INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

University Microfilms International
300 N. Zeeb Road
Ann Arbor, MI 48106

Music, Phyllis Hope Weddington

AN ANALYSIS OF THE RELATIONSHIP BETWEEN SELECTED INSTRUCTIONAL MATERIALS AND THE MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM OBJECTIVES FOR THE PRIMARY GRADES

Michigan State University PH.D. 1982

AN ANALYSIS OF THE RELATIONSHIP BETWEEN SELECTED INSTRUCTIONAL MATERIALS AND THE MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM OBJECTIVES FOR THE PRIMARY GRADES

By

Phyllis Weddington Music

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Administration and Curriculum

1982

ABSTRACT

AN ANALYSIS OF THE RELATIONSHIP BETWEEN SELECTED INSTRUCTIONAL MATERIALS AND THE MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM OBJECTIVES FOR THE PRIMARY GRADES

By

Phyllis Weddington Music

The purpose of this study was fourfold: (1) to determine the relationship between the skill-exercise objectives of three reading series and the MEAP reading objectives tested in grade 4; (2) to compare the percentages of reading-series objectives that matched MEAP objectives, in order to establish congruence with MEAP objectives; (3) to ascertain whether the degree of congruence between reading-series objectives and MEAP objectives affected MEAP-test results; and (4) to determine whether teachers' methods and strategies might have affected MEAP-test results.

For this study, a panel of reading experts was asked to ascertain the purposes of skill exercises in the workbooks and teachers' manuals of three selected reading series. Congruency between MEAP objectives and reading-series objectives was determined according to the percentage of skill-exercise objectives that matched MEAP objectives; then collective scores from Classroom Listing Reports were used to ascertain whether members of the treatment group that used the most congruent series had the highest MEAP-test scores.

Finally, a questionnaire was administered to primary-school teachers in each district, and the results were compared to determine whether their methods and strategies influenced test scores.

Although the Houghton-Mifflin series objectives were found to be the most congruent with MEAP objectives, the treatment group using this series did not perform better than the other two groups on the MEAP test. Collective MEAP-test scores revealed that classes in the three districts performed equally well on four out of five test categories. On the category of critical reading, however, the Holt, Rinehart, and Winston treatment group performed best.

Since the three treatment groups performed equally well on four categories of the MEAP test, it seemed that the methods and strategies of teachers did not affect the test scores for those categories. It is possible, however, that the methods and strategies of teachers did affect the test results for category 4, critical reading, as indicated by the higher scores of the Holt, Rinehart, and Winston treatment group on this category and by their teachers' self-perceptions.

To the memory of my beloved parents,

Crit and Ella Weddington,

both of whom surely share this accomplishment.

ACKNOWLEDGMENTS

Sincere appreciation and gratitude are extended to many individuals and groups for their encouragement and contributions throughout this study. Among those I would like to acknowledge are: my husband, Gene, for his love, encouragement, and assistance each day; my son, Mark Edward, for understanding and enduring my preoccupation with my work; my committee chairperson, Dr. Janet Alleman-Brooks, for her friendship, guidance, and invaluable assistance throughout my doctoral program; committee members Dr. Sheila Fitzgerald, Dr. Peggy Reithmiller, and Dr. Richard Featherstone for their willingness to serve; my mother-in-law, Mildred Johnson, for her constant faith, caring, and assistance; my sisters in Delta Kappa Gamma International, who provided both monetary and moral support; administrators and teachers participating in this study; and my family and friends, both old and new, for their companionship and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION TO THE STUDY

This chapter contains a statement of the need for the study, a statement of the problem, testable hypotheses, a statement of the purpose, the significance of the study, definition of terms, limitations and assumptions, and summary.

Statement of the Need for the Study

The need for this study was evidenced by public concern for accountability of educators and by the back-to-basics movement. These issues have given rise to the need for specifying what is being taught and precisely what is being accomplished, i.e., what students are actually learning.

Various research studies, national polls, and the news media have indicated that the citizenry is concerned about the quality and quantity of instruction in the classroom.[1] "Accountability" seems to be the watchword; people want to be sure their tax dollars are really providing the services that will ensure students have the competencies that are essential for daily living.[2]

---

[1]Robert L. Ebel, "The Case for Minimum Competency Testing," Phi Delta Kappan (April 1978): 546.

[2]Ned B. Lovell, Rodney P. Riegle, and Clinton R. Bunke, "Minimal Competency Testing: Hopes, Fears, and Fallacies," The Educational Forum 45 (January 1981): 199-206.

The need to emphasize fundamental learning has resulted in the back-to-basics movement. This has caused concerned school officials to implement policies that are designed to raise student achievement. The curriculum does make a difference in what students learn; educators, therefore, need to address the increasing concerns for the quality of education by demonstrating that the basic objectives of education are being accomplished through the use of selected curricular materials.

Students are expected to demonstrate, through various types of evaluative programs, what they have learned. Robert L. Ebel, an authority on educational measurement, maintains that the best way to determine how much learning has occurred is to observe, through testing, how successfully the student can cope with the tasks that require learning.[3] Roger Farr, an authority on reading, suggests that if an achievement test is adequate, it must sample the same behaviors as those developed in the instructional program.[4] Achievement cannot be expected on topics that are tested but not taught.[5] A high correlation between curriculum and test content is essential if the test is to measure adequately what has been taught. If the correlation between

---

[3]Robert L. Ebel, The Uses of Standardized Testing, Fastback 93 (Bloomington, Indiana: Phi Delta Kappa Educational Foundation, 1977), p. 8.

[4]Roger Farr, Reading: What Can Be Measured?, An IRA/ENA Knight Research Fund Monograph (Newark, Delaware: The International Reading Association, Inc., 1969), pp. 7-36.

[5]Donald Freeman, Therese Kuhs, Lucy Knappen, and Andrew Porter, A Closer Look at Standardized Tests (East Lansing: Institute for Research on Teaching, Michigan State University, November 1978), pp. 1-10.

the two is low, the test will be insensitive to achievement gains that the curriculum produces.[6] Thus, the test will have little value in the basic purposes of evaluation--namely, decision making.[7] Inconsistencies between these two may result in inaccurate estimations of student achievement of the goals of the curriculum.[8]

A study of the relationship between reading objectives and the objectives tested in the Michigan Educational Assessment Program Reading Test is important because reading is basic; it is the communication process on which success in school depends.[9] The basic purpose of reading is comprehension, which is essential for learning most other content. The components of comprehension identified by Farr include: (1) knowledge of word meanings, (2) ability to select appropriate meanings of words or phrases in a contextual setting, (3) ability to follow passage organization and to identify antecedents and references in it, (4) ability to select the main thought, (5) ability to answer questions that are specifically stated in a passage, (6) ability to use other words to answer a question concerning a passage, (7) ability to draw inferences from a passage about its

---

[6]Kasten G. Tallmadge and Donald P. Horst, "The Use of Different Achievement Tests in the ESEA Title I Evaluation System" (paper presented at the 62nd Annual Meeting of the American Education Research Association, Toronto, Ontario, Canada, 1978), pp. 4-8.

