. * Indicates significance demonstrated at .01 level $p < .01 = 9.210$

For the population of boys, significant differences at the $p < .01$ level were found between each of the four

traditional predictor variables T, TS, TR, TSR and the CG criterion variable in estimating the students' instructional reading levels. For the population of boys, nonsignificant differences were found between the two qualitative predictor variables CGQ and PR and the criterion variable CG in estimating the students' instructional reading levels. Therefore, Hypothesis 1C was refuted with respect to the traditional predictor variables T, TS, TR, TSR and supported with respect to the predictor variables CGQ and PR.

As with Tables XIII and XIV, the traditional predictor variables T, TS, TR, TSR continue to place a high percentage of students below the expected instructional levels, that is about 59% to 65% as compared to the expected estimates of the research-based criterion CG (5%). From 34% to 40% of the population of boys were placed at instructional level, and 1 to 2% above instructional level, both results below the level necessary to indicate an accurate and "practical" scoring system.

On the other hand, the CGQ and the PR predictor variables appear to be practical and accurate in placing the population of boys at instructional levels compared to the CG criterion variable, as indicated by the 98% and 91% results observed in Table XV. Furthermore, the CGQ estimated a student placement of 1% below and above, while the PR estimated a student placement of 6% below and

4% above the instructional level estimated by the CG criterion variable. Discussion of this close agreement will be reserved for Chapter V.

Hypothesis 1D

CG Criterion--Population 2nd Grade. For the second grade, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the CG estimates. Frequencies, percents and Chi square values for the predictor variables with the CG criterion variable, 2nd grade population, are presented in Table XVI.

For the 2nd grade population, significant differences at the $p < .01$ level were found between the four traditional predictor variables T, TS, TR, and TSR and the criterion variable CG in estimating the students' instructional reading levels. For the 2nd grade population, non-significant differences were found between the two qualitative predictor variables, CGQ and PR, and the criterion variable CG in estimating the students' instructional reading levels. Therefore, Hypothesis 1D was refuted with respect to the predictor variables T, TS, TR, TSR and supported with respect to the predictor variables CGQ and PR.

TABLE XVI

CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED FREQUENCIES
FOR THE CRITERION VARIABLE CG FOR THE POPULATION 2ND GRADE
(N = 69)
1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
3.45	5	62.1	90	3.45	5		
OBSERVED							
Predictor Variables	Below N	Below %	At N	At %	Above N	Above %	Chi Square
T	35	50.7	33	47.8	1	1.4	303.9*
TS	35	50.7	33	47.8	1	1.4	303.9*
TR	35	50.7	33	47.8	1	1.4	274.8*
TSR	33	47.8	35	50.7	1	1.4	239.6*
CGQ	2	2.9	67	97.1	0	0.0	4.4
PR	4	5.8	62	89.9	3	4.3	0.1

Note. * Indicates significance demonstrated at .01 level p <.01 = 9.210

For the second grade population, the traditional predictor variables were comparable in the amount of students placed at and below the instructional level. That is, from 47% to 50% of the second grade population were placed below and at the expected levels by the traditional predictor variables.

Furthermore, the trend established by the three previous tables of the traditional predictor variables placing a relatively small percentage above the instructional level continues. That is, 1.4% were placed above the instructional level. This finding will be further discussed in Chapter V.

Both the CGQ and PR qualitative predictor variables continue to agree with the CG research-based criterion variable in placing students at the instructional level. That is, the CGQ predictor variable placed 2nd grade students at instructional level 97.1% of the time, while the PR predictor variable placed 2nd grade students at instructional level 89.9%, as compared to the CG research-based criterion variable. Furthermore, the CGQ predictor placed 2.9% below and 0% above instructional level, and the PR placed 5.8% below and 4.3% above the instructional level indicated by the CG criterion. The implications of these comparisons with the CG criterion will be discussed in Chapter V.

Hypothesis 1E

CG Criterion--Population 5th Grade. For the 5th grade population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the

observed frequencies are the number of agreements between the predictor estimates and the CG estimates. Frequencies, percents, and Chi square values for the predictor variables with the CG criterion variable, 5th grade population, are presented in Table XVII.

TABLE XVII

CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED FREQUENCIES FOR THE CRITERION VARIABLE CG FOR THE POPULATION 5TH GRADE (N =34) 1990

Predictor Variables	EXPECTED						Chi Square
	Below		At		Above		
	N	%	N	%	N	%	
	1.7	5	30.6	90	1.7	5	
Predictor Variables	OBSERVED						Chi Square
	N	%	N	%	N	%	
T	26	76.5	7	20.6	1	2.9	365.8*
TS	24	70.6	9	26.5	1	2.9	308.1*
TR	25	73.5	8	23.5	1	2.9	313.7*
TSR	24	70.6	9	26.5	1	2.9	286.5*
CGQ	0	0.0	34	100.0	0	0.0	3.8*
PR	2	5.9	32	94.1	0	0.0	1.8

Note. * Indicates significance demonstrated at .01 level $p < .01 = 9.210$

For the 5th grade population, significant differences at the $p < .01$ level were found between the four traditional predictor variables T, TS, TR, TSR and the CG criterion variable in estimating the students' instructional reading levels. For the 5th grade population, a nonsignificant difference was found between both the CGQ and PR, and the research based criterion variable CG, in estimating the students' instructional reading levels. Therefore, Hypothesis 1E was refuted with respect to the traditional predictor variables T, TS, TR, TSR, and supported with respect to the qualitative predictor variables, CGQ and PR.

It should be noted that the agreement at the instructional levels between the traditional predictor variables T, TS, TR, TSR and the criterion was from 20% to 26%. That is, only a fourth of the time were the predictor variables "accurate" in the sense that they estimated "as well as" the CG system, the research-based criterion. Furthermore, 2.9% were placed above, and over two thirds, that is 70% to 76%, of the 5th grade population were placed below the expected level by the four predictor variables.

Also shown in Table XVI is the fact that for the 5th grade population, nonsignificant differences were found between both the CGQ and PR qualitative predictor variables in estimating the students' instructional reading levels.

It should also be noted that the agreement between both the PR and CGQ predictors and the CG criterion on the

instructional levels was above the level estimated by the research-based criterion. The CGQ placed 100% at the same instructional level as the CG criterion, and the PR placed 94.1% at the instructional level. The PR placed 5.9% below and 0% above the "correct" level, thus achieving an error total well below the specified 10% for the required accuracy of a scoring system. These findings fit the "practicality" requirement for a scoring system. Further discussion of these findings will occur in Chapter V.

Hypothesis 1F

CG Criterion--Population 3rd Grade. For the third grade population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the observed frequencies are the number of agreements between the predictor estimates and the CG estimates. Frequencies, percents, and Chi square values for the predictor variables with the CG criterion variable, third grade population are presented in Table XVIII.

For the third grade population, significant differences were found at the $p < .01$ level between the four traditional predictor variables T, TS, TR, TSR and the CG criterion variable in estimating the students' instructional reading levels. Nonsignificant differences were found between the

two qualitative predictor variables, the CGQ and the PR, and the criterion variable CG. Therefore, Hypothesis 1F was refuted with respect to the traditional predictor variables T, TS, TR, TSR and supported with respect to the qualitative predictor variables CGQ and PR.

TABLE XVIII

CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED FREQUENCIES FOR THE CRITERION VARIABLE CG FOR THE POPULATION 3RD GRADE (N = 53) 1990

Predictor Variables	EXPECTED						Chi Square
	Below		At		Above		
	N	%	N	%	N	%	
	2.65	5	47.7	90	2.65	5	
OBSERVED							
Predictor Variables	N	%	N	At %	N	Above %	Chi Square
T	38	71.7	15	28.3	0	0.0	496.6*
TS	36	67.9	15	28.3	0	0.0	442.3*
TR	36	67.9	17	32.1	0	0.0	411.4*
TSR	34	64.2	19	35.8	0	0.0	362.1*
CGQ	0	0.0	53	100.0	0	0.0	5.9
PR	5	9.4	47	88.7	1	1.9	3.1

Note: * Indicates significance demonstrated at .01 level p <.01 = 9.210

Continuing the trend established by the previous data tables, less than a third of the third grade population were placed at the instructional level by the traditional predictor variables as compared to the research-based CG criterion variable. That is, from 28% to 36% of the third grade population were placed at instructional level, while from 64% to 71% of the third grade population were placed below the instructional level. No students were placed above the instructional level by the traditional predictor variables.

Shown in Table XVIII is the fact that the CGQ predictor variable placed 100% of the 3rd grade population at the instructional level as compared to the CG research based criterion variable. These results will be discussed in Chapter V.

Also shown in Table XVIII is the 88.7% agreement at the instructional level between the PR predictor and the CG criterion variable. The PR placed 9.4% below and 1.9% above the expected criterion level. This 11.3% is slightly above the error total allowed or an accurate and practical scoring system. These findings will be further discussed in Chapter V.

Hypothesis 1G

CG Criterion--Population 4th Grade. For the 4th Grade population, there will be no significant differences between observed and expected frequencies of students placed below,

at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the observed frequencies are the number of agreements between the predictor estimates and the CG estimates. Frequencies, percents, and Chi square values for the predictor variables with the CG criterion variables for the 4th grade population are presented in Table XIX.

For the population 4th grade, significant differences at the $p < .01$ level were found between each of the four traditional predictor variables T, TS, TR, TSR and the CG criterion variable in estimating the students' instructional reading levels. Nonsignificant differences were found between the two qualitative predictor variables CGQ and PR, and the research-based criterion variable CG in estimating the 4th grade students' instructional reading levels. Therefore, Hypothesis 1G was refuted with respect to the predictor variables T, TS, TR, TSR and supported with respect to the predictor variables CGQ and PR.

The agreement between the traditional predictor variables and the CG criterion variable was consistent with prior findings. That is, the agreement at the instructional level between the traditional predictor variables and the CG criterion variable occurred a little more than a third of the time or from 30% to 34% of the time. Additionally, 2% to 4.5% were placed above the instructional level, and 64%

to 68% were placed below the instructional level as compared to the CG criterion variable. Overall, the traditional predictor variables appear to be far less "accurate" than the research-based criterion in estimating the 4th grade students' instructional reading levels.

TABLE XIX

CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED FREQUENCIES FOR THE CRITERION VARIABLE CG FOR THE POPULATION 4TH GRADE (N = 44) YEAR

Predictor Variables	EXPECTED						Chi Square
	Below		At		Above		
	N	%	N	%	N	%	
	2.2	5	39.6	90	2.2	5	
OBSERVED							
Predictor Variables	N	%	N	At %	N	Above %	Chi Square
T	30	68.2	13	29.5	1	2.3	369.0*
TS	29	65.9	15	34.1	0	0.0	344.0*
TR	29	65.9	14	31.8	1	2.3	318.1*
TSR	28	63.6	14	31.8	2	4.5	293.5*
CGQ	0	0.0	43	97.7	1	2.3	3.1
PR	3	6.8	40	90.9	1	2.3	0.9

Note. * Indicates significance demonstrated at .01 level p < .01 = 9.210

On the other hand, the CGQ qualitative predictor variable placed 97.7% at the instructional level; 2.3% above the instructional level and 0% below the instructional level, well within the estimates established by the CG research-based criterion CG.

Furthermore, the PR qualitative predictor variable placed 90.9% at the instructional level, 6.8% below and 2.3% above the instructional level, within the range of the "practicality" requirement for the PR scoring system. These final results, as well as a general discussion of the implications of all the results related to Hypotheses 1A through 1G, will be included in Chapter V.

Hypothesis 2A

AE Criterion--Total Population. For the total population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the observed frequencies are the number of agreements between the predictor estimates and the AE (administrator's estimates). Frequencies, percents and Chi square values for the predictor variables with the AE criterion variable, total population, are presented in Table XX.

TABLE XX

CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
FREQUENCIES FOR THE CRITERION VARIABLE AE
FOR THE TOTAL POPULATION
(N = 200)
1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
10	5	180	90	10	5		
OBSERVED							
Predictor Variables	Below N	% %	At N	% %	Above N	% %	Chi Square
T	127	63.5	70	35.0	3	1.5	1441.0*
TS	122	61.0	75	37.5	3	1.5	1320.6*
TR	125	62.5	72	36.0	3	1.5	1284.2*
TSR	118	59.0	80	40.0	2	1.0	1128.4*
CGQ	11	5.5	182	91.0	7	3.5	1.0
PR	7	3.5	192	96.0	1	0.5	9.8*

Note. * Indicates significance demonstrated at .01 level p < .01 = 9.210

For the total population, significant differences were found at the $p < .01$ level between five of the six predictor variables T, TS, TR, TSR, PR and the criterion variable AE, in estimating the students' instructional reading levels. For the same total population, nonsignificant differences were found between the predictor variable CGQ and the

criterion variable AE in estimating the students' instructional reading levels. Therefore, Hypothesis 2A was refuted with respect to the predictor variables T, TS, TR, TSR, and PR, and supported with respect to the predictor variable CGQ.

