

EFFECTS OF A DIAGNOSTIC-PRESCRIPTIVE SYSTEM ON  
THE READING COMPREHENSION SCORES OF  
FOURTH AND FIFTH GRADE STUDENTS

THESIS

Submitted to the Graduate Committee of the  
Department of Curriculum and Instruction

Faculty of Education

State University College at Brockport

in Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Education

by

Mary Ann Grazioplena

State University College at Brockport

Brockport, New York

August, 1978

APPROVED BY:

F. Morrey White 9/1/78  
Graduate Advisor Date

Arthur E. Smith 9/1/78  
F. Morrey White 9/1/78  
Project Advisor Date

Joseph Jordan 9/1/78  
Graduate Director Date

## Abstract

The purpose of this quasi-experimental study was to investigate the effects of a diagnostic-prescriptive teaching system on the total reading comprehension scores of fourth and fifth grade students of varying achievements in reading. The study also examined which achievement level of students benefited the most from exposure to a diagnostic-prescriptive teaching system for reading. Three hypotheses dealing with method of instruction, achievement, and their interactions were tested. A sample population of 108 (54 males, 54 females) fourth and fifth grade students was used. The control group utilized various basal programs. The experimental group employed various basal programs that were augmented by a diagnostic-prescriptive system for reading (Houghton-Mifflin, IPMS). Instrumentation included the Stanford Diagnostic Reading Test (Levels I and II, Form W) as well as pretests, interim mastery tests, and posttests from the IPMS system. Significant differences and interactions were assessed by analysis of variance. There was no significant difference in the comprehension mean gain scores of both groups with respect to method of instruction. At both grade levels for both groups, there was a highly significant effect on comprehension mean gain scores with respect to achievement. Interaction effects were significant at the fourth grade level but not at the fifth grade level.

## Table of Contents

	Page
List of Tables . . . . .	iii
Chapter	
I. Statement of the Problem . . . . .	1
Purpose of the Study . . . . .	1
Questions of the Study . . . . .	1
Need for the Study . . . . .	2
Definition of Terms . . . . .	5
Limitations of the Study . . . . .	6
Summary . . . . .	6
II. Review of the Literature . . . . .	7
Purpose of the Study . . . . .	7
Criterion-Referenced Testing--An Overview . . . . .	7
Definition of Criterion-Referenced Testing . . . . .	9
History of Criterion-Referenced Testing . . . . .	10
Assumptions of Criterion-Referenced Testing . . . . .	11
Benefits and Limitations of Criterion-Referenced Testing . . . . .	14
Research and the Need for Further Research . . . . .	17
Summary . . . . .	19
III. Research Design . . . . .	20
Purpose of the Study . . . . .	20
Hypotheses . . . . .	20
Methodology . . . . .	21
Summary . . . . .	24
IV. Analysis of Data . . . . .	26
Purpose of the Study . . . . .	26
Analysis of Findings and Interpretation of Data . . . . .	26
Summary . . . . .	32
V. Conclusions and Implications . . . . .	34
Purpose of the Study . . . . .	34
Conclusions . . . . .	34
Implications for Classroom Practice . . . . .	37
Implications for Further Research . . . . .	38
Summary . . . . .	39
References . . . . .	41

## List of Tables

Table	Page
1. Comparison Data for Control and Experimental Groups . . . . .	21
2. Descriptive Statistics--Gain Scores of Experimental and Control Groups . . . . .	27
3. Preliminary Data for Analysis of Variance . . . . .	28
4. Fourth Grade Summary: Unweighted Means Analysis of 2 X 2 Factorial Analysis of Variance . . . . .	31
5. Fifth Grade Summary: Unweighted Means Analysis of 2 X 2 Factorial Analysis of Variance . . . . .	31
6. Diagnostic-Prescriptive Results . . . . .	33

## Chapter I

### Statement of the Problem

#### Purpose of the Study

The purpose of this study was to investigate the effects of a diagnostic-prescriptive teaching system on the total reading comprehension scores of fourth and fifth grade students. An auxiliary purpose was the determination of which achievement section of students (above average/average or average/below average) benefited the most from exposure to basal reading augmented by a diagnostic-prescriptive teaching system.

#### Questions of the Study

The study sought to answer the following questions:

1. Does a criterion-referenced testing system for reading used in conjunction with a basal-reader program enhance the total reading comprehension scores of fourth and fifth grade students?
2. Does an above average/average or average/below average achievement group of fourth and fifth grade students benefit the most from exposure to a basal-reader program augmented by a criterion-referenced system for reading?
3. What is the effect of the interaction between method of instruction and student achievement on the total comprehension scores of fourth and fifth grade students?

### Need for the Study

Pressures for increased efficiency in education have led to the search for new evaluation procedures which are more significant measurements of instructional outcomes than traditional standardized tests. Recently, there has been a call for measures which are relevant to a school system's program objectives and which can provide information about the merits of its programs. Criterion-referenced testing has gained recognition as an assessment procedure which meets this need.

Interest in criterion-referenced testing has been stimulated by such developments as the National Assessment of Educational Progress, by the expanded use of programmed instruction and computer-assisted instruction, and by an increasing preoccupation with educational accountability and performance contracting (Prescott, 1971). Educational assessment efforts have become a part of a larger movement in the United States to obtain dependable information about the performance of students (Elsner, 1973).

Criterion-referenced testing is an outgrowth of the mastery learning theory developed by Morrison more than forty years ago. By 1955, the influence of this philosophy had significantly waned even though Guttman (1944) and Tucker (1952) elaborated on the subject, and Flanagan (1951) called for a distinction to be made between standards of performance and norm-referenced measurement. Ebel (1962) extended this distinction and presented two schemes for developing tests whose scores could be interpreted objectively and

meaningfully without the use of norms. Glaser (1962) called for the specification of the type of behavior the individual is required to demonstrate with respect to the content. The standard (or criterion) against which a student's performance is compared is the behavior which defined each point along the achievement continuum, hence, the term criterion-referenced testing. Recently, the concept of mastery has been reintroduced into educational discussions as a corollary of various systems of individually prescribed instruction (Carroll, 1963; Bloom, 1968; Block, 1971, 1975). Criterion-referenced testing has been supported by some educators and it has been criticized by others.