[7]Ralph W. Tyler and Sheldon H. White, Testing, Teaching, and Learning (Washington, D.C.: The National Institute of Education, U.S. Department of Health, Education, and Welfare, 1979), pp. 4, 10.

[8]Ibid., p. 10.

[9]Mary Anne Hall, Jerilyn K. Ribovich, and Chris J. Rameg, Reading and the Elementary School Child (New York: D. Van Nostrand Co., 1979), p. 4.

content, (8) ability to recognize literary devices used in a passage and to determine its mood and intent, and (9) ability to determine the writer's purpose, intent, and point of view, i.e., to draw inferences about a writer.[10]

The officials in Michigan's State Department of Education have responded to public concerns regarding basic skills by developing minimal reading-performance standards and by developing a criterion-referenced testing program with which to sample the reading behaviors that students should have acquired by a given grade level. For example, all fourth-grade students are tested on those objectives they should have achieved in the previous grades, first through third. If the students have not attained the specified objectives, this should be accomplished as soon as possible. In this way, data that are commensurate with the instructional content and purposes of Michigan's schools can be produced.[11]

The State Department of Education holds the position that a common core of objectives, which transcend local district boundaries, do exist and that schools are responsible for helping students attain them.[12] The Department has identified the common core of objectives through input from representatives of the Michigan Reading Association, the Michigan Council of Teachers of Mathematics, the Michigan Department

---

[10]Farr, pp. 53-54.

[11]Michigan Educational Assessment Program, Technical Report, Vol. 1 (Lansing: Michigan State Board of Education, 1980), p. 28.

[12]Ibid.

of Education, teachers, curriculum specialists, school administrators, and other citizens.[13] It can serve, therefore, as a credible guide for determining the essential objectives for reading.

The purpose of the Michigan Educational Assessment Program (MEAP) is to assess validly the outcome of Michigan's instructional programs.[14] These programs are largely determined by the curriculum materials that are purchased and used.[15] It is important, therefore, to determine the extent to which the Michigan Educational Assessment Program's objectives are included in the instructional reading series and whether the district using the series most congruent with the MEAP objectives will produce the best MEAP test results.

Statement of the Problem

In this study the writer compared the 25 MEAP minimal reading objectives with those represented in the grades one through three basal reading exercises of the three most widely used series in Michigan, as determined by Market Data Retrieval, Inc.,[16] and Nicholas P. Criscuolo.[17] Specifically, the researcher attempted to ascertain the following:

---

[13]Ibid. [14]Ibid.

[15]Marda Woodbury, Selecting Instructional Materials, Fastback 110 (Bloomington, Indiana: Phi Delta Kappa Educational Foundation, 1978), p. 26.

[16]Market Data Retrieval, Inc., Reading K-8 Survey, HM Co. Market Research Report No. 17 (New York: Market Data Retrieval, Inc., 1977), p. 97.

[17]Nicholas P. Criscuolo, Improving Classroom Reading Instruction (Worthington, Ohio: Charles A. Jones Publishing Co., 1973), p. 23.

1. How many objectives of the reading-skill exercises found in workbooks and teachers' manuals match MEAP objectives?

2. What reading series has the greatest percentage of skill-exercise objectives that are congruent with MEAP objectives, as determined by a panel of reading experts?

3. Will the degree of congruence between reading-exercise objectives and MEAP objectives affect MEAP test results?

4. Are methods and strategies used in teaching reading likely to influence the MEAP test results?

Hypotheses

General Hypothesis I

There will be no difference between the MEAP reading-test mean scores of classes in the district using the most congruent reading series and the mean scores of the classes in the districts using the other two reading series on each of the dependent variables.

Operational H1a: There will be no difference between the MEAP reading-test mean scores of the classes in the district using the most congruent reading series and the MEAP reading-test mean scores of the classes in the districts using the other two reading series on vocabulary meaning.

Operational H1b: There will be no difference between the MEAP reading-test mean scores of the classes in the district using the most congruent reading series and the MEAP reading-test mean scores of the classes in the districts using the other two reading series on literal comprehension.

Operational H1c: There will be no difference between the MEAP reading-test mean scores of the classes in the district using the most congruent reading series and the MEAP reading-test mean scores of the classes in the districts using the other two reading series on inferential comprehension.

Operational H1d: There will be no difference between the MEAP reading-test mean scores of the classes in the district using the most congruent reading series and the MEAP reading-test mean scores of the classes in the districts using the other two reading series on critical reading.

Operational H1e: There will be no difference between the MEAP reading-test mean scores of the classes in the district using the most congruent reading series and the MEAP reading-test mean scores of the classes in the districts using the other two reading series on related study skills.

General Hypothesis II

There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on each of the dependent variables, according to mean scores on teacher-questionnaire responses.

Operational H2a: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on classroom enrichment, according to mean scores on teacher-questionnaire responses.

Operational H2b: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on homework, according to mean scores on teacher-questionnaire responses.

Operational H2c: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on teacher-made and other materials, according to mean scores on teacher-questionnaire responses.

Operational H2d: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on parental involvement, according to mean scores on teacher-questionnaire responses.

Operational H2e: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on time, according to mean scores on teacher-questionnaire responses.

Operational H2f: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on organizational patterns, according to mean scores on teacher-questionnaire responses.

Operational H2g: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on independent learning, according to mean scores on teacher-questionnaire responses.

Operational H2h: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on student interaction, according to mean scores on teacher-questionnaire responses.

Operational H2i: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on outside enrichment, according to mean scores on teacher-questionnaire responses.

Operational H2j: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on basal materials, according to mean scores on teacher-questionnaire responses.

Operational H2k: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on staff support, according to mean scores on teacher-questionnaire responses.

Operational H2l: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on language development, according to mean scores on teacher-questionnaire responses.

Operational H2m: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on written work, according to mean scores on teacher-questionnaire responses.

Operational H2n: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on test uses, according to mean scores on teacher-questionnaire responses.

Operational H2o: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on rewards, according to mean scores on teacher-questionnaire responses.

Statement of the Purpose

A criterion-referenced MEAP testing procedure was designed to relate student achievement to specific standards of performance according to a predetermined criterion or set of criteria. The MEAP testing instrument was constructed to yield results that would be interpreted in terms of essential minimal reading behaviors. These criteria had to be based on the reading programs used in the classroom because research has shown that what is taught is determined primarily by the instructional-program materials used in a school system.[18]

This study was conducted to provide information concerning the degree to which primary-grade minimal reading objectives, as determined by the MEAP, are represented in teachers' manuals and workbooks of the three most widely used basal reading series in

---

[18]William A. Mehrens and Robert L. Ebel, Some Comments on Criterion-Referenced and Norm-Referenced Achievement Tests, NCME Measurement in Education, Vol. 10, No. 1 (Washington, D.C.: National Council on Measurement in Education, Winter 1979), pp. 4-5.

in Michigan, as determined by Market Data Retrieval, Inc.,[19] and Nicholas P. Criscuolo.[20] These include Ginn and Co.; Holt, Rinehart, and Winston Co.; and Houghton-Mifflin Co. The district that used the reading series that was most congruent with the MEAP objectives was expected to show higher achievement scores than were those districts using the series less closely aligned with MEAP objectives.