Table XX shows the results for the total population of 200 children from grades two through five. Statistically significant differences were found between the instructional placements estimated by the four quantitative scoring systems and the AE criterion system. With the quantitative systems the instructional estimates coincided with those of the AE criterion only 35% to 40% of the time. From 59% to 63.5% of the time the quantitative systems estimated below the AE criterion estimate, and from 1.0% to 1.5% of the time they estimated above.

There was no statistically significant difference between the CGQ qualitative placements and the AE criterion placements. The CGQ system placed students the same as the AE system 91% of the time, with 5.5% below and 3.5% above. Since no more than 10% of the students were "misplaced," the CGQ system met the "practicality" requirement of the study.

Although there was a statistically significant difference between the PR qualitative placements and the AE criterion, it was very small and caused by the success of the PR system in predicting the criterion estimates of instructional level 96% of the time. The observed

frequencies were 7 below, 192 the same, and 1 above while the expected were 10, 180, and 10, respectively. The observed yielded 24 cases different from the expected and thus the small but significant difference from a statistical standpoint. The PR system met the practicality requirement of no more than 10% error.

Hypothesis 2B

AE Criterion--Population Girls. For the 2nd grade population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading levels, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the AE estimates. Frequencies, percents, and Chi square values for the predictor variables with the AE criterion variable, population girls are presented in Table XXI.

For the population of girls, significant differences were found at the $p < .01$ level between each of the four traditional predictor variables T, TS, TR, TSR and the AE criterion variable in estimating the students' instructional reading levels. Nonsignificant differences were found between the two qualitative predictor variables CGQ and PR, and the criterion variable AE in estimating the students' instructional reading levels. Therefore, Hypothesis 2B was

refuted with respect to the traditional predictor variables T, TS, TR, and TSR, and supported with respect to the two qualitative predictor variables CGQ and PR.

TABLE XXI
CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
FREQUENCIES FOR THE CRITERION VARIABLE AE
FOR THE POPULATION GIRLS
(N = 100)
1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
5	5	90	90	5	5		
OBSERVED							
Predictor Variables	Below N	Below %	At N	At %	Above N	Above %	Chi Square
T	62	62.0	38	38.0	0	0.0	684.8*
TS	57	57.0	42	42.0	1	1.0	569.6*
TR	60	60.0	40	40.0	0	0.0	587.8*
TSR	56	56.0	45	45.0	0	0.0	502.7*
CGQ	4	4.0	92	92.0	4	4.0	0.4
PR	5	5.0	95	95.0	0	0.0	5.3

Note. * Indicates significance demonstrated at .01 level $p < .01 = 9.210$

It should be noted that the four traditional predictor variables T, TS, TR, TSR when compared to the AE criterion variable, placed about 38% to 45% of the population of girls

at the same instructional level, about 57% to 62% below, and less than 1% of the students above the instructional level. Only a little over a third of the time were the traditional predictor variables "accurate" in the sense that they placed students at the level recommended by the criterion variable AE. These findings do not fit either the "practicality" or the accuracy required for a scoring system.

On the other hand, the CGQ and the PR predictor variables placed 92% and 95% of the population of girls, respectively, at the same level of instruction as the AE criterion variable, above the 90% amount required to fit the accuracy and "practicality" requirement for a scoring system. Furthermore, the CGQ placed 4% both above and below the recommended instructional level, while the PR estimated 5% below and 0% above. Implications of these results will be discussed in Chapter V.

Hypothesis 2C

AE Criterion--Population Boys. There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the AE estimates. Frequencies, percents, and Chi square values for

the predictor variables with the AE criterion variable, population boys are presented in Table XXII.

TABLE XXII
CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
FREQUENCIES FOR THE CRITERION VARIABLE AE
FOR THE POPULATION BOYS
(N = 100)
1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
5	5	90	90	5	5		
OBSERVED							
Predictor Variables	N	Below %	N	At %	N	Above %	Chi Square
T	65	65.0	32	32.0	3	3.0	758.2*
TS	65	65.0	33	32.0	2	2.0	757.9*
TR	65	65.0	32	32.0	3	3.0	700.2*
TSR	62	62.0	35	35.0	2	2.0	630.2*
CGQ	7	7.0	90	90.0	3	3.0	1.6
PR	2	2.0	97	97.0	1	1.0	5.5

Note. * Indicates significance demonstrated at .01 level $p < .01 = 9.210$

For the population of boys, significant differences at the $p < .01$ level were found between each of the four traditional predictor variables T, TS, TR TSR and the AE criterion variable in estimating the students' instructional

reading levels. For the population of boys, nonsignificant differences were found between the two qualitative predictor variables CGQ and PR and the criterion variable AE in estimating the students' instructional reading levels. Therefore, Hypothesis 2C was refuted with respect to the traditional predictor variables T, TS, TR, TSR and supported with respect to the predictor variables CGQ and PR.

As with Tables XX and XXI, the traditional predictor variables T, TS, TR, TSR continue to place from 62% to 65%, or over two thirds of the students below the expected instructional levels, as compared to the 5% estimates of the criterion AE. Additionally, from 32% to 35% of the population of boys were placed at instructional level, below the recommended amount to indicate an accurate and "practical" scoring system.

On the other hand, the CGQ and PR predictor variables appear to be practical and accurate in placing the population of boys at instructional levels compared to the AE criterion variable. The 90% and 97% results at instructional level; the 3% to 7% results below, and the 1% to 2% results above instructional level support the CGQ and PR systems as meeting the "practicality" requirement of the study.

Hypothesis 2D

AE Criterion--Population 2nd Grade. For the second grade, there will be no significant differences between

observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the AE estimates.

Frequencies, percents, and Chi square values for the predictor variables with the AE criterion variable, population 2nd Grade are presented in Table XXIII.

For the second grade population, significant differences at the $p < .01$ level were found between the four traditional predictor variables T, TS, TR, and TSR and the criterion variable AE in estimating the students' instructional reading levels. For the second grade population, nonsignificant differences were found between the two qualitative predictor variables, CGQ and PR, and the criterion variable CG in estimating the students' instructional reading levels. Therefore, Hypothesis 2D was refuted with respect to the predictor variables T, TS, TR, TSR and supported with respect to the predictor variables CGQ and PR.

For the second grade population, the traditional predictor variables were slightly different in the percentage of students placed at and below the instructional level as compared to the AE criterion variable. That is, 46% to 49% of the second grade population were placed below

the expected levels by the traditional predictor variables, and 47% to 52% were placed at the instructional level by the same traditional predictor variables.

TABLE XXIII

CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
FREQUENCIES FOR THE CRITERION VARIABLE AE
FOR THE POPULATION 2ND GRADE
(N = 69)
1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
3.45	5	62.1	90	3.45	5		
OBSERVED							
Predictor Variables	Below		At		Above		Chi Square
	N	%	N	%	N	%	
T	34	49.3	33	47.8	2	2.9	284.8*
TS	32	46.4	35	50.7	2	2.9	248.7*
TR	34	49.3	33	47.8	2	2.9	255.7*
TSR	32	46.4	36	52.2	1	1.4	222.9*
CGQ	5	7.2	61	88.4	3	4.3	0.8
PR	1	1.4	68	98.6	0	0.0	5.8

Note. * Indicates significance demonstrated at .01 level p < .01 = 9.210

Furthermore, the trend established by the previous tables, of the traditional predictor variables placing a relatively small percentage above the instructional level

continues. That is, 1.4% to 2.9% were placed above the instructional level. These findings do not fit the "practicality" requirement for a scoring system.

On the other hand, both the CGQ and the PR qualitative predictor variables agreed with the AE criterion variable in placing students at the instructional level. That is, the CGQ predictor variable placed 2nd grade students at instructional level 88.4% of the time, (1.6% below the required practicality level) while the PR predictor variable placed 2nd grade students at instructional level 98.6% (8.6% above the practicality requirement) as compared to the AE criterion variable. Furthermore, the 7.2% below, and the 4.3% above instructional level achieved by the CGQ, as well as the 1.4% below and the 0% above achieved by the PR also fit the placement requirements needed to agree with the AE criterion.

Hypothesis 2E

AE Criterion--Population 5th Grade. For the 5th grade population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the observed frequencies are the number of agreements between the predictor estimates and the AE estimates. Frequencies, percents, and Chi square values for the predictor variables with the AE criterion,

variable for the 5th grade population are presented in Table XXIV.

TABLE XXIV
CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
FREQUENCIES FOR THE CRITERION VARIABLE AE
FOR THE POPULATION 5TH GRADE
(N =34)
1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
1.7	5	30.6	90	1.7	5		
OBSERVED							
Predictor Variables	Below N	Below %	At N	At %	Above N	Above %	Chi Square
T	25	73.5	8	23.5	1	2.9	336.3*
TS	24	70.6	9	26.5	1	2.9	308.1*
TR	25	73.5	8	23.5	1	2.9	313.7*
TSR	23	67.6	10	29.4	1	2.9	260.4*
CGQ	1	2.9	33	97.1	0	0.0	2.2
PR	2	5.9	32	94.1	0	0.0	1.8

Note. Indicates significance demonstrated at .01 level p <.01 = 9.210

For the 5th grade population, significant differences were found between the four traditional predictor variables T, TS, TR, TSR and the CG criterion variable in estimating the students' instructional reading levels. A

nonsignificant difference was found between CGQ, PR, and the criterion variable AE, in estimating the students' instructional reading levels. Therefore, Hypothesis 2E was refuted with respect to the traditional predictor variables T, TS, TR, TSR, and supported with respect to the qualitative predictor variables CGQ and PR.

It should be noted that the agreement at the instructional levels between the traditional predictor variables T, TS, TR, TSR and the AE criterion was from 23% to 29%. That is, less than a fourth of the time were the predictor variables "accurate" in the sense that they estimated "as well as" the AE criterion. Furthermore, 64% to 73% of the students were placed below the instructional level by the traditional predictor variables and 2.9% were placed above the level indicated by the AE criterion variable. This placement does not fit the accuracy or "practicality" requirement of this study for a scoring system.

On the other hand, the CGQ and PR predictor variables placed 97% and 94% of the 5th grade population, respectively, at the same level of instruction as the AE criterion, above the level expected by the AE criterion variable. The CGQ and PR placed 2.9% and 5.9% below, and 0% above the "correct" level, thus achieving an error total below the specified 10% for the required accuracy of a

scoring system. This finding also fits the "practicality" requirement for a scoring system.

Hypothesis 2F

AE Criterion--Population 3rd Grade. For the third grade population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the observed frequencies are the number of agreements between the predictor estimates and the AE estimates.

Frequencies, percents, and Chi square values for the predictor variables with the AE criterion variable for the 3rd grade population are presented in Table XXV.

For the third grade population, significant differences were found at the $p < .01$ level between the four traditional predictor variables T, TS, TR, TSR and the AE criterion variable in estimating the students' instructional reading levels. Nonsignificant differences were found between the two qualitative predictor variables, the CGQ and the PR, and the criterion variable AE. Therefore, Hypothesis 2F was refuted with respect to the traditional predictor variables T, TS, TR, TSR and supported with respect to the qualitative predictor variables CGQ and PR.

Continuing the trend established by the previous data tables, less than a third of the third grade population were

placed at the instructional level by the traditional predictor variables as compared to the AE criterion variable. That is, 28% to 32% of the third grade population were placed at instructional level, while 67% to 71% of the

TABLE XXV

CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
FREQUENCIES FOR THE CRITERION VARIABLE AE
FOR THE POPULATION 3RD GRADE
(N = 53)
1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
2.65	5	47.7	90	2.65	5		
OBSERVED							
Predictor Variables	Below N	Below %	At N	At %	Above N	Above %	Chi Square
T	38	71.7	15	28.3	0	0.0	496.6*
TS	37	69.8	16	30.2	0	0.0	469.0*
TR	37	69.8	16	30.2	0	0.0	437.3*
TSR	36	67.9	17	32.1	0	0.0	411.4*
CGQ	4	7.5	46	86.8	3	5.7	0.4
PR	3	5.7	49	92.5	1	1.9	1.1

Note. * Indicates significance demonstrated at .01 level $p < .01 = 9.210$

third grade population were placed below the instructional level. No students were placed above the instructional level by the traditional predictor variables.

Shown in Table XXV is the fact that the CGQ predictor variable placed 86.8% of the third grade population at the instructional level as compared to the AE criterion variable. The CGQ predictor variable also placed 7.5% below and 5.7% above the level expected by the AE criterion.

Finally, Table XXV indicates the 92.5% agreement at the instructional level between the PR predictor and the AE criterion variable. The PR placed 5.7% below and 1.9% above the expected AE criterion level, well within the error total allowed for an accurate and "practical" scoring system. Less than 10% of the students were "misplaced", therefore, the PR met the "practicality" requirement of this study.