While the literature reflects descriptive analyses of criterion-referenced testing, there is, however, a minimum of research based on experimental data. Elsner (1973) contended that the literature reveals limited research on the effects of test results on teacher and student behavior. In reference to the lack of investigations dealing with the significance of testing procedures, Elsner (1973) quotes Kirkland (1971) stating that, "only a few small scale and some peripheral empirical studies appear" (p. 730).

Several authors promote empirical research on various aspects of criterion-referenced testing. In their elaboration on skill management systems (categorically, criterion-referenced systems) Johnson and Pearson (1975) proposed that one could determine whether or not mastery of a set of subskills contributed to increased oral reading fluency or comprehension of written discourse and that

such a validation study seems critical to the whole notion of skill management systems. They further contended that until some basic research and evaluation of these systems is conducted, the skill management systems must be viewed only as one of many alternatives to the teaching of reading.

In adding impetus for evaluating criterion-referenced objectives, Bloom, Hastings, and Madaus (1971) suggested that these objectives be judged in relation to learner outcomes. The alignment between teachers and subject matter experts should be explored.

Hambleton and Novick (1973) stated that at this stage of development of a theory of criterion-referenced measurement, the establishment of proficiency levels is primarily a value judgment. Research might usefully be undertaken to provide guidelines for this judgment.

The findings of the present study contributed to the evaluation of one criterion-referenced system. During the 1971-72 school year, a K-12 language arts evaluation was conducted in the participating school district. The outcome of this evaluation emphasized the establishment of a comprehensive K-12 Language Arts/Reading program that promoted a diagnostic-prescriptive approach. As a result, the Language Arts Guide, together with critical objectives and scope and sequencing of skills, was developed. A criterion-referenced system for reading based on the Houghton-Mifflin Individual Pupil Monitoring System (IPMS) was developed during July-August, 1977 and implemented during the 1977-78 school year. The present investigation was part of this implementation.

### Definition of Terms

For the purposes of this study it was necessary to define several terms.

Criterion-referenced measurement is a means of identifying an individual's status with respect to an established standard of performance or criterion. It is a specific indication of the effectiveness of instruction (Popham, Husek, 1969; Block, 1971).

Diagnostic-Prescriptive Instruction is the assessment of academic skill development with the use of criterion-referenced tests and the prescription of instruction based on the learner's observed behavior.

A basal-reader program is a systematic, sequential teaching of reading skills through the use of specific reading series with their accompanying workbooks and worksheets. Supplemental materials including skill labs and language activities such as creative writing may also be used.

Total reading comprehension refers to a composite grade equivalent score received on the Stanford Diagnostic Reading Test Form W (Level I or II) that includes literal and inferential comprehension skills.

Above average/average students in reading are students scoring in the ninth through sixth stanine on subtests of the Stanford Diagnostic Reading Test Form W (Level I or II) administered in September, 1977.

Average/below average students in reading are students scoring in the fifth through first stanine on subtests of the Stanford Diagnostic Reading Test Form W (Level I or II) administered in September, 1977.

Method of instruction refers to the sole use of a basal reader program or to the use of a basal-reader program augmented by criterion-referenced testing and prescriptive teaching.

#### Limitations of the Study

The present study was limited to investigating the effects of one diagnostic-prescriptive system on the total comprehension scores of fourth and fifth grade students following 22 weeks of instruction. The sample population involved 108 students and was limited to a specific school environment in a middle-class suburban setting. The methods of reading instruction utilized in the investigation were limited to specific basal series and the Houghton-Mifflin Individual Pupil Monitoring System (IPMS).

#### Summary

This study questioned the effects of a diagnostic-prescriptive system on the reading comprehension scores of fourth and fifth grade students. Researchers have indicated a need for studies dealing with criterion-referenced testing (Elsner, 1973; Johnson and Pearson, 1975; Hambleton and Novick, 1973).

The study investigated the reading comprehension scores of fourth and fifth grade students using a basal reader program and a criterion-referenced testing system for reading.

## Chapter II

### Review of the Literature

#### Purpose of the Study

This study examined the effects of a diagnostic-prescriptive teaching system on the reading comprehension scores of fourth and fifth grade students. Effects of the diagnostic-prescriptive system with respect to achievement level were also investigated.

The research related to this study has been divided into the following six categories:

Criterion-Referenced Testing--An Overview

Definition of Criterion-Referenced Testing

History of Criterion-Referenced Testing

Assumptions of Criterion-Referenced Testing

Benefits and Limitations of Criterion-Referenced Testing

Research and the Need for Further Research in Criterion-Referenced Testing

#### Criterion-Referenced Testing--An Overview

Educators have recognized that effective reading instruction must focus on each individual student's strengths and weaknesses. One of the most recent results of the search for better ways to individualize reading instruction effectively is the diagnostic-prescriptive, criterion-referenced approach. Publishers are

presently promoting criterion-referenced reading systems in an effort to make instruction attuned to individual needs a reality.

Other factors have influenced recent interest in criterion-referenced testing. The National Education Association (1975) reported that at least three factors have contributed to the emergence of the criterion-referenced testing concept: first, is a strong and rising dissatisfaction with tests in general; second, is the inadequacy of traditional tests for diagnostic and instructional purposes. Third, there is some clamor for evaluating instruction and teachers as part of the accountability movement.

Mather (1977) contended that there appear to be two reasons for the influx of skill management systems. The first reason is the increased public interest in reading scores, and the second is that teachers are being made accountable for the learning of their students. The Florida State Education Department (1977) noted that recent legislation has dealt with accountability and has supported diagnostic-prescriptive techniques.

Mione (1977) stated that the accountability movement "Back to Basics" proponents, decline of national test scores, rising interest in applied performance testing along with a concern about the quality of a high school diploma have led, since January of 1976, to thirty-three states giving consideration to the establishment of some form of minimal competency testing.