Importance of the Study

The validity of the Michigan Assessment Test depends on whether the test samples the same behaviors as those developed through the reading curriculum or course of study. It is generally agreed that the purpose of criterion-referenced testing is to determine whether students have learned what they have been taught. The extent of the relationship between instructional materials and the MEAP should, therefore, be valuable to the following: (1) those revising the Michigan Educational Assessment Test, (2) local school officials who are using the test to ascertain whether students are attaining minimal objectives, (3) teachers who are striving to identify and define more clearly what is essential for students to learn, and (4) publishers who need to know what is essential to be included in the planning and designing of reading series.

Definition of Terms

The following definitions apply throughout this research project.

---

[19]Market Data Retrieval, Inc., p. 97.

[20]Criscuolo, p. 23.

Basal reading series is a set of instructional materials, which is provided in elementary schools and which offers systematic guidance in developing basic reading skills by carefully planned sequences in each level of the series.

Basic reading skills are those skills that are fundamental for later learning and achievement. These include word study, comprehension, and study habits.

Content validity is the degree to which a test measures what it purports to measure. It is sometimes defined as truthfulness.

Criterion-referenced test is a test in which performance is judged, relative to some absolute definition of mastery or adequacy, on a predetermined criterion or set of criteria. It is interpreted with respect to the proportion of correct responses. The term "objective referenced" is used synonymously.

Curriculum, as defined for this study, is the formal course of study used in schools.

Evaluation is the systematic collection and interpretation of information regarding student progress or lack of progress toward educational objectives.

Instructional validity refers to the adequacy of the time and the degree of emphasis used in teaching a specified skill or concept.

Measurement is the quantification of observed learner behavior. The results are commonly used to underscore alleged educational deficiencies or accomplishments.

Minimal competency test is a device used for measuring the achievement of those skills assumed to be necessary for living. It

is a criterion-referenced instrument by which an individual is assessed relative to a predetermined standard or criteria.

Norming population refers to an exemplary population that is selected as representating the larger population for whom the test will eventually be used. This group is tested before the release of the test, and the scores are used as the standard by which to measure the performance of other students of the same age and grade.

Primary grades, as defined for this study, are grades one through three. Primary years of schooling may be used synonymously.

Reliability refers to the consistency with which equivalent forms of a test measure the same skills or concepts.

Skill exercise, as defined for this study, is all subject matter grouped under one title or heading in workbooks and teachers' manuals. A skill exercise covers one or more pages in a workbook; in the teachers' manuals, each skill exercise has a title and a general set of directions.

Standardized achievement tests are measurement devices designed by professional test constructors and drawn from a broad sampling of general content to assess pupils' knowledge and skills at a particular time. Standardized reading achievement test measure word recognition, vocabulary, comprehension, related study skills, and sometimes the rate of reading.

Survey tests, as defined for this study, are norm-referenced tests that are designed to cover a broad range of achievement and to designate generally what students have learned as compared to a specified norming population.

Limitations and Assumptions

The findings of this research study were limited by the following:

1. The Michigan Educational Assessment Program standards for identifying minimal competencies may not include all essential skills for daily living.

2. Other reading experts may not agree that the objectives of the skill exercises are the same as were determined by the panel of reading experts used in this study.

3. Only basal reading workbooks and teachers' manuals were used to determine whether students were taught the essential MEAP objectives.

4. Disproportionate numbers of classes and teacher responses in the three districts provided a greater likelihood of sampling error in some treatment groups than in others.

5. Variation in types of students for whom collective scores were reported on the MEAP Classroom Listing Reports may have affected mean scores of the treatment groups.

6. The quality of instruction, as determined by the teachers' methods, may limit or enhance the degree to which the skill exercises help students attain the objectives that are tested for on the MEAP test.

7. Teachers used a variety of strategies and materials; these were considered through the use of a questionnaire to determine teachers' perceptions of their own instructional methods and strategies. Instructional validity was determined by each instructor individually.

But how an instructor perceived his/her own methods may not necessarily have represented what actually occurred in the classroom.

8. Differences in students' environments, abilities, attitudes, economic opportunities, and the district's ability to provide educational experiences may partially explain the differences in students' achievements on the MEAP test. These may not have been influenced by teaching methods and strategies or materials used.

9. Gross reporting of scores may have accounted for indiscernible differences in MEAP test results.

10. The teachers' manuals and workbooks that were used represented the state of knowledge at the time they were published. Methods recommended for teaching skill exercises may not have been closely aligned with current teaching practices.

11. Differences in teacher responses may be attributed to the manner in which the questionnaires were administered. In District Y, principals preferred to collect responses, but in Districts Z and X, responses were mailed directly to the researcher.

Summary

The first and second research questions are based on the need to include basic-skill objectives in the reading programs that are used in most schools. Concern about accountability has given rise to the need for specifying what has been taught and precisely what has been learned. Research has supported the need for clearly defining those objectives that students need to attain.

The first hypothesis, based on the third research question, has been supported by authorities in both the field of measurement and the field of reading. This hypothesis was founded on the supposition that correlation of curriculum and test content is necessary to measure adequately what has been taught. Supportive research indicated that what is taught is largely determined by the reading series that is selected and used for instructional purposes. Achievement gains can hardly be expected to be meaningful if the objectives of a testing program are not congruent with the skills and concepts taught.

The second hypothesis originated with the fourth research question. This hypothesis was based on the premise that teaching methods and strategies, instructional materials, and factors outside the school environment influence learning. Research has not shown, however, whether individual learning depends more on one variable than another.

Although this study dealt with only some of the variables that influence how and what students learn in reading, it was nonetheless important. Each piece of data that contributes to a greater understanding of how students become proficient readers provides useful information. It was, therefore, important to examine the relationship between teachers' perceptions of what occurs in the classroom and the reading objectives that are tested.

CHAPTER II

REVIEW OF LITERATURE

Introduction

Considerable controversy has been generated over the various means of testing and over whether these tests are valid--whether they actually test what is taught in schools. Concern about accountability has given rise to concern about testing, content taught, and whether these two are related.

One of the major goals for primary students is proficiency in reading because it is an essential tool for further learning. Teachers are expected to teach students to become fluent readers and to prove that this goal has been accomplished. Testing is one measurement device that is commonly used in evaluating student reading performance. The relationship between what is taught and the testing instrument used is important if achievement is to be adequately assessed.

This chapter, therefore, addresses issues related to the following: (1) beginning reading, (2) criterion-referenced tests, and (3) Michigan's Accountability Model, which was designed to ascertain whether students were becoming competent readers.

Beginning Reading

Hall, Ribovich, and Rameg defined reading as a communication process in which written language symbols are used to arrive at a meaningful interpretation of the author's intentions, attitudes, beliefs, and/or feelings, as well as to arrive at a meaningful interpretation of the reader's skill, experience, and intellectual purpose.[1] Its function is shared with writing, speaking, and listening. Both Hall[2] and Chall[3] concurred that reading is central to educational achievement because success in reading is essential for learning other content. In the same vein, Criscuolo reiterated that a basic reading program is most important because it develops skills needed to function effectively in all curricular areas.[4]

Chall cited Bloom's studies as evidence that reading plays a dominant role in the later success of children's academic achievement. In these studies, Bloom found that failure to learn to read by the end of grade one is predictive of later failure.[5] This study, however,

---

[1]Mary Anne Hall, Jerilyn K. Ribovich, and Chris J. Rameg, Reading and the Elementary School Child (New York: D. Van Nostrand Co., 1979), p. 8.