Hypothesis 2G

AE Criterion--Population 4th Grade. For the 4th Grade population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the observed frequencies are the number of agreements between the predictor estimates and the AE estimates.

Frequencies, percents, and Chi square values for the predictor variables with the AE criterion variable for the 4th grade population are presented in Table XXVI.

TABLE XXVI
CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
FREQUENCIES FOR THE CRITERION VARIABLE AE
FOR THE POPULATION 4TH GRADE
(N = 44)
1990

Predictor Variables	EXPECTED						Chi Square
	Below		At		Above		
	N	%	N	%	N	%	
	2.2	5	39.6	90	2.2	5	
Predictor Variables	OBSERVED						Chi Square
	N	%	N	%	N	%	
T	30	68.2	14	31.8	0	0.0	370.0*
TS	29	65.9	15	34.1	0	0.0	344.0*
TR	29	65.9	15	34.1	0	0.0	319.4*
TSR	27	61.4	17	38.6	0	0.0	272.1*
CGQ	1	2.3	42	95.5	1	2.3	1.5
PR	1	2.3	43	97.7	0	0.0	3.1

Note. * Indicates significance demonstrated at .01 level $p < .01 = 9.210$

For the population 4th grade, significant differences at the $p < .01$ level were found between each of the four traditional predictor variables T, TS, TR, TSR and the AE

criterion variable in estimating the students' instructional reading levels. Nonsignificant differences were found between the two qualitative predictor variables CGQ and PR, and the criterion AE in estimating the 4th grade students' instructional reading levels. Therefore, Hypothesis 2G was refuted with respect to the predictor variables T, TS, TR, TSR and supported with respect to the predictor variables CGQ and PR.

The agreement between the traditional predictor variables and the AE criterion variable was consistent with prior findings. That is, the agreement at the instructional level between the traditional predictor variables and the AE criterion variable occurred from 31% to 38%, a little more than a third of the time. Furthermore, no students were placed above instructional level, and 61% to 68% were placed below the instructional level, as compared to the AE criterion variable. Overall, the traditional predictor variables appear to be less accurate than the AE criterion variable in estimating the 4th grade students' instructional reading levels.

On the other hand, the CGQ and the PR qualitative predictor variables appear to be practical and accurate in placing the 4th grade population at instructional levels compared to the AE criterion variable, as indicated by the results observed in Table XXVI. The CGQ placed 95.5% at the instructional level, 2.3% below, and 2.3% above the

instructional level, within the range of the "practicality" requirement for a scoring system. The PR placed 97.7% at the instructional level, 2.3% below and 0% above, also within the range of an accurate and practical scoring system as compared to the AE criterion variable. These final results, as well as a general discussion of the implications of all the results related to Hypotheses 2A through 2G will be included in Chapter V.

Hypothesis 3A

TE Criterion--Total Population. For the total population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the observed frequencies are the number of agreements between the predictor estimates and the TE (Teacher) estimates. Frequencies, percents, and Chi square values for the predictor variables with the TE criterion variable, total population are presented in Table XXVII.

As shown in Table XXVII for the total population, the six predictor variables were significantly different at the $p < .01$ level from the TE criterion variable in estimating the students' instructional reading levels. Therefore, Hypothesis 3A was refuted.

TABLE XXVII

CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
 FREQUENCIES FOR THE CRITERION VARIABLE TE
 FOR THE TOTAL POPULATION
 (N = 200)
 1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
10	5	180	90	10	5		
OBSERVED							
Predictor Variables	Below N	Below %	At N	At %	Above N	Above %	Chi Square
T	117	58.5	60	30.0	23	11.5	1241.8*
TS	115	57.5	63	31.5	22	11.0	1193.0*
TR	112	56.0	67	33.5	21	10.5	1123.4*
TSR	99	49.5	71	35.5	30	15.0	898.1*
CGQ	28	14.0	88	44.0	84	42.0	627.0*
PR	26	13.0	84	42.0	90	45.0	716.8*

Note. * Indicates significance demonstrated at .01 level p < .01 = 9.210

Table XXVI shows the results for the total population of 200 children from grades two through five. Statistically significant differences were found between the instructional placements estimated by the four quantitative scoring systems and the TE criterion system. With the quantitative systems the instructional estimates coincided with those of

the TE criterion only 30% to 36% of the time. From 50% to 59% of the time, the quantitative systems estimated below the TE criterion estimate, and 11% to 15% of the time they estimated above. Statistically significant differences were also found between the qualitative systems and the TE criterion system. With the CGQ and PR systems the instructional estimates coincided with those of the TE criterion only 42% to 44%, with 26% to 28% below and 42% and 45% above.

Hypothesis 3B

TE Criterion--Population Girls. For the population girls, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the TE estimates.

Frequencies, percents, and Chi square values for the predictor variables with the TE criterion variable for the population girls are presented in Table XXVIII.

As shown in Table XXVIII for the population of girls, the six predictor variables; four traditional and quantitative, T, TS, TR, TSR, and two qualitative; CGQ and PR, were significantly different at the $p < .01$ level from the TE criterion variable in estimating the students'

instructional reading levels. Therefore, Hypothesis 3B was refuted.

TABLE XXVIII

CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
FREQUENCIES FOR THE CRITERION VARIABLE TE
FOR THE POPULATION GIRLS
(N = 100)
1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
5	5	90	90	5	5		
OBSERVED							
Predictor Variables	N	Below %	N	At %	N	Above %	Chi Square
T	52	52.0	33	33.0	15	15.0	497.9*
TS	51	51.0	33	33.0	15	15.0	479.3*
TR	51	51.0	35	35.0	15	15.0	476.8*
TSR	40	40.0	39	39.0	21	21.0	325.1*
CGQ	15	15.0	27	27.0	58	58.0	625.9*
PR	13	13.0	33	33.0	54	54.0	529.1*

Note. * Indicates significance demonstrated at .01 level p < .01 = 9.210

The agreement between the four traditional predictor variables and the TE criterion variable was from 33% to 39%, while the agreement between the two qualitative predictor

variables and the TE criterion variable tended to be slightly lower, from 27% to 33%.

Furthermore, for the population of girls, the traditional predictor variables continued to place more students below the TE criterion variable's instructional level estimates, from 40% to 52%, and fewer students above the TE criterion variable's instructional level estimates, that is, 15% to 21%.

The two qualitative predictor variables obtained slightly different results in the amount of students placed at the instructional levels, 27% and 33%, as compared to the TE criterion variable. Both the CGQ and the PR placed a higher percentage of students above instructional level, 58% and 54%, and a lower percentage of students below instructional level, 15% and 13%, as compared to the TE criterion variable. Overall, there was very little agreement between the six predictor variables and the criterion variable, TE.

Hypothesis 3C

TE Criterion--Population Boys. There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of

agreements between the predictor estimates and the TE estimates.

Frequencies, percents, and Chi square values for the predictor variables with the TE criterion variable for the population of boys are presented in Table XXIX.

TABLE XXIX
CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
FREQUENCIES FOR THE CRITERION VARIABLE TE
FOR THE POPULATION BOYS
(N = 100)
1990

Predictor Variables	EXPECTED						Chi Square
	Below		At		Above		
	N	%	N	%	N	%	
	5	5	90	90	5	5	
OBSERVED							
Predictor Variables	N	%	N	%	N	%	Chi Square
T	65	65.0	27	27.0	8	8.0	765.9*
TS	64	64.0	30	30.0	7	7.0	737.0*
TR	61	61.0	32	32.0	6	6.0	664.8*
TSR	59	59.0	32	32.0	9	9.0	623.8*
CGQ	13	13.0	31	31.0	45	45.0	371.5*
PR	13	13.0	40	40.0	47	47.0	393.4*

Note. * Indicates significance demonstrated at .01 level p < .01 = 9.210

As shown in Table XXIX for the population of boys the six predictor variables, four traditional; T, TS, TR, TSR, and two qualitative; CGQ and PR, were significantly different at the $p < .01$ level from the TE criterion variable in estimating the students' instructional reading levels. Therefore Hypothesis 3C was refuted.

It should be noted that the range of agreement between the six predictor variables and the TE criterion was between 27% through 32% at the instructional level. Furthermore, the traditional predictor variables placed 59% through 65% of students below instructional level, and 6% through 9% above instructional level, as compared to the TE criterion variable. This lack of agreement will be further discussed in Chapter V.

On the other hand, the two qualitative predictor variables, the CGQ and PR, tended to place more students above the TE criterion variable's instructional level estimates (45% to 47%), and less students below the TE criterion variable's instructional level estimates (13%). The CGQ placed 31% at instructional level and the PR placed 40% at instructional level as compared to the TE criterion. Overall, there was very little agreement between the six predictor variables' estimates and the TE criterion variable estimates below, at, or above instructional levels.

Hypothesis 3D

TE Criterion--Population 2nd Grade. For the second grade, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the TE estimates.

Frequencies, percents, and Chi square values for the predictor variables with the TE criterion variable for the 2nd Grade are presented in Table XXX.

For the 2nd grade population, significant differences at the $p < .01$ level were found between the six predictor variables T, TS, TR, TSR, CGQ, PR and the criterion variable TE in estimating the students' instructional reading levels. Therefore, Hypothesis 3D was refuted with respect to the six predictor variables.

For the 2nd grade population, the four traditional predictor variables were comparable to other population groups in the percentage of students placed at, below and above instructional level as compared to the TE criterion variable. From 46% to 58% , that is a little more than 50% of the students were placed below the instructional level by the four traditional variables as compared to the TE criterion variable. Furthermore, a little more than a third

of the students were placed at instructional level by the four traditional predictor variables. That is from 29% to 39% were placed at instructional level, and 13% to 14.5% were placed above instructional level by the four traditional predictor variables, as compared to the TE criterion.

TABLE XXX

CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
FREQUENCIES FOR THE CRITERION VARIABLE TE
FOR THE POPULATION 2ND GRADE
(N = 69)
1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
3.45	5	62.1	90	3.45	5		
OBSERVED							
Predictor Variables	Below N	Below %	At N	At %	Above N	Above %	Chi Square
T	40	58.0	20	29.0	9	13.0	424.7*
TS	37	53.6	23	33.3	9	13.0	359.8*
TR	37	53.6	23	33.3	9	13.0	359.8*
TSR	32	46.4	27	39.1	10	14.5	268.5*
CGQ	13	18.8	40	58.0	16	23.2	80.0*
PR	12	17.4	40	58.0	17	24.6	82.3*

Note. * Indicates significance demonstrated at .01 level $p < .01 = 9.210$

The two qualitative predictor variables, the CGQ and PR, have comparable results at, below and above instructional levels when compared to the TE criterion variable. That is, exactly 58% were placed at instructional level; 17% to 18% were placed below instructional level; and 23% to 24% were placed above instructional level by both the CGQ and the PR predictor variables when compared to the TE criterion variable. The significance of these findings will be discussed in Chapter V.

Hypothesis 3E

TE Criterion--Population 5th Grade. For the 5th grade population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the observed frequencies are the number of agreements between the predictor estimates and the TE estimates.

Frequencies, percents, and Chi square values for the predictor variables with the TE criterion variable for the population girls are presented in Table XXXI.

As demonstrated in Table XXXI for the 5th grade population, significant differences were found between the six predictor variables and the TE criterion variable in estimating the students' instructional reading levels.

Therefore, Hypothesis 3E was refuted with respect to the six predictor variables.

TABLE XXXI
 CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
 FREQUENCIES FOR THE CRITERION VARIABLE TE
 FOR THE POPULATION 5TH GRADE
 (N = 34)
 1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
1.7	5	30.6	90	1.7	5		
OBSERVED							
Predictor Variables	Below		At		Above		Chi Square
	N	%	N	%	N	%	
T	18	52.9	12	35.3	4	11.8	170.7*
TS	17	50.0	11	32.4	5	14.7	156.7*
TR	17	50.0	12	35.3	5	14.7	155.4*
TSR	15	44.1	10	29.4	9	26.5	149.3*
CGQ	3	8.8	14	41.2	17	50.0	147.7*
PR	3	8.8	12	35.3	19	55.9	188.4*

Note. * Indicates significance demonstrated at .01 level $p < .01 = 9.210$

For the 5th grade population, the agreement at the instructional level between the traditional predictor variables T, TS, TR, TSR and the criterion variable TE was from 29% to 35%. From 44% to 53% of the students were

estimated below instructional level, and from 12% to 27% were estimated to be above instructional level by the traditional predictor variables, as compared to the TE criterion variable.

Table XXXI also reports a significant difference between both the CGQ and the PR qualitative predictor variables and the TE criterion in estimating the 5th grade students' instructional reading levels. The CGQ placed 8.8% below, 41% at, and 50% above the instructional level estimates obtained by the TE criterion variable.