### Definition of Criterion-Referenced Testing

A review of the literature reveals a myriad of interpretations of what constitutes a criterion-referenced test. The term "criterion-referenced system" is used interchangeably with "diagnostic-prescriptive system," "skill management system," and "objective-based system." However, some basic notions prevail.

Millman (1972) noted that criterion-referenced measurement meaning comes from a comparison of the student's performance relative to the skills being assessed by test questions. Popham and Husek (1969) reported that criterion-referenced tests are used to identify an individual's status with respect to an established standard of performance. Mione (1977) said that criterion-referenced tests indicate what an individual can or cannot do with certain specific requirements specified as a criterion score which does not depend on the scores the other students obtain on the test.

The Florida State Education Department (1977) extended the criterion-referenced testing notion by indicating that diagnostic-prescriptive reading instruction is a method whereby each student's major reading strengths and weaknesses are assessed (diagnosed) by both formal and informal means. Based on this assessment, instruction (treatment) is prescribed which enables students to work on their own specific needs at their own rate in a variety of appropriate materials.

### History of Criterion-Referenced Testing

Diagnostic-prescriptive instruction is not a new concept. More than forty years ago, Morrison developed a method of teaching based on the mastery of "adaptations" of understanding, appreciation or ability. Guttman (1944) mentioned an ideal test whose items were tied to a criterion. Tucker (1955) elaborated on this view.

Flanagan (cited in Lindquist, 1951) distinguished between "standards of performance" and "norm-performance." He defined a "standard of performance" as a minimum goal that an individual should attain. A "norm performance" was the present average attainment with respect to a specific population.

During the 1960's, advocacy for adopting criterion-referenced tests from an educational or philosophical point of view proliferated in spite of the reservations of classical measurement theories. Ebel (1962) outlined two models for developing tests whose scores could be interpreted without the use of norms. Glaser (cited in Gagne, 1962) initiated the term criterion-referenced testing. He defined the standard (or criterion) against which a student's performance is compared as the behavior which is defined along each point of an achievement continuum. Coulson and Cogswell (1965) emphasized the need for criterion-referenced tests with individualized instruction. Glaser and Cox (1968) also stressed a similar need when it is important to differentiate between those who have mastered the objective and those who have not.

Support for criterion-referenced testing flourished in the 1970's. Millman (1970) advocated the use of a criterion-referenced marking system for the reporting of student progress. Block (1972) indicated that the maintenance of different standards was likely to maximize student learning depending upon the criterion by which the learning is operationalized. Heines (1975) concluded that criterion-referenced testing was perhaps the most significant development in the evaluation of instruction since norm-referenced testing was implemented on a large scale in the early 1900's.

#### Assumptions of Criterion-Referenced Testing

Prescott (1971) contended that there are three assumptions underlying the criterion-referenced approach. The first upholds that mastery is a reasonable criterion. The problem of determining mastery levels on a criterion-referenced test is a complex issue, and it has received much attention in the research literature. It has been proposed that a criterion-referenced measure is related to a student's acquisition of knowledge along a continuum from no proficiency, or mastery, to perfect mastery performance (Glaser, 1963, 1971; cited in Thorndike, 1971; Nitko, 1971).

In presenting a model based upon four tenets of mastery theory, Emrick (1971, p. 7) listed:

1. Learning of fundamental skills can be considered all or none.
2. Each item response on a single skill test represents an unbiased sample of the examinee's true mastery status.

3. Measurement error occurring on the test can be of only one type for each examinee.
4. Through practical and theoretical considerations of evaluation error costs and item error characteristics, an optimal mastery criterion can be calculated.

Adams (1974) also concluded that the learner gives a correct response if he or she is in the mastery state. The learner gives an incorrect response if he or she is in a non-mastery state. However, error of testing occurs when learner performance on a test item does not reflect true competence. He debated Emrick's third principle by stating that Type I Error occurs when the learner is in a non-mastery state but gives a correct response. Type II Error occurs when the learner is in a mastery state but gives an incorrect response.

Prescott (1971) contended that absolute mastery is a basic weakness of criterion-referenced test interpretation, one that severely limits its applicability because it is not universally true. Mione (1977) summarized by stating that no valid and reliable data or conclusions have been reached on the issue of mastery levels.

The second assumption underlying criterion-referenced testing outlined by Prescott (1971) was that each item in a test has inherent worth. Thompson and Dziuban (1973) maintained that the most fundamental weakness of skill management systems is that all of the behavioral objectives are not related to the skill of reading, or in some cases, the relationship may be so minimal that time spent learning the subskill may be better applied directly to skills with stronger relationships to reading. Johnson and Pearson (1975)

raised the question of whether mastery of each skill transfers to something called "reading."

The final assumption of criterion-referenced testing proposed by Prescott (1971) was a definite hierarchy of skills and knowledge exists in any skill or content area. He maintained that the criterion-referenced approach is of little value unless the assumption is made that mastery of one skill is essential for mastering another skill of a somewhat similar character at a higher level of difficulty or complexity.

Proponents of this assumption such as Otto and Samuels (cited in Mather, 1977) present reading as a complex task with a variety of skills and subskills which must be mastered before one can read. Davis (1971) concluded that reading comprehension is composed of separate skills and abilities. The hypothesis that a natural hierarchical sequence of learning skills exists and can be replicated by the task analyst was upheld by Resnick, et al. (1973), Wang (1973), and Wang, et al. (1971).

Opponents of the hierarchy of skills issue such as Smith (1971) believe that reading is a holistic or total process that cannot easily be divided into sets of skills and subskills. Johnson and Pearson (1975) indicated that the fragmentation of reading inherent in all skills monitoring systems denies the nature of language. They elaborated that the whole notion of sequence or hierarchy of skills is, at best, a pedagogical convenience and there is little evidence to support the existence of separate skills that

can be placed into a sequential hierarchy. Thompson and Dziuban (1973) stated that the assumption in criterion-referenced testing that there exists an empirically derived hierarchy of reading skills is unsubstantiated. The assumption is faulty since no evidence is available to support the contention. Chall (1977) maintained that the most far-reaching theoretical question on reading comprehension is whether it is a general skill or whether it is made up of a number of specific, identifiable skills. She noted that these two theories of reading comprehension can lead to considerable different approaches to testing and teaching.