[2]Ibid., p. 4.

[3]Jeanne Chall, Reading 1967-1977: A Decade of Change and Promise, Fastback 97 (Bloomington, Indiana: The Phi Delta Kappa Educational Foundation, 1977), p. 6.

[4]Nicholas P. Criscuolo, Improving Classroom Reading Instruction (Worthington, Ohio: Charles A. Jones Publishing Co., 1973), p. 23.

[5]B. S. Bloom, Human Characteristics and School Learning, quoted by Jeanne Chall in Reading 1967-1977: A Decade of Change and Promise, p. 6.

lacked substance because it failed to address the major issue of the original cause of failure.

The philosophy of teaching reading and theories of how children learn to read have changed considerably over the past two decades. Vocabulary in basal reading materials is no longer rigidly controlled, and units of meaning are believed to be of greater value in comprehending what is read.

During the early part of the twentieth century, rigid vocabulary control in children's basal readers was advocated by Thorndike and other pioneer researchers.[6] It was believed that young children should be given stories and textbooks that contained easier words and a lot of repetition; the purpose of this was to alleviate frustration in the process of learning to read. Acceptance of this theory of vocabulary control was evident, as the publishing industry employed Thorndike's techniques in word-frequency counts.

Now that much more is known about the process of learning to read, researchers and theorists question the use of controlled vocabulary and simple syntax in basal readers. It is now believed that natural-sounding sentences provide contextual cues that help children recognize a wide range of words, not just core vocabulary or frequently encountered sight words. Children may, in fact, respond to printed messages with immediate understanding, rather than going through the laborious process of identifying individual words. Present-day advocates of this whole-language approach do not consider instant

---

[6]John Guthrie, "Vocabulary Control," The Reading Teacher 33 (November 1979): 240-42.

recognition of core vocabulary relevant to measuring the reading process.[7] Instead, they feel that measuring oral-reading errors in relation to self-correction strategies is more appropriate.

English is a positional language. The order of sentences as well as the order of words and phrases within sentences plays an important role in reading comprehension. Meaning is derived from the interaction of sentences. Researchers have found that attempts to make reading easier by using simple sentence structure have only served to complicate the process.[8] Omission of explicit connective words--such as because, instead, for, and while--makes the relationship between sentences vague or incomprehensible if the young reader cannot discern implicit relationships. If the reader fails to understand the implicit relationships integral to the text, both recall and comprehension become more difficult.

Lundsteen studied the effect of simple and complex reading materials on the level of thinking of 190 Berkeley elementary students.[9] She sought to determine whether it was the materials or the level of student maturity that made the difference in their selection of abstract, functional, or concrete word and paragraph meanings. She used three measures having abstract, functional, and concrete choices. These included Choose a Meaning Test, Depth of Meaning

---

[7]Joyce Hood, "Sight Words Are Not Going Out of Style," The Reading Teacher 33 (January 1977): 379-82.

[8]Ibid., pp. 380-82.

[9]Sara W. Lundsteen, "Levels of Meaning in Reading," The Reading Teacher 28 (December 1974): 268-72.

Test, and Creative and Critical Reading (paragraphs) Test. She found that both primary- and upper-grade students made significantly more abstract choices with paragraph measures than with word measures. She concluded that cognitive-level reactions were more relevant to higher-level thoughts that are encompassed in larger units of meaning, regardless of maturity level.

Lundsteen's study suggested that students, both younger and older, need exposure to abstract reading material and to provocative questions about the material. From these tactics, a child can develop a repertoire of various kinds of meanings, which will encourage problem-solving behavior.

As early as 1969, Clay studied the oral-reading strategies of children who had been taught to read by the whole-language approach. She demonstrated children's achievements by reporting errors and self-correction rates for the median children in four ability groups.[10] The highest group had an accuracy rate of 95 percent in self-corrections. She concluded that this group was using other contextual cues without systematic repetition of a core vocabulary. Using the same self-correction system, she found that the lowest group was able to correct only 5 percent of its errors. This lowest group evidently needed a larger core vocabulary to provide contextual clues for the recognition of less-familiar words. This suggested that a percentage of children do need a core vocabulary of known words.

---

[10]Marie M. Clay, "Reading Errors and Self-Correction Behavior," British Journal of Educational Psychology 39 (1969): 47-56.

Biemiller researched the importance of teaching children the basic strategy of risk taking in word pronunciation.[11] He analyzed the oral-reading errors of first graders and found that the errors could be categorized in three phases. At first, beginning readers primarily used context clues to guide their guesses. In the second phase, children frequently made no attempt to guess at all. In the third stage, the students used both graphic and contextual restraints to aid in word identification; in this stage, the number of nonresponse errors decreased.

According to this study, children first approached the reading task confidently but gradually became less willing to risk guessing as they became aware of the constraints placed on their guesses. In the third phase, children, though hesitant, used multiple clues to word meanings and made intelligent guesses with few errors. The children needed a great deal of encouragement and training to develop the confidence to take risks necessary to make the educated guesses that successful readers make.

The psycholinguistic model of reading stresses the interaction between thought and language. It is based on units of meaning that are derived from the whole context, rather than emphasizing word meaning and/or pronunciation in isolation. This process combines cognitive psychology and linguistics to analyze and understand the language and thinking process. Psycholinguists contend that, since

---

[11] A. Biemiller, "The Development of the Use of Graphic and Contextual Information as Children Learn to Read," Reading Research Quarterly 6 (Fall 1970): 75-96.

identification of meaning is the goal of reading, the human brain must rely on syntactical rules to bridge the processing system between visual information and the deeper structure of the brain from which meaning is derived.[12] The interaction of words in sentences depends on the syntax of the language. Even though the speech of young children reflects basic syntactic information, there is evidence that children continue to grow throughout elementary-school years in their ability to understand and use syntactic structures. It is now believed that the failure of beginning readers to organize an author's words into meaningful units creates comprehension problems. Psycholinguists, therefore, advocate the use of whole units of meaning rather than learning or using words in isolation.

Even though a great deal more is known today about the reading process, Hall stated that basals are still the most widely used materials for teaching reading in the United States since the McGuffey Readers were introduced in 1840.[13] McGuffey Readers were characterized by the use of sequenced materials and controlled vocabulary, according to the grade level for which the book in the series was designed.

Woodbury substantiated the fact that textbooks are still the major means of conveying curricular content.[14] They are influential

---

[12]Charles Cooper and Anthony Petrosky, "A Psycholinguistic View of the Fluent Reading Process," Journal of Reading 20 (1976): 184-207.

[13]Hall, Ribovich, and Rameg, p. 8.