Furthermore, the PR predictor variable placed 8.8% below, 35% at, and 56% above the level estimated by the TE criterion variable. The implications of these results will be discussed in Chapter V.

Hypothesis 3F

TE Criterion--Population 3rd Grade. For the third grade population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the observed frequencies are the number of agreements between the predictor estimates and the TE estimates.

Frequencies, percents, and Chi square values for the predictor variables with the TE criterion variable for the 3rd grade population are presented in Table XXXII.

TABLE XXXII

CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
 FREQUENCIES FOR THE CRITERION VARIABLE TE
 FOR THE POPULATION 3RD GRADE
 (N = 53)
 1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
2.65	5	47.4	90	2.65	5		
OBSERVED							
Predictor Variables	Below N	%	At N	%	Above N	%	Chi Square
T	34	64.2	12	22.6	7	13.2	404.7*
TS	35	66.0	13	24.5	5	9.4	422.2*
TR	31	58.5	17	32.1	5	9.4	294.2*
TSR	29	73.6	19	35.8	5	9.4	489.3*
CGQ	8	15.1	20	37.7	25	47.2	215.4*
PR	7	13.2	19	35.8	27	50.9	248.2*

Note. * Indicates significance demonstrated at .01 level $p < .01 = 9.210$

For the 3rd grade population, significant differences were found at the $p < .01$ level between the six predictor variables and the TE criterion in estimating the students' instructional reading levels. Therefore, Hypothesis 3F was refuted with respect to the six predictor variables.

For the 3rd grade population, the agreement at the instructional level between the traditional predictor variables T, TS, TR, TSR and the TE criterion variable was from 23 to 36%; below instructional level was from 59% to 74%; and above instructional level was 9% to 13%.

Table XXXII, for the 3rd grade population, also reports a significant difference between both the CGQ and the PR qualitative predictor variables and the TE criterion in estimating the 3rd grade students' instructional reading levels. The CGQ placed 15% below, 37.7% at, and 47% above the instructional level estimated by the TE criterion variable.

Furthermore, the PR predictor variable placed 13% below, 35.8% at, and 51% above the levels estimated by the TE criterion variable. These results will be discussed in Chapter V.

Hypothesis 3G

TE Criterion--Population 4th Grade. For the 4th Grade population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the observed frequencies are the number of agreements between the predictor estimates and the TE estimates.

Frequencies, percents, and Chi square values for the predictor variables with the TE criterion variable for the 4th grade population are presented in Table XXXIII.

TABLE XXXIII

CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
FREQUENCIES FOR THE CRITERION VARIABLE TE
FOR THE POPULATION 4TH GRADE (N = 44)
1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
2.2	5	39.6	90	2.2	5		
OBSERVED							
Predictor Variables	Below		At		Above		Chi Square
	N	%	N	%	N	%	
T	25	56.8	16	36.4	3	6.8	250.6*
TS	25	56.8	16	36.4	3	6.8	250.6*
TR	26	59.1	15	34.1	3	6.8	273.0*
TSR	23	52.3	15	34.1	6	13.6	218.5*
CGQ	4	9.1	14	31.8	26	59.1	275.5*
PR	4	9.1	13	29.5	28	63.6	321.9*

Note. * Indicates significance demonstrated at .01 level p < .01 = 9.210

For the population 4th grade, significant differences at the $p < .01$ level were found between each of the six predictor variables, T, TS, TR, TSR, CGQ, PR and the TE

criterion variable in estimating the students' instructional reading levels. Therefore, Hypothesis 3G was refuted with respect to the six predictor variables.

For the 4th grade population, the traditional predictor variables placed 52% to 59% below the instructional level; 34% to 36% at instructional level; and 7% to 14% above the instructional level estimated by the TE criterion variable.

Table XXXII also reports a significant difference between both the CGQ and the PR qualitative predictor variables and the TE criterion in estimating the 4th grade students' instructional reading levels. The CGQ and the PR had comparable results with the CGQ and PR placing 9.1% below instructional level; 30% and 32% at instructional level; and 59% and 64% above the instructional level estimated by the TE criterion variable. The implications of these results and the results of all the tables related to this hypothesis will be discussed in Chapter V.

Hypothesis 4A

SR Criterion--Total Population For the total population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the observed frequencies are the number of agreements between the predictor estimates and the SR (Silent Reading) estimates.

Frequencies, percents, and Chi square values for the predictor variables with the SR criterion variable for the total population are presented in Table XXXIV.

TABLE XXXIV

CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
FREQUENCIES FOR THE CRITERION VARIABLE SR
FOR THE TOTAL POPULATION
(N = 200)
1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
10	5	180	90	10	5		
OBSERVED							
Predictor Variables	N	Below %	N	At %	N	Above %	Chi Square
T	144	72.0	40	20.0	16	8.0	8190.7*
TS	141	70.5	40	20.0	19	9.5	7890.7*
TR	142	71.0	39	19.5	19	9.5	8003.5*
TSR	136	68.0	42	21.0	22	11.0	7386.8*
CGQ	90	45.0	76	38.0	34	17.0	3519.5*
PR	98	49.0	66	33.0	36	18.0	4169.2*

Note. * Indicates significance demonstrated at .01 level p <.01 = 9.210

As indicated in Table XXXIV for the total population, significant differences were found at the $p < .01$ level between the six predictor variables T, TS, TR, TSR, CGQ, PR

and the criterion variable SR, in estimating the students' instructional reading levels. Therefore, Hypothesis 4A was refuted.

Table XXXIV shows the results for the total population of 200 children from grades two through five. Statistically significant differences were found between the instructional placements estimated by the four quantitative scoring systems and the SR criterion system. With the quantitative systems the instructional estimates coincided with those of SR criterion only 19% to 21% of the time. From 68% to 72% of the time the quantitative systems estimated below the SR criterion estimate, and from 8% to 11% of the time they estimated above.

Statistically significant differences were also found between the instructional placements estimated by the two qualitative scoring systems and the SR criterion system. With the qualitative systems the instructional estimates coincided with those of the SR criterion 33% to 38% of the time, with 45% to 49% below and 17% to 18% above. These results will be discussed along with the other population results for this criterion variable in Chapter V.

Hypothesis 4B

SR Criterion--Population Girls. For the population girls, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of

six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the SR estimates.

Frequencies, percents, and Chi square values for the predictor variables with the SR criterion, for the population girls, are presented in Table XXXV.

The results of the Chi square analysis using the criterion variable SR for the population of girls as seen in Table XXXV, yielded significant differences at the $p < .01$ level of significance as compared to the six predictor variables T, TS, TR, TSR, CGQ, and PR. Therefore, Hypothesis 4B was refuted.

It should be noted that for the population of girls, the four traditional predictor variables, T, TS, TR, TSR placed a high frequency of students below the instructional level indicated by the SR criterion variable. That is, two thirds or 65% to 67% of the girls were placed below the instructional level indicated by the SR criterion. Furthermore, the four traditional predictor variables placed from 17% to 18% at and 16% to 18% above the instructional level indicated by the SR criterion variable.

For the population of girls, the two qualitative predictor variables, the CGQ and PR placed 41% and 42% below, 43% and 41% at, and 16% and 17% above the instructional level indicated by the SR criterion. That is,

the CGQ and PR observed similar frequencies of students at, below and above the instructional level, but significantly different frequencies from the levels expected by the SR criterion.

TABLE XXXV
CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
FREQUENCIES FOR THE CRITERION VARIABLE SR
FOR THE POPULATION GIRLS
(N = 100)
1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
5	5	90	90	5	5		
OBSERVED							
Predictor Variables	Below N	Below %	At N	At %	Above N	Above %	Chi Square
T	67	67.0	17	17.0	16	16.0	852.2*
TS	65	65.0	17	17.0	18	18.0	813.0*
TR	66	66.0	17	17.0	17	17.0	832.2*
TSR	65	65.0	18	18.0	17	17.0	806.4*
CGQ	41	41.0	43	43.0	16	16.0	307.9*
PR	42	42.0	41	41.0	17	17.0	329.3*

Note. * Indicates significance demonstrated at .01 level $p < .01 = 9.210$

Hypothesis 4C

SR Criterion--Population Boys. There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the SR estimates.

Frequencies, percents, and Chi square values for the predictor variables with the SR criterion for the population of boys are presented in Table XXXVI.

For the population boys, as displayed in Table XXXVI, significant differences were found at the $p < .01$ level between the six predictor variables T, TS, TR, TSR, CGQ, and PR, and the criterion variable SR, in estimating the students' instructional reading levels. Therefore, Hypothesis 4C was refuted.

For the population of boys, the agreement at the instructional level between the traditional predictor variables T, TS, TR, TSR and the criterion SR was from 22% to 24%; below instructional level was from 71 to 77%; and above instructional level was 0% to 5%.

Table XXXVI also reports a significant difference between both the CGQ and the PR predictor variables and the SR criterion variable. The CGQ and the PR placed 49% and

61% below instructional level; 33% and 26% at instructional level; and 18% and 13% above instructional level as estimated by the SR criterion variable. These results will be discussed in Chapter V.

TABLE XXXVI
 CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
 FREQUENCIES FOR THE CRITERION VARIABLE SR
 FOR THE POPULATION BOYS
 (N = 100)
 1990

Predictor Variables	EXPECTED						Chi Square	
	Below		At		Above			
	N	%	N	%	N	%		
	5	5	90	90	5	5		
OBSERVED								
Predictor Variables	N	%	N	At	%	Above	%	Chi Square
T	77	77.0	23	23.0	0	0.0		1091.7*
TS	76	76.0	23	23.0	1	1.0		1061.3*
TR	76	76.0	22	22.0	2	2.0		1061.4*
TSR	71	71.0	24	24.0	5	5.0		919.6*
CGQ	49	49.0	33	33.0	18	18.0		457.1*
PR	61	61.0	26	26.0	13	13.0		685.5*

Note. * Indicates significance demonstrated at .01 level p < .01 = 9.210

Hypothesis 4D

SR Criterion--Population 2nd Grade. For the second grade, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the SR estimates.

Frequencies, percents, and Chi square statistics for the predictor variables with the SR criterion variable, 2nd grade population, are presented in Table XXXVII.

The results of the Chi square analysis using the criterion variable SR for the 2nd grade population as seen in Table XXXVII yielded significant differences at the $p < .01$ level as compared to the 6 predictor variables T, TS, TR, TSR, CGQ and PR. Therefore, Hypothesis 4D was refuted.

It should be noted that the traditional predictor variables each placed from 18.8% to 21.7% of the 2nd grade population at the same instructional level as the SR criterion variable; 5.8% above the instructional level; and from 72.4% to 75% or three fourths of the 2nd grade population below the instruction level estimated by the SR criterion variable.

TABLE XXXVII
 CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
 FREQUENCIES FOR THE CRITERION VARIABLE SR
 FOR THE POPULATION 2ND GRADE
 (N = 69)
 1990

EXPECTED									
		Below		At		Above			
		N	%	N	%	N	%		
		34.5	5	62.1	90	3.45	5		
OBSERVED									
Predictor Variables	Below	%	N	At	%	Above	%	Chi Square	
T	52	75.4	13	18.8	4	5.8	722.1*		
TS	51	73.9	14	20.3	4	5.8	692.7*		
TR	52	75.4	13	18.8	4	5.8	722.1*		
TSR	50	72.4	15	21.7	4	5.8	663.9*		
CGQ	36	52.2	28	40.6	5	7.2	326.5*		
PR	37	53.6	24	34.8	8	11.6	355.6*		

Note. * Indicates significance demonstrated at .01 level p < .01 = 9.210

On the other hand, the CGQ qualitative predictor variable placed 52.2% below, 40.6% at, and 7.2% above the instructional level estimated by the SR criterion variable. The PR qualitative predictor variable demonstrates placement results similar to the CGQ predictor variable when compared to the SR criterion variable. That is, 53.6% were placed

below, 34.8% were placed at, and 11.6% were placed above instructional level when compared to the SR criterion variable. These results will be discussed further in Chapter V.

Hypothesis 4E

SR Criterion--Population 5th Grade. For the 5th grade population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the observed frequencies are the number of agreements between the predictor estimates and the SR estimates.

Frequencies, percents and Chi square values for the predictor variables with the SR criterion variable, for the 5th grade population are presented in Table XXXVIII.

The results of the Chi square analysis using the criterion variable SR for the 5th grade population as seen in Table XXXVII yielded statistical differences at the $p < .01$ level of significance as compared to the 6 predictor variables. Therefore, Hypothesis 4E was refuted.

It should be noted that the traditional predictor variables T, TS, TR, TSR each placed from 8.8% to 11% of the 5th grade population at the same level of instruction as the SR criterion variable; 14.7% above, and from 73.5% to 76.5%, or nearly three fourths of the 5th grade population below

the instructional level estimated by the SR criterion variable.