#### Benefits and Limitations of Criterion- Referenced Testing

The values of criterion-referenced testing are heralded by some educators as the key to individualization. Mione (1977) commented that criterion-referenced testing fits well with programs emphasizing individualized instruction. He further reported that some educators argue that competition can be eliminated and student learning greatly facilitated by criterion-referenced testing. Thompson and Dziuban (1973) stated that the nature of criterion-referenced tests makes them of special interest to reading teachers, since reading teachers have been in the forefront of individualizing their instruction.

Other benefits of criterion-referenced testing have been noted. Millman (1972) indicated that decisions about future

instruction should be made on the basis of what a child can do now. Criterion-referenced information seems indispensable to making such instructional decisions. Thompson and Dziuban (1973) contended that the principal strength of criterion-referenced tests is the focus on specific skill behaviors. Knowledge of children's specific skill needs will help the teacher in prescribing appropriate instructional strategies. Wilson (1975) proposed that the value of task analysis for reading comprehension diagnosis is increased by the fact that we do not all learn to comprehend through the same set of subskills. Task analysis provides a highly suitable strategy for assessing the unique learning styles of each pupil.

In an evaluation study of the criterion-referenced Wisconsin Design, White and Damos (1977) reported that the Design provided follow-through for instruction from year to year; facilitated awareness of individual differences; provided strong organization; and helped screen out advanced students and those with learning problems. Teachers became more aware of different learning styles of children.

An analysis of the related literature also reveals that criterion-referenced testing has met with much criticism. Ebel (1971, p. 287) delineated major limitations of criterion-referenced tests:

1. They do not tell us all we need to know about achievement.
2. They are difficult to obtain on any sound basis.
3. They are necessary for only a small fraction of important educational achievements.

He also pointed out the weaknesses of behavioral objectives, the problem of defining mastery scores, and the difficulty of test construction.

Boehm (1973) mentioned that there is a lack of accepted theory and procedures for determining test reliability and validity. Ten Brink (1974) disputed the small number of items measuring each objective. He also noted that little variance in criterion-referenced scores makes reliability estimates and item analysis difficult.

In reviewing the disadvantages of criterion-referenced tests, Johnson and Pearson (1975, p. 758) discussed: (a) their psycholinguistic naiveté, (b) their assembly-line underpinnings, (c) their concern for skill at the expense of interest, (d) their advocacy of sequencing separable reading skills, (e) the validity of their assessment instruments, and (f) the very notion of mastery itself.

Mione (1977) contended that the teacher may feel locked into stated objectives when a new, equally important or more important objective may come to mind. Teaching for a criterion-referenced test might become a sole aim for the instructor and students then could view mastery as the primary focus at the expense of retention and transfer of learning.

Chall (1977) found that in some schools where a criterion-referenced reading system has been adopted, teachers have complained that with all the testing, checking, and exercising, the children have no time for reading books. She felt that this was a serious matter for it has been agreed upon by all who study and teach

reading that mature reading is not possible without the reading of books.

General negative comments concerning the implementation of the Wisconsin Design were reported by White and Damos (1977). Lack of time for planning, testing, teaching and record keeping were major problems. Lack of materials and sufficient personnel to carry out steps and keep skill groups at workable sizes was also cited.

#### Research and the Need for Further Research

There is a limited amount of empirical research in the area of criterion-referenced testing. Athey and O'Reilly (1975) found the hypothesis that criterion-referenced tests would prove to be more sensitive measures of reading improvement over a length of time compared to norm-referenced tests, was partially confirmed at the fifth grade level. Some researchers determined what skills are prerequisites to competence in reading (McNeil, 1974; Carducci, Bolchazy, 1975). Others have investigated the functional word attack ability of students instructed with an objective-based approach (Kurth, 1975). Smead (1977) noted previous studies that have investigated whether learning proceeds in small steps or depends on massive general experience. He also cited that some studies have been directed to the validity of hierarchical orderings of specific skill prerequisites.

The evaluation study of White and Damos (1977) includes the results of questions concerning the comprehension element of the

Wisconsin Design. Of the 24 schools implementing the comprehension element, only 5 rated their success as excellent, while 10 rated success as good, 7 as moderate, 3 as fair and 1 as poor.

Several authors have advocated further research in criterion-referenced testing and diagnostic-prescriptive systems. Elsner (1973) clarified that implications for further research point to continued use of criterion-referenced assessment as an alternative and more in-depth study of classroom uses. The National Educational Association (1975) stated that many objective-referenced tests have not been extensively studied. Even when investigated, frequently only a handful of students are involved. Criterion-referenced testing should be thoroughly field tested.

Ewing and Brecht (1977) noted that the legitimacy of the task-analysis approach to diagnostic-prescriptive teaching has not been conclusively determined through research, and much of the research reported has been criticized for methodological reasons. Educators recognize the need for better designed research to determine the usefulness of task analysis.

The pressing need for research to investigate the effectiveness of the criterion-referenced approach was stated by White (1977). In reference to skill management systems, Mather (1977) reported that research is limited and that studies must be made to determine whether one management system is better than any other reading program. Smead (1977) indicated that the major tenets of task

analysis must be subjected to further research and that studies using control groups taught the same content according to different approaches are desperately needed.

#### Summary

Recently, criterion-referenced testing has been receiving much attention in educational discussions. Even though various interpretations of criterion-referenced testing have been reported, basic understandings of what constitutes a criterion-referenced test as an integral part of diagnostic-prescriptive teaching exist. There are several assumptions underlying criterion-referenced systems with proponents and opponents of these tenets. Criterion-referenced testing has been supported by some educators and it has been criticized by others. There has been a minimum of research investigating various aspects of criterion-referenced testing, and the literature indicates the need for further empirical research in this area.