[14]Marda Woodbury, Selecting Instructional Materials, Fastback 110 (Bloomington, Indiana: Phi Delta Kappa Educational Foundation, 1978), p. 26.

in standardizing the curriculum by grade levels, and they influence the teaching approach that is used. Often they are validated by field tests in schools before publication. Woodbury further stated that since standardized achievement tests tend to be based on textbook content, scores on achievement tests are correlated with mastery of texts.[15]

In a survey of 1,300 teachers throughout the United States, Spache found that 95 to 98 percent of all primary-grade teachers and 80 percent of the intermediate-grade teachers used basal readers.[16] Spache further noted that seven out of ten elementary classes used the basal reading materials as their only major instructional program.

Rosecky studied the major instructional tools used by 61 teachers in two districts and found evidence suggesting that, not only are basals the core of the reading in most American elementary schools, but teachers are also expected to rely on the basal guidebooks for daily reading instruction.[17] Of those teachers studied, she found that 90 percent used the basal guidebook suggestions most or all of the time. Within that same group, 75 percent omitted all or part of the directed reading activities. The most frequently omitted were such follow-up enrichment activities as dramatization, composing,

---

[15]Ibid., p. 26.

[16]George D. Spache, The Teaching of Reading (Bloomington, Indiana: Phi Delta Kappan, Inc., 1972), pp. 35-36.

[17]Marion Rosecky, "Are Teachers Selective When Using Basal Guidebooks?" The Reading Teacher 31 (January 1978): 273-74.

and reading to students. The reason cited was lack of time for these additional activities, which were considered "extras."

Coordinated textbooks, workbooks, and teachers' manuals form the core of American reading programs.[18] The overall purpose of these programs is to provide comprehensive sequential reading materials. The textbooks are accompanied by workbooks, which contain important supplementary features and which are designed to strengthen and extend the skills presented in textbooks. These skills include the following: (1) decoding activities, (2) syllabication, (3) structural and textual analysis, (4) measurement skills, and (5) comprehension skills.[19]

Spache and Spache studied reports of research done in 36 classes of 18 Los Angeles schools; in these reports, comparisons were made between standardized reading test scores when teachers used workbooks and standardized reading test scores when teacher-made materials were used.[20] It was reported that students using workbooks in the second and third grades had higher vocabulary and comprehension scores, but fifth and sixth graders showed no difference in reading test scores, with or without workbooks. The groups that did not use workbooks used teacher-prepared materials.

After Spache and Spache studied the reports, they studied teachers' attitudes toward preparing their own teaching materials.

---

[18]George D. Spache and Evelyn B. Spache, Reading in Elementary Schools (Boston: Allyn & Bacon, Inc., 1973), pp. 146-70.

[19]Albert J. Harris and Edward R. Sipay, How to Increase Reading Ability (New York: David McKay Co., 1976).

[20]Spache and Spache, p. 170.

They found that teachers preferred not to prepare their own materials but wanted to use workbooks that were closely related to their reading programs. The teachers believed that coordinated workbooks provided concrete evidence of reading growth in both work habits and mechanical skills at primary levels. Spache and Spache concluded that the quality of teacher-made materials might have been affected by teachers' attitudes, thereby affecting students' achievement. The conclusions they reached might have been substantially different if a third control group had used a combination of teacher-made materials and workbooks; attitudes as well as test results might have been different.

Whether materials are made or selected from among various commercial materials that are available, particular attention needs to be given to materials used in the primary grades because researchers have found that most basal comprehension skills are initiated in the primary grades. Jenkins and Pany stated that comprehension skills that received the greatest attention during the middle grades were introduced and taught during the first years of instruction.[21] Hall, Ribovich, and Rameg also discovered that intermediate- and middle-grade students generally are refining the basic reading strategies that were learned at the primary levels.[22] Rosenshine, too, concluded that most

---

[21]Joseph R. Jenkins and Darlene Pany, Teaching Reading Comprehension in the Middle Grades, Reading Education Report No. 4 (Urbana-Champaign: University of Illinois, January 1978), p. 1. (ED 151 756)

[22]Hall, Ribovich, and Rameg, p. 186.

skills practiced in the later elementary years were not new but just a continuation of the progress begun earlier.[23]

Since the initial introduction to desired reading behaviors begins in the primary grades, it is important to consider not only the appropriate content, but also the proper use of that content. Both are important in producing the desired outcomes.

McClutcheon researched the question of how teachers plan activities for students. She found that teachers did not first formulate objectives and then choose the appropriate learning activities, organize them, and select evaluative procedures accordingly.[24] Rather, they focused on the activities suggested or provided for in the teachers' manuals before considering the objectives of learning. McClutcheon cited this as one of the major weaknesses of using basals when trying to provide a balanced program.

Inflexible use of basals and indiscriminate use of the accompanying workbooks were two of the problems that Criscuolo found in his research to be related to inappropriate use of basals.[25] He surveyed 940 elementary students in 34 classrooms and found that only 9 percent changed reading groups during the school year. Immobility

---

[23]B. Rosenshine, "Skill Hierarchies in Reading Comprehension," in Theoretical Issues in Reading Comprehension, ed. R. J. Spiro, B. C. Bruce, and W. F. Brewer (Hillsdale, N.J.: Erlbaum, 1978).

[24]Gail McClutcheon, "How Do Elementary School Teachers Plan? The Nature of Planning and Influences on It," The Elementary School Journal 81 (September 1980): 4-10.

[25]Criscuolo, pp. 75-77.

was diagnosed as resulting from inflexible use of a single set of basal reading materials. He stated that teachers tend to use a single reader for the entire class and to keep students in the same reading group regardless of growth.

Some authorities in the field advocate the use of basals as appropriate reading materials.[26] Basals continue to be a major factor in reading instruction in the nation's schools because they provide the basic organizing thread to ensure that pupils are exposed to an ordered sequence of reading skills.[27]

Criscuolo emphasized that it is advantageous to use basal series because they are adaptable to the theory of continuous child growth and development and because they provide economy of instruction for teachers who need the guidance of a basic reading program.[28] Duffy and Sherman, likewise, advocated the use of basals because there is a defined structure in them that does not depend on teachers' judgments to determine the sequence or content of instruction.[29]

Spache and Spache recommended the use of basal reading series because they have many advantages, including the following: (1) systematic guidance in planned sequential learning, (2) subject matter based on common childhood experiences, (3) a superior program to that

---

[26]Spache and Spache, pp. 160-63.

[27]Jack Bagford, Instructional Competence in Reading (Columbus, Ohio: Charles E. Merrill Publishing Co., 1975), pp. 51-52.

[28]Criscuolo, p. 23.

[29]Gerald G. Duffy and George B. Sherman, How to Teach Reading Systematically (New York: Harper & Row, 1973), pp. 238-39.

which a teacher can create, (4) the development of readiness procedures in sequential steps, (5) core vocabulary, (6) material scaled in difficulty, (7) well-rounded selections, and (8) useful accompanying workbooks to practice and test the development of skills.[30]

Representing the opposing point of view, Durkin questioned whether basal reading series should be used. She stated that the problem with basals is that "what is done with comprehension is commonly concerned with assessment rather than with instruction. Whatever the focus, most activities lead to written exercises."[31]

Huey, too, is still concerned with comprehension rather than specific vocabulary as he was more than 40 years ago. He said that the rigid vocabulary control in basals is contrary to the natural process of learning to read.[32] Children's reading vocabulary should come from their own environment and should pave the way for reflective thinking, he contended. He further emphasized that it is not necessary for children to be able to pronounce words, spell, or write them in order to grasp their meanings in context.