TABLE XXXVIII
 CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
 FREQUENCIES FOR THE CRITERION VARIABLE SR
 FOR THE POPULATION 5TH GRADE
 (N = 34)
 1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
1.7	5	30.6	90	1.7	5		
OBSERVED							
Predictor Variables	Below		At		Above		Chi Square
	N	%	N	%	N	%	
T	26	76.5	3	8.8	5	14.7	378.6*
TS	25	73.5	4	11.8	5	14.7	348.9*
TR	25	73.5	4	11.8	5	14.7	348.9*
TSR	25	73.5	4	11.8	5	14.7	348.9*
CGQ	11	32.4	14	41.2	9	26.5	91.2*
PR	14	41.2	12	35.3	8	23.5	123.6*

Note. * Indicates significance demonstrated at .01 level p < .01 = 9.210

On the other hand, the CGQ predictor variable placed 32.4% below, 41.2% at, and 26.5% above the instructional level estimated by the SR criterion variable. The PR qualitative predictor variable obtained placement results

similar to the CGQ predictor variable. That is, 41.2% were placed below, 35.3% were placed at, and 23.5% were placed above instructional level when compared to the criterion variable SR. These results will be further discussed in Chapter V.

Hypothesis 4F

SR Criterion--Population 3rd Grade. For the third grade population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the observed frequencies are the number of agreements between the predictor estimates and the SR estimates. Frequencies, percents, and Chi square values for the predictor variables, with the SR criterion, for the 3rd grade population, are presented in Table XXXIX.

The results of the Chi square analysis using the criterion variable SR for the population 3rd grade as seen in Table XXXIX yielded statistical differences at the $p < .01$ level of significance as compared to the six predictor variables. Therefore, Hypothesis 4F was refuted.

It should be noted that the traditional predictor variables T, TS, TR, TSR placed from 24.5% to 26.4% of the 3rd grade population at the same instructional level as the SR criterion variable; from 5.7% to 11.3% above; and 62.% to

68%, or nearly two thirds of the 3rd grade population below the instructional level estimated by the SR criterion variable.

TABLE XXXIX
 CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
 FREQUENCIES FOR THE CRITERION VARIABLE SR
 FOR THE POPULATION 3RD GRADE
 (N = 53)
 1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
2.65	5	47.7	90	2.65	5		
OBSERVED							
Predictor Variables	Below N	%	At N	%	Above N	%	Chi Square
T	36	67.7	14	26.4	3	5.7	433.6*
TS	36	67.7	13	24.5	4	7.5	445.6*
TR	36	67.7	13	24.5	4	7.5	445.6*
TSR	33	62.3	14	26.4	6	11.3	375.6*
CGQ	21	39.6	21	39.6	11	20.8	168.3*
PR	24	45.3	17	32.1	12	22.6	224.8*

Note. * Indicates significance demonstrated at .01 level p <.01 = 9.210

On the other hand, the CGQ predictor variable placed 39.6% at and below instructional level, and 20.8% above the instructional level estimated by the SR criterion level. The

qualitative predictor variable, PR placed 45.3% below, 32.1% at, and 22.6% above instructional level when compared to the criterion variable, SR. The implications of these results will be discussed in Chapter V.

Hypothesis 4G

SR Criterion--Population 4th Grade. For the 4th Grade population, there will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables, when the expected frequencies are 5%, 90%, and 5% respectively, and the observed frequencies are the number of agreements between the predictor estimates and the SR estimates.

Frequencies, percents, and Chi square values for the predictor variables, with the SR criterion, for the 4th grade population, are presented in Table XXXX.

As demonstrated in Table XXXX, for the population of 4th grade, significant differences were found at the $p < .01$ level between the six predictor variables T, TS, TR, TSR, CGQ and PR, and the criterion variable SR, in estimating the students' instructional reading levels. Therefore, Hypothesis 4G was refuted.

Each of the four traditional predictor variables placed from 63.6% to 68.2% of the 4th grade students below instructional levels as compared to the criterion variable SR. For these same traditional predictor variables, 20.5%

to 22.7% of the students were placed at instructional level, as compared to the SR criterion.

TABLE XXXX
 CHI SQUARE ANALYSIS OF EXPECTED AND OBSERVED
 FREQUENCIES FOR THE CRITERION VARIABLE SR
 FOR THE POPULATION 4TH GRADE
 (N = 44)
 1990

EXPECTED							
Below		At		Above			
N	%	N	%	N	%		
2.2	5	39.6	90	2.2	5		
OBSERVED							
Predictor Variables	Below		At		Above		Chi Square
	N	%	N	%	N	%	
T	30	68.2	10	22.7	4	9.1	374.9*
TS	29	65.9	9	20.5	6	13.6	356.7*
TR	29	65.9	9	20.5	6	13.6	356.7*
TSR	28	63.6	9	20.5	7	15.9	336.7*
CGQ	22	50.0	13	29.5	9	20.5	217.7*
PR	23	52.3	13	29.5	8	18.2	229.8*

Note. * Indicates significance demonstrated at .01 level p <.01 = 9.210

The traditional predictor variables also placed from 9.1% to 15.9% of the 4th grade students above the level of instruction estimated by the SR criterion variable.

The CGQ and PR qualitative predictor variables placed 50% and 52% below instructional level; 29.5% at instructional level and from 18.2% to 20.5% above the instructional levels indicated by the SR criterion variable.

CHAPTER V

SUMMARY AND CONCLUSIONS

IMPLICATIONS AND RECOMMENDATIONS

Introduction

Chapter V consists of three sections. The first section includes a summary of the purpose, research questions, and procedures. Additionally, the four hypotheses are repeated with a discussion of the important and significant findings for each of the seven populations.

The second section includes four general conclusions which respond to the four research questions and hypotheses. These conclusions are supported by the findings and review of the literature.

The third section concludes with implications and recommendations the researcher offers to others who may wish to pursue further investigations of this subject.

Summary

The purpose of this descriptive study was to determine the accuracy and "practicality" of the PR (poor reader) scoring system for the IRI (informal reading inventory), an individual assessment device designed to determine a student's instructional reading level. The PR was compared

with four traditional and largely quantitative scoring systems (T, TS, TR, TSR), and one qualitative system, the CGQ. These results were further compared to four criterion scores (CG, AE, TE, SR). The data were examined through the use of Chi square tests of significance to determine if there were any differences between the results of the various scoring systems.

Research Questions

The study answers the following research questions:

1. Is the qualitative PR (poor reader) scoring system more practical and accurate than four traditional scoring systems (T, TS, TR, TSR) in placing students in instructional materials?
2. Is the PR scoring system as practical and accurate as May's CG (context-graphophonic) scoring system based on miscue and IRI research findings?
3. Is the PR scoring system as accurate as an experienced and knowledgeable reading coordinator's estimates of the student's level of instruction?
4. How does the PR scoring system compare with two other criterion scores, TE (teacher estimates) and SR (silent reading) scores, in placing students in instructional materials?

Procedures

Ten Chapter I teachers each selected 20 second through fifth grade students from their population to be tape recorded reading the Silvaroli Informal Reading Inventory. Two hundred oral reading samples were collected, miscues coded, and the results scored seven times using the four traditional and three qualitative methods already described. Chi square was used to test the differences between expected and observed frequencies of students being "accurately" placed with respect to each criterion. Leading up to the Chi square computations, the six predictor variables (T, TS, TR, TSR, CGQ, PR) were first compared with each of the four criterion variables (CG, AE, TE, SR) for each of the populations (Total, Girls, Boys, 2nd, 5th, 3rd, 4th grade).

The four hypotheses followed by a summary of the conclusions for each of the sub-populations are now presented.

Statistical Hypotheses and Results

Hypothesis 1--CG as Criterion. There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables and each of the seven population groupings, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements

between the predictor estimates and the estimates made through the CG system (May's context-graphophonic scales).

Statistically significant differences were found between the instructional placements estimated by the four quantitative scoring systems and the CG criterion system at the $p < .01$ level for all the population groups including Total, Girls, Boys, 2nd, 5th, 3rd, and 4th grades. Rounding off to the nearest percent, with the quantitative systems the instructional estimates coincided with those of the CG criterion only 21 to 40% of the time. From 48 to 77% of the time the quantitative systems estimated below the CG criterion estimates, and from 0 to 3% of the time they estimated above.

The 5th grade had the largest percent of students placed below the expected level estimated by the research-based criterion CG while the 2nd grade had the smallest percent. There were no statistically significant differences between the PR qualitative placements and the CG criterion placements. The PR system placed students the same as the CG system from 89 to 94% of the time; below 6 to 9% of the time and above, 3% of the time. Since no more than 10% of students were "misplaced," the PR system met the "practicality" requirement of the study. The CGQ placements were significantly different from the CG placements. However, this difference turned out to be a statistical anomaly, caused by the CGQ system's success in estimating as

well as the CG system. Only 0 to 3% of the students were placed lower and 1.5% of the students were placed above the CG expected placement. The observed frequencies were 3 below, 196 the same, and 1 above while the expected were 10, 180, and 10, respectively. The observed yielded 34 cases different from the expected and thus the difference from a statistical standpoint. The CGQ system met the "practicality" requirement, with 99% of the placements the same as those of the CG criterion.

Hypothesis 2--AE as Criterion. There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables and each of the seven population groupings, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the (AE) administrator estimates made by the IRI administrator. Statistically significant differences were found between the instructional placements estimated by the four quantitative scoring systems and the AE criterion estimates at the $p < .01$ level for all of the population groupings including Total, Girls, Boys, 2nd, 5th, 3rd and 4th grades. Rounding off to the nearest percent for all population groups, with the quantitative systems the instructional estimates coincided with those of the AE criterion only 20 to 51% of the time.

From 49 to 79% of the time the quantitative systems estimated below the AE criterion estimate, and from 0% to 7%, they estimated above.

There were no statistically significant differences between the CGQ qualitative placements and the AE criterion placements. The CGQ system placed students the same as the IRI administrator estimates (AE) from 86 to 97% of the time, with 3 to 8% below, and 0 to 6% above. Since no more than 10% of students were "misplaced," the CGQ system met the "practicality" requirement of the study.

Although there was a statistically significant difference between the PR qualitative placements and the AE criterion for the total population, it was very small and caused by the success of the PR system in predicting the criterion estimates of instructional level 96% of the time. The observed frequencies were 7 below, 192 the same, and 1 above, while the expected were 10, 180, and 10, respectively. The observed data yielded 24 cases different from the expected, and thus the small but significant difference from a statistical standpoint. The PR system met the practicality requirement of no more than 10% error. From 91 to 99% of the subpopulation groups were placed at the same instructional level as the AE criterion by the PR predictor. From 2 to 9% were placed below and from 0 to 2% were placed above. The PR met the practicality requirement for a scoring system.

Hypothesis 3--TE as Criterion. There will be no significant differences between observed and expected frequencies of students placed below, at, and above their estimated reading level for each of six predictor variables and each of the seven population groupings, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the teacher estimates made by the reading teachers.

Statistically significant differences at the $p < .01$ level for all of the population groups including Total, Girls, Boys, 2nd, 5th, 3rd, and 4th grades were found between the instructional placements estimated by the four quantitative scoring systems and those estimated by the reading teachers (TE). With the quantitative systems the instructional estimates coincided with those of the TE criterion for all of the populations only 23 to 39% of the time. From 40 to 74% of the time the quantitative systems estimated below the TE criterion, and from 7 to 27% of the time they estimated above. The 3rd grade demonstrated the least agreement with the traditional scoring systems, while the 4th grade had the highest agreement.

Statistically significant differences were also found between the qualitative systems and the TE criterion system. For the CGQ and PR systems the instructional estimates coincided with those of the TE criterion only 27 to 58%,

with 9 to 19% being estimated below, and 42 to 64% estimated above the TE criterion placement. The traditional scoring systems placed a higher frequency of students below the instructional estimate of the TE, while the two qualitative scoring systems placed more students above the instructional estimates of the TE.

Hypothesis 4--SR as Criterion. There will be no significant differences between observed and expected frequencies of students placed below, at, and above their instructional reading level, for each of six predictor variables and each of seven population groupings, when the expected frequencies are 5%, 90%, and 5% respectively and the observed frequencies are the number of agreements between the predictor estimates and the estimates produced through the silent reading scores on form B.

Statistically significant differences at the $p < .01$ level for all of the population groups, Total, Girls, Boys, 2nd, 5th, 3rd, and 4th grade were found between the instructional placements estimated by the four quantitative scoring systems and the silent reading estimates (SR). For the quantitative systems the instructional estimates coincided with those of the SR criterion only 9 to 24% of the time. From 62 to 77% of the time the quantitative systems predicted an instructional level below the predictions based on silent reading scores; from 0 to 18% of

the time the quantitative predictions were above the estimated levels of the SR criterion.