## Chapter III

### Research Design

#### Purpose of the Study

The purpose of this quasi-experimental study was to investigate the effects of a diagnostic-prescriptive teaching system on the total reading comprehension scores of fourth and fifth grade students of varying achievements in reading. A secondary purpose was an assessment of which achievement section (above average/average or average/below average) benefited the most from exposure to basal reading and a diagnostic-prescriptive teaching system as evidenced in higher comprehension scores.

#### Hypotheses

The null hypotheses investigated in this study were:

1. There are no significant main effects for method of instruction with respect to the total comprehension gain scores of fourth and fifth grade students.
2. There are no significant main effects of achievement with respect to the total comprehension gain scores of fourth and fifth grade students.
3. There is no significant interaction between method of instruction and student achievement with respect to the total comprehension gain scores of fourth and fifth grade students.

## Methodology

### Selection of Subjects

A sample population of 108 students, 54 males and 54 females, was drawn from fourth and fifth grade classes in a middle-income, suburban school district. They comprised four heterogeneously grouped classes, one teacher to each class. Two classes were randomly assigned to the control group, and two classes were randomly assigned to the experimental group. Comparison data obtained from permanent records and Spring 1977 Metropolitan Achievement Test Scores are presented in Table 1.

Table 1

Comparison Data for Control and Experimental Groups

Control		Experimental	
$\bar{I}Q$ (Otis Lennon) =	113	$\bar{I}Q$ (Otis Lennon) =	113
$\bar{X}$ Comprehension =	5.4	$\bar{X}$ Comprehension =	5.1
$\bar{X}$ Total Reading =	5.4	$\bar{X}$ Total Reading =	5.3
s.d. Total Reading =	1.61	s.d. Total Reading =	1.43
Sex M = 28	F = 26	Sex M = 26	F = 28
$\bar{X}$ Age =	9.9	$\bar{X}$ Age =	9.8
No. of 4th graders =	30	No. of 4th graders =	25
No. of 5th graders =	24	No. of 5th graders =	29

An examination of these data revealed distinct similarities for both groups in intelligence, prior reading achievement, age, and sex. Therefore, the two groups were judged to be comparable for the purpose of this study.

### Instrumentation

The Stanford Diagnostic Reading Test (Levels I and II) Form W was administered to all subjects during the first week of school (Fall, 1977). Determination of which students received Level I and which students received Level II was based on age, reading performance on the previous Metropolitan Reading Achievement Test, and on the basal text that was used for instruction the previous year.

In the control group and in the experimental group, the students were sectioned according to stanine scores received on subtests of the Stanford Diagnostic Reading Test. Above average/average, and average/below average achievement sections were formed for reading instruction.

The Stanford Diagnostic Reading Test was chosen because it renders a total comprehension score and because it is consistently used as a diagnostic tool (pretest) by the teachers involved.

A criterion-referenced testing system for reading was used (the comprehension element of the Houghton-Mifflin Individual Pupil Monitoring System (IPMS)). A group of the participating district's teachers concluded that the objectives of this commercially prepared system satisfactorily correlated with many of the district's critical reading objectives.

## Procedures

Each group received reading instruction once a day, for forty-five minutes, for twenty-two weeks. The length of the treatment period was determined by the amount of time that was needed for all students in the experimental group to master all or nearly all of the criterion-referenced objectives. The control group was instructed with basal reading programs using the series published by Houghton-Mifflin, Lippincott, Merrill, Ginn, and Harper-Row Companies. Supplemental reading materials and activities were also employed.

The experimental group was instructed with basal series published by Houghton-Mifflin, Scott Foresman, and Ginn augmented by the Houghton-Mifflin IPMS system. Initially, the IPMS criterion-referenced pretest (Form B) was administered. Results from this testing were used to chart each child's reading profile of skills--strengths and weaknesses. On the basis of this diagnosis, whole-class, reading group, and individual prescriptive teaching of specific skills took place using the basal series (cross-referenced with the criterion-referenced testing objectives), and other materials correlated with the criterion-referenced objectives. Continued monitoring of student progress was made through the use of IPMS interim mastery tests (Form B Tests) and through extensive record-keeping. The testing situations were regarded as routine classroom interaction. Throughout the treatment period, the teachers of the experimental group made an effort to be cognizant of emphasizing specific skill development.

Although the groups were housed in separate areas of the same building, similar daily programs and instructional time periods were maintained for both groups. Six reading series were implemented to control any advantage that one series may give to either of the groups. Availability of supplementary materials was equal for both groups.

At the conclusion of the treatment period, the IPMS criterion-referenced posttest (Form A) was administered to the experimental group. Both groups were retested with the comprehension section of the Stanford Diagnostic Reading Test (Levels I and II), Form X.

#### Statistical Analysis

Significant differences in total comprehension scores with respect to method of instruction and achievement were assessed by analysis of variance. Significant interaction between method of instruction and student achievement with respect to total comprehension was also assessed by analysis of variance. Separate analyses of variance were conducted for fourth and fifth grade students.

#### Summary

A quasi-experimental research design was employed in this study. Three hypotheses dealing with method of instruction, achievement, and their interactions were tested. A sample population of 108 students was used. Intelligence, achievement, age,

and sex variables were controlled. Significant differences and significant interactions were assessed by analysis of variance.

## Chapter IV

### Analysis of Data

#### Purpose of the Study

The purpose of this study was twofold. First, the study investigated the effects of a diagnostic-prescriptive teaching system on the total reading comprehension scores of fourth and fifth grade students. Second, the study determined which achievement level of students benefited the most from exposure to a basal reading program enhanced by a diagnostic-prescriptive teaching system for reading.

#### Analysis of Findings and Interpretation of Data

Initially, descriptive statistics techniques were employed. The total mean gain on the Stanford Diagnostic Reading Test for the experimental group was .4. The total mean gain for the control group was .6. It appeared from this data that the basal group outperformed the diagnostic-prescriptive group (see Table 2).

Preliminary data to conduct the analysis of variance were computed. It appeared from this data that the above average/average fourth grade students performed better in the diagnostic-prescriptive program than in the basal program. However, it appeared that the above average/average fifth grade students performed better in the basal program than in the diagnostic-prescriptive program (see Table 3).