Hall, Ribovich, and Rameg also cited some disadvantages of the basal reading series: They (1) present a limited vocabulary; (2) include stilted language for beginners; (3) included, until the 1970s, sexual stereotyping; (4) present a lack of literary merit;

---

[30]Spache and Spache, pp. 160-63.

[31]Dolores Durkin, Teaching Young Children to Read, 3rd ed. (Boston: Allyn & Bacon, Inc., 1980), pp. 447-48, 482-83.

[32]Edmund Burke Huey, The Psychology and Pedagogy of Reading (Cambridge, Mass.: The M.I.T. Press, 1977), pp. 348-49.

(5) are often used for pronunciation rather than reading for meaning; (6) result in a routinized reading program; (7) result in continuous ability grouping; (8) often result in enrichment materials being ignored; and (9) leave little motivation for learning, if used in a routine manner.[33]

In summary, reading is essential to success in most educational endeavors because the skills that are inherent in reading are the same as those that are essential for learning other content. Textbooks are the major vehicles for conveying curriculum content.

Basal reading programs are used by a majority of schools in the United States because they provide a framework whereby to structure the curriculum. At least 85 percent of the nation's primary teachers use basal reading programs all or most of the time.

There are both advantages and disadvantages to using basal reading programs. Spache and Spache, Criscuolo, Duffy, and others recommended the use of basals because they felt that the advantages far outweighed the disadvantages. Durkin, Huey, and others felt that the basal approach was too narrow, in that it limited many aspects of the reading process.

Much more is known about the reading process now than even a decade ago. Thus the available research can serve as a guide for teachers in determining the manner in which reading will be taught.

---

[33]Hall, Ribovich, and Rameg, pp. 100-113.

Criterion-Referenced Achievement Tests

Today there is great concern about educator accountability. This issue has given rise to the need for determining whether the testing instruments used to measure student achievement are adequate and appropriate. Ebel stated, "No school can do a good job and show it is doing a good job without systematically auditing the results it is getting."[34] If test content and instructional content are not congruent, however, test results could reflect a distorted picture of the effects of instruction.[35] This suggests that tests should measure what is taught, if an accurate representation of instruction is the purpose for administering a test. If a test instrument does not measure achievement in a specific educational program, the growth or lack of growth measured by the test may be unrelated to the instruction that the student received. Only to the extent that the objectives of both the test and the instructional program coincide is the testing instrument a valid means of measuring the success of a program.[36]

---

[34]R. L. Ebel, The Uses of Standardized Testing, Fastback 93 (Bloomington, Indiana: The Phi Delta Kappa Educational Foundation, 1977), p. 31.

[35]Robert E. Floden, Andrew C. Porter, William H. Schmidt, and Donald J. Freeman, Don't They All Measure the Same Thing? Consequences of Selecting Standardized Tests (East Lansing: Institute for Research on Teaching, Michigan State University, July 1978), p. 1.

[36]J. Wayne Wrightstone, Thomas P. Hogan, and Muriel M. Abbott, "Accountability in Education and Associated Measurement Problems," in Readings in Measurement and Evaluation in Education and Psychology, ed. William A. Mehrens (New York: Holt, Rinehart, and Winston, 1976), p. 325.

For many years, the general public has supported the use of standardized tests for internal and external management purposes.[37] Externally, tests permit the public to keep track of what pupils and schools are doing; internally, they can be used for placing children in appropriate instructional programs or groups. Standardized testing responds to the American society's long-standing need to use human resources efficiently, to reward talent regardless of social origins, and to moderate and/or limit the autonomy of local school districts.[38] It is within this framework of American values that precision in measurement has developed.

Defining objectives and measuring outcomes of learning are in step with the trend toward greater accountability. Out of this has grown the need for criterion-referenced tests. These tests are fulfilling an important and vital role, so the demands made on test developers to design appropriate tools to facilitate the measurement process have increased. The growing demands to assess the outcome of instructional programs and to hold teachers and administrators responsible for actual gains in student performance have given impetus to criterion-referenced testing.[39] To date, however, test developers have not devised a method of measuring other factors that enter into

---

[37]Lauren B. Resnick, "Introduction: Research to Inform a Debate," Phi Delta Kappan 62 (May 1981): 623-24.

[38]Daniel P. Resnick, "Testing in America: A Supportive Environment," Phi Delta Kappan 62 (May 1981): 625-28.

[39]Stephen P. Klein and Jacquiline Kosecoff, "Issues and Procedures," in Readings in Measurement and Evaluation in Education and Psychology, ed. William A. Mehrens (New York: Holt, Rinehart, and Winston, 1976), p. 277.

the learning process and affect student performance--such factors as attitude, environment, and child development, over which the school has only partial influence.

The need for specific information relating to the student's mastery of or deficiency in skills has caused a major shift from norm-referenced to criterion-referenced testing.[40] Criterion-referenced procedures are gaining in popularity because they relate student achievement to predetermined objectives, and the results are interpreted in terms of specific standards of performance. During the test-construction process, behavioral objectives are stated first, and then items are designed to sample student ability in relationship to the objectives that are presented. Tyler said that it is valuable to identify and define educational objectives in terms of behavior; this helps teachers recognize more clearly what they expect students to learn.[41] The performance standard set for the mastery of objectives is usually 80 percent.[42] Many people, however, believe that mastery should be determined on the basis of whether the objective is an essential step for further learning. If accomplishment of an objective is necessary for further learning or if it is essential for the next higher step in the learning process, 80 percent is inadequate because 20 percent of the essential component is missing.

---

[40]Bagford, pp. 91-96.

[41]Kappan Interview, "The Father of Behavioral Objectives Criticizes Them: An Interview With Ralph Tyler," Phi Delta Kappan 55 (September 1973): 55-56.

[42]Hall, Ribovich, and Rameg, pp. 286-87.

Criterion- and norm-referenced testing instruments have many similarities as well as differences. Both kinds of tests are used for assessing behaviors and making decisions. A demonstration of proficiency is required in both. It is the interpretation, or anchor point, of the two that is different, according to Farr.[43] Norm-referenced tests are centered on the typical, or average, scores; performance tends to be spread out to make comparisons possible. In criterion-referenced tests, it is not the spread of scores that is of interest but whether students perform well enough to master the objectives. Because more items are used in criterion-referenced measures, the tests are more conducive to diagnosing strengths and weaknesses in limited areas of learning, but they require more testing time.

Norm-referenced and criterion-referenced tests both have special advantages that depend on the purposes for which the tests are used; but Airasian and Madaus, in response to E. L. Thorndike, concurred that criterion-referenced instruments may prevail if education follows the course that developed in the physical sciences.[44] They cited four reasons for the growth of criterion-referenced testing: (1) growing criticism of standardized tests of achievement, (2) growing controversy surrounding grades, (3) growth of instructional technology, and (4) continuing belief that all or almost all students can achieve

---

[43]Roger Farr, "Standardized Reading Tests," in Reading for All, ed. Robert Karlin (Newark, Delaware: The International Reading Association, 1972), pp. 200-206.