Statistically significant differences were also found between the instructional placements estimated by the two qualitative scoring systems and the silent reading predictions. With the qualitative systems the instructional estimates coincided with those of the SR criterion only 26 to 43% of the time, with 32 to 61% being placed below the SR estimates and 16 to 27% being placed above the levels estimated by the SR criterion. This lack of agreement between the SR criterion and the predictor variables is interesting in the light of the research and practice related to the use of questions to determine instructional reading levels. These results and their implications will be further discussed along with the other significant results as they relate to the four hypotheses.

Comparisons for all seven population groups using the two qualitative scoring systems and the most qualitative of the quantitative systems as to their agreement with the four criterion systems, are presented in Table XXXXI.

TABLE XXXXI

PERCENTAGE OF AGREEMENT ON INSTRUCTIONAL LEAVE ESTIMATES BETWEEN
CRITERION VARIABLES AND THE PR, CGQ, AND THE MOST
QUALITATIVE TRADITIONAL SCORING SYSTEM
1990

	Criterion Variables for the Study											
	May's CG System (CG)			IRI Administrator (AE)			Reading Teachers (TE)			Silent Reading (SR)		
	PR	TSR %	CGQ	PR	TSR %	CGQ	PR	TSR %	CGQ	PR	TSR %	CGQ
Total	91	39	99	96	40	91	42	36	44	33	21	38
Girls	90	38	99	95	45	92	33	39	27	41	18	43
Boys	91	40	98	97	35	90	40	32	31	26	24	33
2nd	90	51	97	99	52	88	58	39	58	35	22	41
5th	94	27	100	94	29	97	35	29	41	35	12	41
3rd	89	36	100	93	32	87	36	36	38	32	26	40
4th	91	32	98	98	39	96	30	34	32	30	21	30

Comparison for All Seven Population Groups

As shown in Table XXXXI the PR scoring system, the main focus of this report, yielded estimates that coincided with either the research criterion (CG) or the judgment of the IRI administrator (AE) 13 out of 14 times at the practicality level of no more than 10% error. For one of the population groups the error was 11%. Overall for the total group of 200 children in the second through fifth grade, the agreement between PR and CG was 91%; the agreement between PR and AE was 96%.

The other qualitative scoring system, the CGQ, yielded estimates that coincided with either the CG or the AE criterion 12 out of 14 times at the practicality level of no more than 10% error. For the total group of 200 children the agreement between CGQ and AE was 91%. The agreement between CGQ and CG was 99%, which will require a discussion in the next section of the importance of IRI questions. It should be noted, however, that the TRS system, in which the scorer omits self corrections and repeats, thus theoretically making it a more quantitative scoring system, yielded estimates that agreed only 39% with CG and 40% with AE.

Also to be noticed is that the agreement on instructional level estimates between all seven scoring systems and the other two criterion variables, TE and SR, as

already shown in Tables XXVII and XXXIV, ranged from 20 to 38%.

Finally, it should be observed in Table XXXXI that the PR scoring system yielded estimates that agreed at least 90% of the time with both the research criterion (CG) and the IRI administrator (AE) for the most extreme age levels in this study, second grade and fifth grade. The percentages for the second grade group were 90 and 99, respectively; for the fifth grade 94%.

Additional Results

The AE as the data collector found that the PR scoring system took only one fourth of the time of the CG scoring system. The CG required a more complex approach to the analyzing of substitution miscues, both quantitatively and qualitatively, while the PR merely required looking at defaults and meaning denigrating substitutions. This could be done at the time of the taping of the sample. As the data collector and scorer, the AE potentially contributed to a sampling bias.

CONCLUSIONS AND DISCUSSION

The following conclusions, limited to the population of the present study, serve as answers to the four research questions and provide promise and direction for future research.

Conclusion One

The PR (poor reader) qualitative scoring system was more practical and accurate than the four traditional quantitative scoring systems in placing students in instructional materials.

Assuming sufficient similarity between the study's population and any new population to be tested, the PR scoring system would misplace no more than 10%, or three students in a class of thirty. This was demonstrated by the degree to which the PR estimates coincided with the CG and AE criterion variables. The percentage of misplacement for the total population was 9% with respect to the CG criterion and 4% with respect to the AE criterion. Such minor misplacement should be observed in contrast to the most qualitative of the four quantitative systems, the TSR, which misplaced 61% of the students in comparison with the CG criterion and 60% in comparison with the AE criterion.

Conclusion Two

The PR scoring system was about as accurate as, and also more practical than, May's two-scaled procedure, the CG (context-graphophonic) system in placing students in instructional materials. The PR system placed students in instructional materials as accurately (90% or more of the time) as May's more time-consuming but research-based CG scoring system. The CG scoring system requires the use of two scales, one for context sensitivity and one for

graphophonic sensitivity, both of which require numerous judgments. Although the CG system is relatively fast when compared with other qualitative systems developed in the past, it requires the use of two scales, one for context sensitivity and one for graphophonic sensitivity, both of which require numerous judgements. The PR demonstrated its potential as the most practical of the scoring systems in the study since the IRI administrator and scorer (AE) found that the PR system took only about one-fourth of the time of the CG scoring system. The PR demonstrated its potential as the most practical and accurate of the scoring systems in the study.

Conclusion Three

May's PR system was as accurate as the estimates made by a highly experienced reading coordinator and IRI administrator (AE), who had knowledge of the research on miscues and informal reading inventories.

. This has already been demonstrated by the high degree of agreements on estimates of instructional level between PR and AE. The level of agreement with AE, the expert practitioner, was 96%.

Conclusion Four

The PR scoring system did as well as any of the other predictor variables in making estimates that agree either with the TE criterion (reading teachers' estimates) or the

SR criterion (silent reading scores based on answering of questions) in placing students in instructional materials. The estimates derived from the six scoring systems for all seven populations agreed with the two criterion scores (TE and SR), less than 50% of the time. Data to substantiate this observation were obtained from the questionnaire completed by the teachers participating in this study. One item completed by the teachers reflected very little agreement as to what is most important as an information source when determining a student's instructional reading level. Areas such as prior placement, standardized test scores, and classroom teacher referrals were selected as criteria, with no single result surfacing as a more favored criterion.

The questionnaire, administered to the reading teachers before the IRI testing began, revealed that these teachers seemed largely unaware of the research of the seventies and eighties on miscues and informal reading inventories. They appeared to favor mastery criterion tests, an indicator that their view of reading was that of a cumulative sub-skill process rather than a process of "intelligent and adventurous predicting and checking", as supported by the currently accepted definition of reading repeated here: "Reading is the process of constructing meaning through the dynamic interaction [transaction] among the reader's existing knowledge, the information suggested by the written

language, and the context of the reading situation" (Weaver, 1988, p. 5).

This result may also be influenced by the teachers' perception of the district level support of a criterion test as an appropriate performance indicator. The data from the questionnaire describing these teachers' training and experience reflected some specific limitations in their background. A different group of reading teachers, with different training, might have served as a very different criterion group.

This matches the results found by numerous researchers who have attempted to use teacher judgments as a criterion measure. The most notable study with respect to this problem was one done by Haller and Waterman (1985), in which they found that when teachers made judgments on children's level of reading, only 31 to 45% of the time were their judgments based on actual reading ability. Other factors that often weighed more heavily were general academic competence, work habits, behavior and personality, and home background.

With regard to the weakness of SR (silent reading scores) as a criterion, it is probable that the inadequacy of questions as a measure of comprehension is inadequate. Because schemata and memory are factors which influence these results the literature, previously reviewed, regards such an approach to measuring comprehension as an

unreliable, invalid indicator. This approach was used as a criterion because of its prevalent use, not due to its effectiveness. Miscue analysis, on the other hand, can serve as a method of viewing comprehension "in process".

In retrospect, the selection of the use of (TE) teacher estimates and (SR) silent reading scores as criteria for such a study probably needs to be examined because of the limits of their strengths as true performance indicators. This is due to the varied background and experience of the teachers which effected their ratings of the students reading performance. It was also due to the research questioning SR question-answering scores as valid comprehension indicators.

Whether the TE or SR criterion measures used in this study are accepted or not, the PR system proved to be nearly as "accurate" as any other scoring system examined here, both qualitative and quantitative, in yielding instructional estimates similar to those from a group of reading teachers (TE) or from comprehension scores on silent reading passages (SR). The agreement between the TE criterion and the PR, TSR, and CGQ predictors was 42, 36, and 44 percent. The agreement between the SR criterion and the PR, TSR, and CGQ predictors was 33, 21, and 38 percent. It should be remembered however, that none of the seven scoring systems

produced estimates that agreed very much with either of these two criterion measures.

IMPLICATIONS AND RECOMMENDATIONS

The main implications drawn from this study are as follows.

Implication One

It may be advisable for classroom teachers and reading teachers to make use of May's PR scoring system for the IRI as a quick and qualitative way of estimating a student's instructional reading level. Since May's PR scoring system for the IRI placed students for instruction as accurately as a research-based criterion system (CG), as accurately as the estimates of an experienced reading coordinator with extensive knowledge of the research on miscues and IRIs (AE), and much more accurately than any of the four commonly used quantitative scoring methods (T, TS, TR, TSR), it may be advisable for classroom teachers and reading teachers to try this easily applied qualitative system. With this system the teacher can ignore all miscues other than defaults and poor substitutions. When these two miscues, which are true indicators of comprehension, occur more than 5% of the time in a passage, it is likely that the student has reached his or her frustration level. Furthermore, the teacher need not ask any questions, thus cutting the time for administration of the IRI in half.

Implication Two

Since the PR scoring system has met, in this study, the following eight stringent criteria established prior to the study for a "qualitative IRI scoring system," researchers may wish to employ this quick system as one of the independent variables used in studies related to improving informal reading inventories.

For classroom teachers and reading teachers the PR system was quick to use as well as clear and easy to follow. This was determined during the pilot conducted prior to this study through the used of two-hour workshops with over two hundred teachers. In all such training sessions, the participants were able to agree on instructional placements for children, with the workshop instructor and with each other, 90% of the time or better.

The PR system showed promise of utility over a wide range of grades. In this study the instructional range of 2% to 5% poor reader errors seemed appropriate, according to the reading coordinator who administered the IRI's, at all of the grades tested--from second through fifth. The agreement of instructional placement, furthermore, between AE, the reading coordinator's judgments, and PR, May's qualitative scoring system, was 96% for all the grades combined. For second grade alone the agreement was 99%; for fifth grade alone the agreement was 94%. The error of both

second and fifth grade was within the practicality limit of 10%.

The PR system required the use of oral reading rather than silent reading, thus allowing the researcher to obtain diagnostic miscue-analysis evidence of a student's reading concept and reading strategies, whether the text was easy, average, or difficult. This was true in spite of the fact that the person doing the scoring was only allowed to examine defaults and meaning-denigrated substitutions. Through examination of poor substitutions, for example, the scorer was able to determine the extent to which the letters and phonemes in the reader's substitution matched those of the author's word; the scorer could also determine the extent to which the reader relied on the first-letter phonics strategy (get the first one and guess the rest); and the scorer could infer whether the reader thought that reading was a search for meaning or that the reading was mainly a decoding and get-the-job-done operation. The analysis of the use of such strategies allows the teacher to apply current theory and research about miscue analysis and the use of IRI's to determine whether the reader is making "good" or "poor" reader miscues. Through examination of defaults the scorer was capable of determining the extent to which the student was a courageous predictor, a careful checker, a "beggar," and/or a person lacking strategies for detecting phonograms, syllables, and affixes. For nearly

every child tested, the production of several poor reader errors was the clearest sign of the frustration level.

The PR system reflected the research on the meaning of various miscues. By ignoring insertions, omissions, and repeats, this system allowed the researchers to utilize the research findings, discussed in the review of the literature that these three types of miscues do not interfere with comprehension. By ignoring self-corrections as well, this procedure allowed the researcher to concentrate on the two types of miscues that truly differentiate a poor reader from a good reader, namely defaults and meaning-denigrating substitutions. Even good readers, it was found, resort to these two when faced with very difficult text. As mentioned before, these two forms of "error," can provide teachers with an easily determined indication of frustration level--when committed with respect to over five percent of the words in the text. None of the other forms of miscue provides such a clear, unambivalent indication of frustration.

The PR system used in this study did not penalize good readers in order to get estimates of instructional level low enough for poor readers. This has been one of the arguments used in the past, that we must count everything or "poor reader Johnny" might be placed in frustrating material. This argument has ignored the damage done to good readers who have been held back by estimates of instructional level

that were too low. Such estimates have led to boredom, frustration, and lack of growth--certainly too high a price to pay in order to keep others from being frustrated with difficult text. The PR system made this either-or decision unnecessary. No one, either good readers or poor readers, was penalized for using insertions, omissions, self corrections, or repetitions. Only those who used defaults or meaning-denigrated substitutions were penalized, and this was only those students who were definitely reaching their frustration level anyway.

The PR system placed students in instructional materials as accurately (90% or more of the time) as May's more time-consuming but research-based CG scoring system.