Table 2  
 Descriptive Statistics--Gain Scores  
 of Experimental and Control Groups  
 (Except for N, numerals represent  
 grade equivalent scores)

Experimental Group			
Test	N	Total Gain	$\bar{X}$ Gain
Post	55	306.8	5.6
Pre	55	286.4	5.2 .4 Gain
Control Group			
Test	N	Total Gain	$\bar{X}$ Gain
Post	54	319.6	6.0
Pre	54	286.3	5.4 .6 Gain

Table 3  
Preliminary Data for Analysis of Variance

Fourth Grade Students		
	Basal	Diagnostic-prescriptive
Above Average / Average	$\bar{X}$ Gain = $-.577$ s.d. = 1.510	$\bar{X}$ Gain = $.588$ s.d. = $.756$
Average / Below Average	$\bar{X}$ Gain = $.160$ s.d. = $.759$	$\bar{X}$ Gain = $.492$ s.d. = $.666$
Fifth Grade Students		
	Basal	Diagnostic-prescriptive
Above Average / Average	$\bar{X}$ Gain = $.160$ s.d. = 1.560	$\bar{X}$ Gain = $-.106$ s.d. = 1.031
Average / Below Average	$\bar{X}$ Gain = 1.433 s.d. = 1.223	$\bar{X}$ Gain = $.990$ s.d. = $.692$

### Interpretations of Analysis of Variance

The sample for this study consisted of 108 students from four classrooms in a middle-class suburban school district. The relative effects of program and achievement variables as well as their interactions were examined by an unweighted means analysis of a 2 X 2 factorial design analysis of variance. Separate analysis of variance procedures were conducted for fourth and fifth grade students. The Litton Monroe 1860 Calculator was used.

Hypothesis 1 - Method of Instruction. The data failed to reject the hypothesis that there are no statistically significant main effects for method of instruction with respect to total comprehension gain scores of fourth and fifth grade students at the .05 level of significance. As shown in Tables 4 and 5, there was no overall significant difference between the gain scores of the basal and the diagnostic-prescriptive groups. However, the data for the fifth grade students revealed that there was a tendency for the basal group to score higher than the diagnostic-prescriptive group but this tendency was not statistically significant (see Tables 4 and 5).

Hypothesis 2 - Achievement Level. The hypothesis that there are no significant main effects of achievement with respect to the total comprehension gain scores of fourth and fifth grade students was rejected at the .01 level of significance. Reference to Tables 4 and 5 indicates that there was a highly significant effect with respect to achievement. The average/below average fourth and

fifth grade students gained significantly more than the above average/average achievement group. This finding applied to both the experimental and control groups (see Tables 4 and 5).

Hypothesis 3 - Interaction of Method of Instruction and Achievement. The hypothesis that there is no significant interaction between method of instruction and student achievement with respect to total comprehension gain scores of fourth grade students only was rejected at the .001 level of significance. Table 4 shows that the interaction of achievement and program was significant indicating that the above average/average fourth grade students performed better in the diagnostic-prescriptive program while the average/below average fourth grade students performed better in the basal program.

The findings failed to reject the hypothesis that there is no significant interaction between method of instruction and student achievement with respect to total comprehension gain scores of fifth grade students only at the .05 level of significance. Data from Table 5 indicate that there was no significant interaction of program and achievement for fifth grade students (see Tables 4 and 5).

Table 4  
 Fourth Grade Summary: Unweighted Means Analysis of  
 2 X 2 Factorial Analysis of Variance

Source	DF	SS	MS	F
Program	1	1.003	1.003	1.268
Achievement	1	8.182	8.182	10.343**
Interaction	1	10.279	10.279	12.992***
Error	53	41.958	0.791	
Total	56			

\*\* $p < .01$

\*\*\* $p < .001$

Table 5  
 Fifth Grade Summary: Unweighted Means Analysis of  
 2 X 2 Factorial Analysis of Variance

Source	DF	SS	MS	F
Program	1	1.537	1.537	1.121
Achievement	1	16.872	16.872	12.315***
Interaction	1	0.096	0.096	0.070
Error	48	65.795	1.370	
Total	51			

\*\*\* $p < .001$

### Diagnostic-Prescriptive Analysis

During the treatment period, locally developed criterion-referenced testing pretests and posttests were administered. Results from these data indicate that the experimental group gained a total of 52 percentage points. Since the criterion-referenced objectives had been correlated with the local objectives, considerable progress was made in meeting the objectives of the participating district. Data from the criterion-referenced results are found in Table 6.

Table 6

Diagnostic-Prescriptive Results

Experimental Group	IQ	Metropolitan Achievement Test	Post		CRT	
			Post	CRT	Post	CRT
1	Average = 115 Range = 93-133	Average = 5.7* Range = 4.3-9.0	Average = 69** Range = 25-100	Average = 87** Range = 67-100		
2	Average = 110 Range = 92-123	Average = 4.9 Range = 2.4-4.6	Average = 59 Range = 8-92	Average = 73 Range = 34-100		

\*Grade Equivalent Scores

\*\*Percentage Points

Summary

The data reveal that there was no significant difference in the total comprehension mean gain scores of the experimental and control groups with respect to method of instruction. At the fourth and fifth grade levels for both programs, there was a highly significant effect on total comprehension scores with respect to achievement. The interaction of program and achievement was significant at the fourth grade level but not at the fifth grade level.

## Chapter V

### Conclusions and Implications

#### Purpose of the Study

This study was designed to investigate three hypotheses concerning the effects of a diagnostic-prescriptive teaching system on the total reading comprehension scores of fourth and fifth grade students. The variables of program and achievement level were investigated.

#### Conclusions

The results of this study may be summarized as follows:

1. There were no significant main effects for method of instruction on the total reading comprehension gain scores of fourth and fifth grade students.
2. There were significant main effects of achievement on the total reading comprehension gain scores of fourth and fifth grade students.
3. At the fourth grade level, there was a significant interaction between method of instruction and student achievement on total reading comprehension gain scores. At the fifth grade level, there was no significant interaction between method of instruction and student achievement on the total reading comprehension gain scores.