[44]Peter Airasian and George Madaus, Measurement in Education (May 1972), citing E. L. Thorndike, Seventeenth Yearbook of the National Society for the Study of Education (1918).

competency in all or most subject matter. It is generally agreed that whether or not competency can be attained by all depends on the level or degree of proficiency required for further learning.

Although the trend toward using criterion-referenced testing instruments is growing, Harris and Sipay specified that one should use norm-referenced tests, as well, in a complementary fashion.[45] It should not be either one or the other; both are useful since they provide different information. Norm-referenced tests help educators discern whether pupils are performing relative to their grade and age expectancy in general areas of learning, whereas criterion-referenced instruments are used to determine whether students have or have not mastered specific objectives.

Since it is interpretation that makes the objectives meaningful for mastery or comparative information, Canney stated that interpretation should always be viewed in relation to other informational sources.[46] To be meaningful for instructional purposes, scores should not only reflect what is taught but should be examined carefully, and decisions should always be tentative. Canney further stated that measuring students' abilities always results in some inaccurate information.

Not only is some inaccurate information generated when using criterion-referenced instruments, but many educators have voiced

---

[45]Harris and Sipay, pp. 165-66.

[46]George Canney, "Organizing and Applying Test Results," in Reading Tests and Teachers: A Practical Guide, ed. Robert Schreiner (Newark, Delaware: The International Reading Association, Inc., 1979), pp. 53-56.

concerns that tests cannot measure all important objectives of the reading curriculum.[47] Guthrie[48] and Blachowicz[49] presented opposing views concerning what can be measured adequately. Guthrie stated that tests are limited to concrete cognitive skill areas because adequate measurement instruments have not been devised for higher-order cognitive achievement or affective and emotional development. He expressed concern for whether this may restrict the goals of education to those that are measurable. Blachowicz, on the other hand, acknowledged that it is difficult, but not impossible, to assess affective objectives; she maintained that these can be appropriately measured as cognitions about affective events and that they can be stated and ordered. Blachowicz further stated that this process is important in clarifying goals for developing reading interest and appreciation.

To determine exactly what criterion-referenced tests can or cannot do, the tests' contents need to be examined with respect to the intended purposes. Criterion-referenced tests are measurement instruments with which individuals are assessed relative to a given standard. Criterion-referenced tests permit educators to ascertain whether a student can display a clearly defined set of behaviors. The "criterion" to which the student's score is referenced is the class of

---

[47]"A Position on Minimum Competencies in Reading," Journal of Reading 23 (October 1979): 50-51.

[48]John T. Guthrie, "Research Testing: Uses and Visibility," Journal of Reading 23 (March 1980): 542-44.

[49]Camille L. Blachowicz, "Reading Objectives: A Competency-Based Accountability Model," The Reading Teacher 28 (April 1975): 659-61.

skills that the test is designed to measure. If the criterion-referenced test is well constructed, it is possible to reference a student's performance to a clearly described set of behaviors so that it can be said with certainty whether or not a pupil possesses a particular competency.[50]

Criterion-referenced tests focus on specific objectives that are clearly defined in a particular learning program or specific objectives that are independent of any instructional program. The items must flow directly from the learning specifications, and they must be representative of the class of objectives being assessed. They should provide information on mastery or nonmastery that can be used for diagnosing deficits or ascertaining strengths when measuring an individual's level of proficiency.[51]

When selecting and administering criterion-referenced tests, the primary consideration, once the purpose for assessment is established, is whether the instrument is valid.[52] There are several different kinds of validity; the referent type in this instance is content validity. To the extent that a test measures what it purports to measure, it is said to have content validity.

---

[50]W. James Popham, "Normative Data for Criterion-Referenced Tests?" Phi Delta Kappan 57 (May 1976): 593-94.

[51]George Prescott, "Criterion-Referenced Test Interpretation," in Elementary Reading Instruction: Selected Materials, ed. Althea Beery, Thomas C. Barrett, and William Powell (Boston: Allyn & Bacon, Inc., 1974), pp. 605-13.

[52]William A. Mehrens and Irvin J. Lehmann, Standardized Tests in Education (New York: Holt, Rinehart, and Winston, 1980), p. 42.

Kavali stated that there are three basic factors that should be used to ascertain how well a test measures reading achievement: (1) an agreement about the skills, knowledge, and understanding that comprise the test selector's definition of reading achievement; (2) an examination of the test to determine the skills, knowledge, and understanding required for it; and (3) a determination of whether the test items correlate with the instructional objectives. Kavali emphasized, "To the extent that the course content is represented in the test, the test is considered to possess adequate content validity."[53] One should examine the test instrument carefully and then determine whether it really matches what one is trying to assess.

Mehrens and Lehmann stressed reliability as the second most important quality that a measurement device should possess.[54] They defined reliability as the degree of consistency between two measures of the same thing. For example, if different scorers measure a person's achievement on similar items, one expects the outcomes to be similar. The same theory applies if one is using equivalent forms of the same test.

Criterion-referenced tests are based on the premise that whatever is worth teaching is worth teaching to the point of mastery. Prescott suggested that this is the basic weakness of criterion-referenced

---

[53]Kenneth Kavali, "Selecting and Evaluating Reading Tests," in Reading Tests and Teachers: A Practical Guide, ed. Robert Schreiner (Newark, Delaware: The International Reading Association, Inc., 1979), pp. 9-23.

[54]Mehrens and Lehmann, p. 42.

test interpretation, one that limits its applicability.[55] Many variables--such as pupil ability, instructor effectiveness, and relevance of content--condition mastery.

Specific limitations of criterion-referenced tests are based on the following facts: (1) mastery is a reasonable criterion; (2) each item in the test has inherent worth by design; (3) hierarchy or sequence of skills exists in any content area; and (4) common consensus regarding the objectives selected for testing exists.[56] In areas in which there are defined goals that are absolutely essential for daily living, these might not be defined as limitations.

Many researchers have advocated criterion-referenced tests as useful for evaluating the effectiveness of instruction and for making decisions concerning appropriate instructional programs for individuals. For example, Otto and Chester stressed the need for criterion-referenced tests, although they cautioned against poorly designed instruments.[57] They stressed the need to proceed with great care when (1) writing objectives for affective qualities, (2) determining the performance necessary for proficiency, (3) specifying the universe of the task, and (4) stressing mastery, rather than retention and transfer of what is learned.

---

[55]Prescott, pp. 605-13.

[56]Kavali, pp. 19-23.

[57]Wayne Otto and Robert D. Chester, Objective-Based Reading (Reading, Mass.: Addison-Wesley Publishing Co., 1976), pp. 250-51.

Jenkins and Pany recommended the use of criterion-referenced tests as a means of overcoming "curriculum bias."[58] Their study matched the words presented in seven reading series with those presented in five reading-achievement tests. Curriculum bias was evidenced between tests for a single curriculum as well as in a single test that was used for different reading curricula. As a result of this study, the need to establish clearly the relationship between test and curricula was emphasized and was advocated as the necessary first step in interpreting test results.