The CG scoring system requires the use of two scales, one for context sensitivity and one for graphophonic sensitivity, both of which require numerous judgments. Although the CG system is relatively fast when compared with other quantitative systems developed in the past, the IRI administrator and scorer (AE) in the present study found that the PR system took only about one-fourth of the time.

May's PR system placed students in instructional materials as accurately (90% or more of the time) as a reading coordinator with many years of experience and with knowledge of the research on miscues and informal reading inventories. The actual agreement between the AE

(administrator estimate) criterion variable and the PR scoring system was 96%. At this rate only one student in a class of 25 would be misplaced by the PR system.

For a scoring system to be considered reliable in this study, it had a coefficient of interjudge reliability of at least 0.90.

The PR system produce estimates with an interjudge reliability coefficient of 0. Before computing this coefficient the IRI test administrator trained a teacher for a total of two hours on the use of the four quantitative and three qualitative systems used in this study. The PR system by itself took about 30 minutes of training. Since the resulting data were nonparametric, Kendall's W (coefficient of concordance) was used to estimate consistency. Whereas Pearson's yielded a coefficient of 0.99, Kendall's W, a more conservative technique used for ordinal or nominal data, yielded a coefficient of 0. The results of this interrater reliability study verify the ease of training teachers to use the PR scoring system as well as the practicality and accuracy of the system itself.

In summary all eight of the criteria established before the study as indicators of an effective qualitative scoring system for the IRI were tested. All eight of the criteria for May's PR scoring system were met.

CLOSING STATEMENT

The qualitative approach to the scoring of IRIs has the potential of enhancing the richness of a student's educational experience by providing the classroom teacher and reading teacher with meaningful information upon which to base instructional and placement decisions. Furthermore, process oriented assessment, such as the qualitative approach to the scoring of informal reading inventories, could be an authentic and trust worthy source of data to guide the decision-making of the classroom teacher.

The PR scoring system is such a qualitative approach to assessment and instructional decision making. Because of its practicality and accuracy, and because it is easy to learn to use, as demonstrated in the pilot training as well as in the training of the inter rater scorer, it can be an excellent tool for the practitioner seeking to help students to be successful.

The current focus on meaningful assessment supports the use of such an approach, for it can enable the teacher to focus on the process of reading, not on the output of a product. Such authentic assessment is a way to measure a student's strengths and weaknesses in the act of reading, not while performing meaningless isolated skills.

With May's qualitative approach to assessment, fewer students will be likely to be penalized and labeled as "at risk" learners for "errors" made during oral reading.

Rather the assessment process will be just that, an assessment process, not a focus on a product. Furthermore, the student becomes a self-evaluator and as such gains self-confidence and a sense of control over his or her learning.

Reading assessment which provides information regarding the repertoire of strategies a student uses during the reading act is sound assessment. Often times, current reading assessment practices are time consuming and it is often difficult to obtain and use the results in an efficient and timely manner. May's approach to the qualitative scoring of informal reading inventories provides a method to give quick access to such information to the practitioner.

A recommendation for future practice includes a study of May's qualitative scoring system using passages from the student's regular text, to further authenticate the assessment process. Additionally, a replication of the study, using a stratified random sampling would extend the implications of the results. An equal number of students in each grade level would further validate the results of this study. A future study should include the training of the reading teachers involved in the study enabling them to use the PR qualitative scoring system. This would eliminate the missing element of this study, the lack of involvement of the reading teachers in the scoring system.

The attitude toward qualitative assessment in general needs to be enhanced. The provision of more information regarding its effectiveness would support this need. Qualitative assessment is a powerful tool the classroom teacher could use for "action oriented research" and as an excellent diagnostic, placement, and instruction tool. Often times current reading assessment practices are time-consuming and it is often difficult to obtain and use the results in an efficient manner. This study could provide support for an approach to assessment which could address these current deficiencies.

REFERENCES

- Aaron, I., Hutchison, S., & Norris, J. (1989). Classroom Reading Inventory (4th ed.). In J. Conaly & J. Kramer (Eds.), Tenth Mental Measurements Yearbook (pp. 177-180). Lincoln: University of Nebraska.
- Aaronson, E., & Farr, R. (1988). Issues in assessment. Journal of Reading, 32(2), 174-177.
- Allen, P. (1976). The miscue research studies. In P. Allen & D. Watson (Eds.), Findings of research in miscue analysis: Classroom implications (NCTE Stock No. 17333) (pp. 3-10). Urbana, IL: NCTE.
- Anderson, B., & Joels, R. (1984). Informal reading inventory scoring criteria: Should repetitions count? Journal of Reading Education, 11(1), 20-24.
- Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. (1985). Becoming a nation of readers. (Contract No. 400-83-0057). Washington, DC: National Institute of Education.
- Argyle, S. (1989). Miscue analysis for classroom use. Reading Horizons, 29, 94-1022.
- Bader, L. A., & Wiesendanger, K. D. (1989) Realizing the potential of informal reading inventories. Journal of Reading, 32(5), 402-408.
- Baumann, J. F. (1988). Reading assessment, an instructional decision-making perspective. Columbus, Ohio: Merrill.
- Beebe, M. (1980). The effect of different types of substitution miscues on reading. Reading Research Quarterly, 15(3,) 324-336.
- Bergeron, B. (1990). What does the term whole language mean? Journal of Reading, 22(4), 301-326.
- Betts, E. A. (1946). Foundations of reading instruction. New York: American Book Co.

- Breznitz, Z. (1989). Teacher evaluation of pupils' reading performance: Criteria used for assessment. Reading Improvement, 26(2), 174-180.
- Bristow, P. S., Pikluski, J., & Pelosi, P. L. (1983). A comparison of five estimates of reading instructional level. Reading Teacher, 37, 273-279.
- Brozo, W. (1990). Learning how at-risk readers learn best: A case for interactive assessment. Journal of Reading, 33(7), 522-527.
- Brown, L., & Sherbenou, R. (1981). A comparison of teacher perceptions of student reading ability, reading performance, and classroom behavior. Reading Teacher, 34, 557-560.
- Brown, S. (1964). A comparison of five widely used standardized reading tests and an informal reading inventory for a selected group of elementary school children. Dissertation Abstracts International, 25, ADG 80-28323.
- Burke, C. L. (1976). Reading miscue research: A theoretical position. In P. D. Allen & D. J. Watson (Eds.), Findings of research in miscue analysis: Classroom implications (NCTE Stock No. 17333) (pp. 10-23). Urbana, IL: NCTE.
- Burns, P. & Roe, B. (1985). Informal reading inventory (2nd ed.). Boston: Houghton-Mifflin.
- Bussis, A. M. & Chittendon, E. A. (1987). What the reading tests neglect. Language Arts, 64(3), 302-308.
- Cadenhead, K. (1987). Reading level: A metaphor which shapes practice. Phi Delta Kappan, 68(6), 436-441.
- Calfee, R. (1987). The school as a context for assessment of literacy. Reading Teacher, 40, 738-743.
- Cardarelli, A. (1988). The influence of reinspection of students' IRI results. Reading Teacher, 41(7), 664-667.
- Carroll, J. B. (1987). The national assessments in reading: Are we misreading the findings? Phi Delta Kappan, 68(6), 424-430.
- Christie, J. F. (1979). The qualitative analysis system: Updating the IRI. Reading World, 18, 393-399.

- Clay, M. (1979). The early detection of reading difficulties: A diagnostic survey with recovery procedures. Portsmouth: Heinemann Press.
- Coleman, M., & Harmer, W. R. (1982). A comparison of standardized reading tests and informal placement procedures. Journal of Learning Disabilities, 15(7), 396-398.
- Corbett, M. (1989). Support for the learning and teaching of English (SLATE). Urbana, IL: NCTE.
- Cramer, E. (1980). Informal reading inventories go commercial. Curriculum Review, 19(5), 424-429.
- Cunningham, J. (1984). A simplified miscue analysis for classroom and clinic. Reading Horizons, 24(2), 83-86.
- D'Angelo, K., & Mahalios, M. (1983). Insertion and omission miscues of good and poor readers. The Reading Teacher, 36(8), 778-782.
- Davey, B., & Theofield, M. (1983). An investigation of miscue analysis training for teachers. Reading Improvement, 20(2), 105-110.
- Dunkeld, C. G. (1970). The validity of the informal reading inventory for the general designation of instructional reading levels: A study of the relationships between children's gains in reading achievement and the difficulty of instructional materials. Dissertation Abstracts International, 31, 6274-A. (University Microfilms No. 71-14733).
- Ebel, R. (1977). The uses of standardized testing. Bloomington, Indiana: Phi Delta Kappa Educational Foundation.
- Ekwall, E. (1985). Ekwall Reading Inventory. Boston: Allyn & Bacon.
- Englert, C., & Semmel, M. (1981). The relationships of oral reading substitution miscues to comprehension. The Reading Teacher, 35(3), 373-379.
- Farr, R., & Carey, R. F. (1986). Reading: What can be measured? Newark, Delaware: International Reading Association.
- Fleisher, B. (1987-88). Oral reading cue strategies of better and poorer readers. Reading Research and Instruction, 27(3), 35-50.

- Gay, L. R. (1987). Educational research: Competencies for analysis and application. Columbus, OH: Merrill Publishing Co.
- Gillis, M., & Olson, M. (1987). Elementary IRIs: Do they reflect what we know about text type/structure and comprehension? Reading Research and Instruction, 27(1), 36-44.
- Glazer, S., Searfoss, L., & Gentile, L. (Eds.). (1988). (Eds.) Re-examining reading diagnosis: New trends and procedures. Newark, Delaware: International Reading Association.
- Goodman, K. (1981). Miscues: Windows on the reading process. In K. Goodman (Ed.), Miscue analysis: applications to reading instruction (NCTE Stock No. 31883) (pp. 3-14). Urbana, IL: NCTE.
- Goodman, K. (1976). What we know about reading. In P. Allen & D. Watson (Eds.), Findings of research in miscue analysis: Classroom implications (NCTE Stock No. 17333) (pp. 57-69). Urbana, IL: NCTE.
- Goodman, K., & Burke, C. (1968). Study of childrens' behavior while reading orally (Report No. 147). Washington, DC: US Department of HEW/CEO.
- Goodman, K., & Gollasch, F. (1980). Word omissions: Deliberate and non-deliberate. Reading Research Quarterly, 16(1), 6-31.
- Goodman, Y., & Burke, C. (1972). Reading miscue inventory: Procedure for diagnosis and evaluation. New York: Macmillan.
- Goodman, Y., Watson, D., & Burke, C. (1987). Reading miscue inventory: Alternative procedures. New York: Macmillan.
- Groff, P. (1980). A critique of an oral reading miscue analysis. Reading World, 19(3), 254-64.
- Guzak, F. (1970). Dilemmas in informal reading assessment. Elementary English, 47, 666-670.
- Haller, E., & Waterman, M. (1985). The criteria of reading group assignments. Reading Research, 38(8), 772-781.
- Haney, W., & Madaus, G. (1978). Making sense of the competency testing movement. Harvard Educational Review, 48, 462-484.