Program

It was concluded that the total reading comprehension gain scores of fourth and fifth grade students were not significantly enhanced by a diagnostic-prescriptive system for reading. One possible explanation for this finding is that the participating teachers were implementing a new program. They may have been unable to make the most advantageous use of the diagnostic-prescriptive program because of unfamiliarity. Also, in the experimental group, the time spent on testing procedures reduced actual teaching time by a minimum of nine hours.

At the fifth grade level, there was a tendency for students to perform better in the basal program than in the diagnostic-prescriptive program. This conclusion was the first evidence in the study that there was a "ceiling effect" with respect to specific skills instruction at this level. Even though this finding was not statistically significant, fifth grade students seemed to respond more positively to a "global" approach to the teaching of reading than to a specific skills approach.

It should be noted that both the experimental and control groups gained in comprehension scores as a result of the instruction given. Ruling out the Hawthorne effect, the students in the experimental group were not handicapped. They did not fall behind in their expected reading comprehension growth. Therefore, the researcher concluded that a diagnostic-prescriptive teaching system for reading comprehension can be introduced in a classroom without deleterious effects.

### Achievement Level

In reference to both the experimental and control groups, it was found that the average/below average fourth and fifth grade students gained significantly more than the above average/average achievement group. This conclusion could have been reasonably assumed since it is easier to raise the scores of students performing less well than to raise the scores of students performing well. In previous years, the lower achieving students concentrated on basic word attack and comprehension skills. They would not have had much exposure to the comprehension skills outlined in the basal series and diagnostic-prescriptive system that were used in the study. These students would show a more distinct positive response to instruction in these skills than the higher achieving students.

This significant difference in the gains of the lower achieving students may be a statistical artifact. It may be referred to as regression toward the mean. The lower achieving students have the greatest opportunity to make more significant gains than the higher achieving students.

### Program and Achievement Level

At the fourth grade level, there was a significant interaction between method of instruction and achievement. The above average/average fourth grade students performed better in the diagnostic-prescriptive program. Previously, the above average/average students had mastered basic word attack and comprehension

skills and would have been introduced to the comprehension skills similar to the skills concentrated on in the IPMS system. Therefore, they would respond well to the diagnostic-prescriptive system because of prior exposure.

However, the average/below average fourth grade students performed better in the basal program. It is possible that this program met the needs of these students more because it was a more structured program than the diagnostic-prescriptive system. The average/below average fourth grade students had received prior basal instruction and probably were comfortable with this approach. Stated conversely, while the intrusion of the diagnostic-prescriptive system did not impede the comprehension gain scores of these students, it did not significantly enhance them either.

There were no significant interaction effects between method of instruction and achievement for the fifth grade students. The results of this study suggest that by the time students reached this grade level, they would not benefit significantly from a specific skills system although such a system would not hinder their progress.

#### Implications for Classroom Practice

Students should benefit from practical applications of this study. Teachers should be aware that the use of a diagnostic-prescriptive teaching system does not impair the comprehension gain scores of fourth and fifth grade students. Recommendations

concerning placement into a basal program and into a diagnostic-prescriptive system follow from the findings of this study. At the fourth grade level, it is advisable to place above average/average students in a basal series augmented by a diagnostic-prescriptive system. The average/below average fourth grade students should be placed in a basal series program without a diagnostic-prescriptive system.

At the fifth grade level, students can be placed in either a basal series program alone or in a basal series program used in conjunction with a diagnostic-prescriptive system. These students should show gains in comprehension in either placement. However, since there was a slight trend for the fifth grade students to perform better in the basal program, their placement into this program may be more advantageous.

#### Implications for Further Research

There are several follow-up studies suggested by this study. Since the Stanford Diagnostic Reading Test (Level I) that was used in the study did not give separate grade equivalent scores for literal and inferential comprehension, a total reading comprehension score was used. It would be beneficial to determine the effects of a diagnostic-prescriptive system on the literal and inferential comprehension scores of fourth and fifth grade students. A measurement instrument that lent itself to this type of analysis would have to be utilized.

The effects of a diagnostic-prescriptive system on the comprehension scores of students with respect to variables other than program and achievement could be examined. Sex, age, and grade level could be investigated. In the participating school where the present study was conducted, the same diagnostic-prescriptive system was utilized with high achieving third grade students. Statistical analysis of these students' scores would provide comparison data for the present study. Further placement recommendations for this grade level could be made.

The effects of a diagnostic-prescriptive system on other strands of the reading process may be investigated. Effects concerning word attack, vocabulary, and study skills could be determined.

Other studies similar to the present investigation could be designed to research the effects of other commercially or locally prepared diagnostic-prescriptive systems on the reading comprehension of fourth and fifth grade students. Results obtained would provide information for comparing the effects of different diagnostic-prescriptive systems.

#### Summary

The results of this study demonstrate that total reading comprehension scores of fourth and fifth grade students are not hindered by a diagnostic-prescriptive teaching system.

Achievement had a highly significant effect on total comprehension scores. Previous exposure and statistical regression may account for this effect.

Program and achievement level interacted to affect total comprehension scores at the fourth grade level. Prior experience and program structure may have affected the results. There was no interaction effect at the fifth grade level.

Implications for classroom practice mainly included recommendations for placement in a reading program.

Further research indicates an examination of the effects of a diagnostic-prescriptive system on literal and inferential comprehension scores. Sex, age, and grade level variables could be investigated. The effects of a diagnostic-prescriptive system on other aspects of the reading process could be researched. It would be beneficial to demonstrate the effects of other diagnostic-prescriptive systems on comprehension scores.