- Harker, W. J. (1990). Testing reading: We need a new perspective. Reading Horizons, 30(4), 309-317.
- Harlin, R., & Lipa, S. (1990). Emergent literacy: A comparison of formal & informal assessment methods. Reading Horizons, 30, 209-223.
- Harris, A., & Sipay, E. (1985). How to increase reading ability: A guide to developmental and remedial methods (8th ed.). New York: Longman & Co.
- Harris, L., & Lalik, R. (1987). Teachers' use of informal reading inventories: An example of school constraints. The Reading Teacher, 40(7), 624-631.
- Harris, L., & Niles, J. (1982). An analysis of published informal reading inventories. Reading Horizons, 22, 159-174.
- Harris, T., & Hodges, R. (Eds.). (1981). A dictionary of reading and related terms. Newark, Delaware: International Reading Association.
- Helgren, V. (1984). An analysis of alternate form reliability of three commercially prepared informal reading inventories (Doctoral Dissertation, University of Miami, 1984). Dissertation Abstracts International, 32, ADG 84-16328)
- Henk, W. (1987). Reading assessments of the future: Toward precision diagnosis. The Reading Teacher, 40, 860-867.
- Hoffman, J. V. (1980). Weighting miscues in informal inventories: A precautionary note. Reading Horizons, 20(20), 135-139.
- Holmes, B. (1983). The effect of prior knowledge on the question answering of good and poor readers. Journal of Educational Psychology, 15(4), 1-17.
- Homan, S., & Klesius, J. (1985). A reexamination of the IRI: Word recognition criteria. Reading Horizons, 26, 54-61.
- Hood, J. (1976). Qualitative analysis of oral reading errors: The inter-judge reliability of scores. Reading Research Quarterly, 11(4), 577-98.
- Huey, E. (1968). The Psychology and Pedagogy of Reading. New York: MIT Press. (Original work published 1908)

- Joels, R., & Anderson, B. (1988). Informal reading inventory comprehension questions: Are classification schemes valid? Reading Horizons, 28, 178-184.
- Johns, J. (1982). The dimensions and uses of informal reading assessment. In J. J. Pikulski & T. Shanahan (Eds.), Approaches to the Informal Evaluation of Reading (pp. 1-11). Newark, DE: International Reading Association.
- Johns, J., Gorton, S., Schoenfielder, P., & Skriba, P. (Eds.). (1977). Assessing reading behavior: Informal reading inventories. Newark, Delaware: International Reading Association.
- Johns, J., & Magliari, A. (1989). Informal reading inventories: Are the Betts criteria the best criteria? Reading Improvement, 26(23), 124-132.
- Johns, J. L., & Lunn, M. K. (1983). The informal reading inventory: 1910-1980. Reading World, 23(5), 8-19.
- Johnson, D. A. (1990). The relationship between school integration and student attitude toward residential racial integration. Unpublished doctoral dissertation, Portland State University, Portland, Oregon.
- Johnson, M. S. (1978). Classroom reading inventory (3rd ed.) (pp. 1234-1235). In O. K. Buros (Ed.), Eighth mental measurements yearbook II (3rd ed.) (pp. 1234-1235). Highland Park, NJ: Gryphon Press.
- Johnson, M. S., Kress, R. A., & Pikluski, J. J. (1987). Informal reading inventories. Newark: International Reading Association.
- Johnston, P. H. (1984). Assessment in reading. In P. D. Pearson (Ed.), Handbook of Reading Research (pp. 147-182). New York: Longman.
- Johnston, P. (1987). Teachers as evaluators. Reading Teacher, 40(7), 744-748.
- Klesius, J., Searles, E., & Hines, C. (1988). What's the value of an IRI? Is it being used? Reading Horizons, 28(2), 92-102.
- Kletzien, S., & Bednar, M. (1990). Dynamic assessment for at-risk readers. Journal of Reading, 33(7), 528-533.
- Lesiak, J. & Johnson, B. (1983). Reading assessment for placement and programming. New York: Longman.

- Leu, D. (1982). Oral reading error analysis: A critical review of research and application. Reading Research Quarterly, 17(3), 420-435.
- Lidz, C. (1981). Improving assessment of school children. San Francisco, CA: Jossey-Bass.
- Lindberg, S. (1987). The relationship between product and in-process assessments of reading of third-grade students. Unpublished doctoral dissertation, University of Southern California. (ERIC Document Reproduction Service No. ADG05-61034 8801)
- Little, C. (1989). Validity: Whose construction is it in the teaching and learning context? Educational Measurement Issues and Practice, 8(1), 8-13.
- Long, P. (1987). The effectiveness of reading miscue instruments (Occasional Paper #13). Tucson: University of Arizona, College of Education.
- Marek, A., Goodman, K., & Babcock, P. (1985). Annotated miscue analysis bibliography. Program in language and literacy (Occasional paper #16). Tucson: Arizona University, Tucson College of Education.
- Masztal, N., & Smith, L. (1984). Do teachers really administer IRIs? Reading World, 24, 80-83.
- May, F. (1990). Reading as communication: An interactive approach. Columbus: Merrill.
- McCracken, R. (1964). The development and validation of the standard reading inventory for the individual appraisal of reading performance. In J. A. Figuel (Ed.), Improvement of reading through classroom practice (pp, 310-313). Newark, DE: International Reading Association.
- McKenna, M. (1983). Informal reading inventories: A review of the issues. The Reading Teacher, 36(7), 670-679.
- Mosenthal, P. (1988). Understanding the histories of reading. The Reading Teacher, 41, 64-65.
- Nickerson, R. S. (1989). New directions in educational assessment. Educational Researcher, 18(9), 3-7.
- Nicholson, T., Lillas, C., & Rzoska, A. (1988). Have we been misled by miscues? The Reading Teacher, 41, 6-10.

- Peterson, J., Greenlaw, M., & Tierney, R. (1978). Assessing instructional placement with the IRI, the effectiveness of comprehension questions. Journal of Educational Research, 71, 247-250.
- Pikulski, J. (1990). Informal reading inventories. The Reading Teacher, 43(7), 514-516.
- 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, (pp. 94-116). Newark, DE: International Reading Association.
- Portland Public Schools. (1989). Enrollment Report, October, 1989. Portland, Oregon: Author.
- Potter, F. (1980). Miscue analysis: A cautionary note. Journal of Research in Reading, 3(2), 116-128.
- Powell, W. (1978, May). Measuring reading performance informally. Paper presented at the annual meeting of the International Reading Association, Houston, Texas.
- Powell, W. R. (1960). Reappraising the criteria for interpreting informal reading inventories. In D. Doboer (Ed.), Reading diagnosis and evaluation. (pp. 76-82). Newark, DE: International Reading Association.
- Powell, W. R., & Dunkeld, C. (1971). Validity of the J. R. and reading levels. Elementary English, 48(6), 637-642.
- Readence, R. S. (1987). Interactive assessment. In J. Readence & R. Scott, R., (Ed.), Research in Literacy: Merging Perspectives (pp. 201-205). Rochester, NY: National Reading Conference, Inc.
- Resnick, D. P. (1981). Testing in America: A supportive environment. Phi Delta Kappan, 62(1), 625-628.
- Robinson, G. & Craves, J. (1989). Assessing and grading student achievement. Arlington, Virginia: Educational Research Service.
- Rumelhart, D. E. (1985). Toward an interactive model of reading. In H. Singer & R. B. Ruddell (Eds.), Theoretical Models and Processes of Reading (3rd ed.) (pp. 722-750). Newark, DE: International Reading Association.

- Sadoski, M., & Lee, S. (1986). Reading comprehension and miscue combination scores: Further analysis and comparisons. Reading Research and Instruction, 25(3), 160-167.
- Salmon-Cox, L. (1981). Teachers and standardized achievement tests: What's really happening? Phi Delta Kappa, 62, 631-634.
- Searls, E. (1988). What's the value of an IRI? Is it being used? Reading Horizons, 28, 92-101.
- Shanahan, T. (1983). A critique of P. A. Killgallon's study: A study of the relationships among certain pupil adjustments in reading situations. In L. Gentile, M. Kamil, and J. Blanchard. (Eds.) Reading Research Revisited (p. 581). Columbus, OH: Charles Merrill.
- Silvaroli, N. (1986). Classroom reading inventory. Dubuque, Iowa: Wm. C. Brown, Publ.
- Singer, H. (1985). A century of landmarks in reading research. In H. Singer & R. Ruddell (Eds.), Theoretical Models and Processes of Reading (3rd ed.) (pp. 60-94). Newark, DE: International Reading Association.
- Sipay, E. A. (1964). A comparison of standardized reading scores and functional reading levels. Reading Teacher, 17, 265-268.
- Smith, F. (1979). Reading without nonsense. New York: Teachers College Press.
- Smith, F. (1988). Understanding reading: A psycholinguistic analysis of reading and learning to read (4th ed.). Hillsdale, New Jersey: Lawrence Erlbaum, Associates.
- Stiggins, R. (1988). Revitalizing classroom assessment: The highest instructional priority. The Delta Kappan, 69(5), 363-368
- Thorndike, E. (1917). Reading as reasoning: A study of mistakes in paragraph reading. Journal of Educational Psychology, 8(6), 323-332.
- Tierney, R., & Pearson, P. (1983). Toward a composing model of reading. Language Arts, 60(5), 568-580.

- Tittle, C. (1989). Validity: Whose construction is it in the teaching and learning context? Educational Measurement Issues and Practice, 8(1), 8-13.
- Tyler, R. W. (1988). Evaluation for utilization. In J. Keeves (Ed.) Educational research, methodology and measurement: An international handbook (pp. 155-161). New York: Pergamon Press.
- Valencia, S. (1990). Alternative assessment: Separating the wheat from the chaff. The Reading Teacher, 44(1), 60.
- Valencia, S., & Pearson, P. D. (1987). Reading assessment: Time for change. The Reading Teacher, 40(8), 726-732.
- Valencia, S. W., & Wixon, K. (1991). Diagnostic Teaching. The Reading Teacher, 44(6), 420-423.
- Vincent, D., & Creswell, M. (1976). Reading tests in the classroom. London: N.F.E.R. Publishing Co.
- Weaver, C. (1988). Reading process and practice: From socio-psycholinguistics to whole language. Portsmouth: Heinemann.
- Weaver, C. (1990). Commentary: Phonics first advocates. Education Week, 23, 32.
- Weber, R. M. (1968-1969). The study of oral reading errors: A survey of the literature. Reading Research Quarterly, 4(1), 96-119.
- Wittrock, M., & Schultze, H. (1990). School and classroom evaluation of reading. Paper presented at the meeting of the W.O.R.D. Conference, Tacoma, WA.
- Wixon, K., Peters, C., Weber, E., & Roeber, E. (1989). New directions in statewide reading assessment Reading Teacher, 40(86-87), 749-755.
- Wood, N. (1989). Reading tests and reading assessment. Journal of Developmental Education, 13(2), 14-18.
- Woodley, J. W. (1988). Reading assessment from a whole language perspective. Denton, TX: University of Northern Texas. (ERIC Document Reproduction Service No. ED 296-309)
- Worthen, B. R., & Spandel, V. (1991). Putting the standardized test debate in perspective. Educational Leadership, 48(5), 65-69.

APPENDIX

DEMONSTRATION OF THE CG SCORING PROCEDURE

DEMONSTRATION OF THE CG SCORING PROCEDURE

MISCUE	SCORES	SCORES	EXPLANATION	MARKING MISCUES & SCALES
	C	G		How teacher marks miscues while listening to tape recording
<u>only</u> own	6/6	3/3	C: Self-corrects; 6 out of 6 points G: 3 out of 3 letters	"We can go to into the forest," said Cathy, and cut out our <u>own</u> tree."
<u>in</u> into	5/6	2/4	C: Same meaning; 5 out of 6 points G: 2 out of 4 letters	"What do you mean?" said Pete. "They don't let you just go into the forest and cut trees."
<u>permanent</u> permit	3/6	5/6	C: Same syntax; 3 out of 6 points G: 5 out of 6 letters	"Yes they do," said Cathy, "you have to buy a permit first. Then you can cut your own tree."
<u>tree</u> tree	5/6	3/4	C: Same meaning to message; 5 out of 6 points G: 4 out of 4 letters but additional letter	Next Saturday the whole family drove to the forest. <u>and</u> they found the ranger <u>(station)</u> and bought a permit. It cost a dollar. The rangers had marked off a <u>place</u> for people to cut trees.
<u>d</u> family	0/6	0/6	C: Default; 0 out of 6 points G: 0 out of 6 letters	"Look at that one," said Cathy. "It's perfect."
<u>and</u> they	5/6	0/4	C: Same meaning to message; 5 out of 6 points G: 0 out of 4 letters	"No, it's not," said Pete. "It's way too small."
<u>ranger</u> rangers	5/6	6/7	C: Same meaning to message; 5 out of 6 points G: 0 out of 7 letters	Dad gave Pete the saw. "I think Cathy is right," he said. "Anyway, her tree is <u>very</u> good shape."
<u>the</u> a	5/6	0/1	C: Same meaning; 5 out of 6 points G: 0 out of 1 letter	Pete cut <u>it</u> down. They loaded it onto the roof of the car and drove home. When they got home, Cathy's tree was too big to go through the door. They had to cut <u>about</u> a foot off the bottom.
<u>price</u> place	6/6	5/5	C: Self-corrects; 6 out of 6 points G: 5 out of 5 letters	
<u>That's</u> it's	5/6	2/3	C: Same meaning; 5 out of 6 points G: 2 out of 3 letters	
<u>a</u> the	5/6	0/3	C: Same meaning; 5 out of 6 point G: 0 out of 3 letters	
<u>the</u> it	6/6	2/2	C: Self-corrects; 6 out of 6 points G: 2 out of 2 letters	
<u>then</u> when	4/6	3/4	C: Relevant; 4 out of 6 points G: 3 out of 4 letters	
Totals	60/78	32/52	C: 60 out of 78 G: 32 out of 52	
Percents	77	62	C: score: 77% G: score: 62% CG: score: 70%	

- Context Scale (C)
- 0 = no attempt (default)
 - 1 = nonsense substitution or partial teacher gives word
 - 2 = substitution not in same syntactic class
 - 3 = substitution merely in same syntactic class (e.g. noun for noun)
 - 4 = substitution somewhat relevant to author's message and also in same syntactic class
 - 5 = substitution gives about the same meaning to author's message
 - 6 = self-correction
- Graphophonic Scale (G)
- 0 = no letter the same in substitution as in author's message
 - 1 = one letter the same in substitution as in author's word
 - 2 = two letter the same in substitution as in author's word
 - N = number of letters the same in substitution as in author's word
 - 1 = one or more extra letters