## References

## References

- Adams, E. N. On scoring a mastery learning control test. Journal of Computer-Based Instruction, 1974, 2, 50-58.
- Athey, I., & O'Reilly, R. P. A criterion-referenced model for assessing growth in reading. Paper presented at the Annual Meeting of the National Council for Measurement in Education, Washington, D.C., March 1975. (ERIC Document Reproduction Service No. ED 105 422)
- Block, J. H. Criterion-referenced measurements: potential. School Review, 1971, 79, 289-298.
- Block, J. Student evaluation: toward the setting of rational, criterion-referenced performance standards. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Ill., April 1972. (ERIC Document Reproduction Service No. ED 065 605)
- Block, J. H., & Anderson, W. Mastery learning in classroom instruction. New York: Macmillan, 1975.
- Bloom, B. S. Learning for mastery. Evaluation Comment, 1968, 1, 60-65.
- Bloom, B. S., Hastings, T. J., & Madaus, G. F. Handbook on formative and summative evaluation of student learning. New York: McGraw Hill, 1971.
- Boehm, A. E. Criterion-referenced assessment for the teacher. Teachers College Record, 1973, 75, 117-125.
- Carducci-Bolchazy, M. How necessary are certain skills for reading competence? Paper presented at the National Reading Reference St. Petersburg, Fla., December 1975.
- Carroll, J. B. A model of school learning. Teachers College Record, 1963, 64, 723-733.
- Chall, J. S. Reading 1967-1977: A Decade of Change and Promise. Bloomington, Ind.: Phi Delta Kappa Educational Foundation, 1977. (ERIC Document Reproduction Service No. ED 146 556)
- Coulson, J., & Cogswell, J. F. Effects of individualized instruction on testing. Journal of Educational Measurement, 1965, 2, 59-64.

- Davis, F. B. Mastery test scores. Paper presented at the Annual Meeting of the American Educational Research Association, New York, February 1971. (ERIC Document Reproduction Service No. ED 060 134)
- Ebel, R. L. Content standard test scores. Educational and Psychological Measurement, 1962, 22, 15-25.
- Ebel, R. L. Criterion-referenced measurements: limitations. School Review, 1971, 79, 282-288.
- Elsner, P. J. Criterion-referenced assessment and its classroom uses as viewed by teachers. School Science and Mathematics, 1973, 73, 730-735.
- Emrick, J. A. An evaluation model for mastery testing. Journal of Educational Measurement, 1971, 8, 321-326.
- Ewing, N., & Brecht, R. Diagnostic/prescriptive instruction: a reconsideration of some issues. The Journal of Special Education, 1977, 11, 323-327.
- Florida State Department of Education. Diagnostic prescriptive reading instruction. Washington, D.C.: Office of Education Right to Read Program, 1977. (ERIC Document Reproduction Service No. ED 149 310)
- Gagne, R. (Ed.). Psychological principles in systems development. New York: Holt, Rinehart, and Winston, 1962.
- Glaser, R. Instructional technology and the measurement of learning outcomes. American Psychologist, 1963, 18, 519-521.
- Glaser, R., & Cox, R. Criterion-referenced testing for the measurement of educational outcomes. In R. A. Weisgerber (Ed.), Instructional process and media innovation. Chicago: Rand McNally, 1968.
- Guttman, L. A basis for scaling qualitative ideas. American Sociological Review, 1944, 9, 139-150.
- Hambleton, R. K., & Novick, M. R. Toward an integration of theory and method for criterion-referenced tests. Journal of Educational Measurement, 1973, 10, 159-170.
- Heines, J. M. An examination of the literature on criterion-referenced and computer-assisted testing, 1975. (ERIC Document Reproduction Service No. ED 116 633)

- Johnson, D. D., & Pearson, P. D. Skills management systems: a critique. Reading Teacher, 1975, 28, 757-764.
- Kirkland, M. C. The effects of tests on students and schools. Review of Educational Research, 1971, 41, 351-368.
- Kurth, R. J. Evaluation of an objective-based curriculum in word attack. Madison, Wisc.: Wisconsin University, 1975. (ERIC Document Reproduction Service No. ED 109 665)
- Lindquist, E. F. (Ed.). Educational measurement. Washington, D.C.: American Council on Education, 1951.
- Mather, P. W. Skill management systems in reading: a real revelation. Journal of Reading, 1977, 21, 139-144.
- McNeil, J. D. False prerequisites in the teaching of reading. Journal of Reading Behavior, 1974, 6, 421-427.
- Millman, J. Reporting student progress: a case for criterion-referenced marking systems. Phi Delta Kappan, 1970, 52, 226-230.
- Millman, J. Criterion-referenced measurement: an alternative. The Reading Teacher, 1972, 26, 278-281.
- Mione, S. A. Criterion-referenced testing: a critical perspective, 1977. (ERIC Document Reproduction Service No. ED 147 757)
- Morrison, H. C. The practice of teaching in the secondary school. Chicago: University of Chicago Press, 1926.
- National Education Association. Guidelines and cautions for considering criterion-referenced testing. Washington, D.C.: 1975. (ERIC Document Reproduction Service No. ED 147 361)
- Popham, J. W., & Husek, T. R. Implications of criterion-referenced measurement. Journal of Educational Measurement, 1969, 6, 1-9.
- Prescott, G. Criterion-referenced test interpretation in reading. The Reading Teacher, 1971, 24, 347-354.
- Smead, V. S. Ability training and task analysis in diagnostic/prescriptive teaching. The Journal of Special Education, 1977, 11, 113-125.
- Smith, F. Understanding reading--a psycholinguistic analysis of reading and learning to read. New York: Holt, Rinehart and Winston, 1971.

- Ten Brink, T. D. Evaluation--a practical guide for teachers. New York: McGraw-Hill, 1974.
- Thompson, R. A., & Dziuban, C. D. Criterion referenced reading tests in perspective. The Reading Teacher, 1973, 27, 292-294.
- Thorndike, R. L. (Ed.). Educational measurement. Washington, D.C.: American Council on Education, 1971.
- Tucker, L. Scales for minimizing the importance of reference groups. In Proceedings, invitational conference on testing problems. Princeton, N.J.: Educational Testing Service, 1952.
- White, S., & Damos, E. An objective-based approach to reading instruction: implementation and evaluation. Paper presented at the Annual Meeting of the International Reading Association, Miami Beach, Florida, May 1977. (ERIC Document Reproduction Service No. ED 142 948)
- Wilson, R. M. Comprehension diagnosis via task analysis. Reading World, 14, 178-179.