Tallmadge and Horst studied the relationship between test results and the measurement device used.[59] They wanted to ascertain whether one standardized achievement test was more effective than another in measuring achievement gains. They found that both survey types of reading achievement tests were poorly correlated with three different commercial reading programs and were, therefore, insensitive to achievement gains that the curricula produced. They concluded that the only valid way to assess the effects of an instructional treatment was to use a test in which the items were sampled from the same domain as the instructional exercises.

In 1978 a joint statement concerning testing was issued from the National Conference on Achievement Tests and Basic

---

[58]Joseph R. Jenkins and Darlene Pany, "Curriculum Biases in Reading Achievement Tests," Journal of Reading Behavior 10 (Winter 1978): 345-57.

[59]Kasten G. Tallmadge and Donald P. Horst, "The Use of Different Achievement Tests in the ESEA Title I Evaluation System" (paper presented at the 62nd Annual Meeting of the American Educational Research Association, Toronto, Ontario, Canada, 1978), pp. 4-5.

Skills.[60] Renowned educators, professional test developers, and legislators agreed that tests should be used to supply meaningful information about individuals, promote quality education, and help people learn.

The National Council on Measurement in Education surveyed teachers nationally in 1979 to ascertain their attitudes toward standardized testing in general.[61] Teachers were found to be interested in the knowledge that standardized achievement tests provide. The teachers considered tests of this kind to be helpful, fair, and useful; but at the same time, they did not feel that additional testing instruments were necessary.

Turnbull,[62] Harmer,[63] and others agreed that tests are useful and appropriate measurement devices if the purposes for testing are determined in advance and if the results are used as instructional aids. Kavali advised that it is better not to use any test at all if the purposes for testing have not been clearly established before

---

[60] The National Conference on Achievement Tests and Basic Skills (Washington, D.C.: The National Institute of Education, U.S. Department of Health, Education, and Welfare, February 1979), pp. 25-29. (ED 171 784)

[61] Michael Beck and Frank P. Stetz, "Standardized Testing as Viewed by Test Specialists and Users" (paper presented at the National Council on Measurement in Education, Washington, D.C., 1980).

[62] William Turnbull, quoted by John Guthrie in "Research Testing: Uses and Visibility," Journal of Reading 23 (March 1980): 542-44.

[63] William R. Harmer, "The Selection and Use of Survey Reading Achievement Tests," in The Evaluation of Children's Reading Achievement, ed. Thomas C. Barrett (Newark, Delaware: International Reading Association, Inc., 1967), pp. 53-64.

administering it.[64] Harmer reiterated, "Testing for the sake of testing can not be countenanced."[65] He reasoned that if they are not used to promote further learning or to diagnose learning difficulties, they constitute unnecessary expense and waste time and effort--all of which are ill-afforded by our schools.

Perrone pointed out that teachers need to be mindful of the fact that primary students' growth, both mentally and physically, is sporadic and uneven.[66] The skills that these students need for success in school are in a rather fluid, acquisitional stage. A great deal of fluctuation and inconsistency in learning patterns is inevitable. Perrone advocated the use of many evaluative tools over various time spans.

Farr cautioned against reliance on a single testing instrument because performance on any one test is a sample of behavior in a given situation under a single set of conditions.[67] Price concurred and emphasized that since standardized tests are only one device for obtaining information, they have little value in isolation.[68]

---

[64]Kavali, pp. 9-23.

[65]Harmer, pp. 58-62.

[66]Vito Perrone, The Abuses of Standardized Testing, Fastback 92 (Bloomington, Indiana: The Phi Delta Kappa Educational Foundation, 1977), p. 74.

[67]Roger Farr, Reading: What Can Be Measured?, An IRA/ENA Knight Research Fund Monograph (Newark, Delaware: The International Reading Association, Inc., 1969), p. 80.

[68]Gary Price, "Standardized Achievement Tests for Young Children: An Analysis" (paper presented at the Big Ten Symposium on Early Childhood Education, Madison, Wisconsin, October 1978), pp. 1-18.

McDonald suggested that since an achievement test is a measure of performance on a specific test, the score is inadequate to represent overall reading performance.[69] He recommended using four steps to determine a student's level of proficiency: (1) carefully define the variable being measured, (2) test under conditions likely to secure student cooperation, (3) compare raw scores with suitable norms, and (4) consider the possibility that scores may deviate sizably from the true scores.

In the same vein, Tyler and White acknowledged that it is not possible to test all educational objectives that are deemed important because no way has yet been devised to measure higher-order cognitive achievement or affective and emotional development.[70] They stressed the need to make legislators and the general public aware that measurement is restricted to concrete cognitive skills. They feared that if this is not done, the goals of education may be restricted to those that are measurable.

The International Reading Association agreed that not all important outcomes of the reading curriculum can be measured.[71] They feared that dependence on a single test may dictate a narrowing of school curricula to just test items, thus omitting other important

---

[69]A. S. McDonald, "Using Standardized Tests to Determine Reading Proficiency," Journal of Reading 8 (1964): 58-61.

[70]Ralph W. Tyler and Sheldon H. White, Testing, Teaching, and Learning, Report of a Conference on Research on Testing (Washington, D.C.: National Institute on Education, U.S. Department of Health, Education, and Welfare, August 17-26, 1979). (ED 171 784)

[71]"A Position on Minimum Competencies in Reading, pp. 50-51.

objectives. They advocated collecting and using various kinds of assessment data that are gathered many times during the year.

Two well-known national organizations have opposed the use of standardized tests for any and all purposes. The National Education Association (NEA) and the National Association for the Advancement of Colored People (NAACP) have both advocated a moratorium on achievement testing because of the variety of ways the results are being used.[72] The NAACP opposes testing, mainly on the grounds that tests are culturally biased. The NEA task force studied various aspects of tests and testing over a period of three years and concluded that tests were not only biased, but were being used too heavily to determine who was excluded or who was included in programs or in the allocation of federal funds. Both organizations observed that the primary purposes of testing were tracking and categorizing.

In summary, concern about accountability has given rise to a greater use of test instruments than ever before for auditing purposes. The most widely used kind of test is the norm-referenced survey of achievement. Criterion-referenced testing is becoming more popular, however, because the public is demanding precise, rather than general, knowledge concerning student achievement of predetermined objectives, and the results are interpreted in terms of specific standards of performance.

---

[72]Frances Quinto and B. McKenna, Alternatives to Standardized Testing (Washington, D.C.: National Education Association, 1977), pp. 1-17.

Whether one should use norm-referenced or criterion-referenced instruments depends on the information required. Both are useful and serve to substantiate teacher judgment.

Many of the questions regarding criterion-referenced testing are related to interpretation and use of the results. Research showed that teachers generally find these tests worthwhile because they provide information that can be used in the decision-making process. They are helpful in determining students' instructional needs and in diagnosing specific strengths and/or weaknesses. Many groups and individuals, however, are speaking out strongly against reliance on test data for measuring educational growth and categorizing students for instructional purposes.

Michigan's Accountability Model

Since 1969, Michigan has responded to the public's concern about accountability by developing a six-step accountability model with which to establish and implement an assessment program.[73] The six steps are as follows:

1. Involve persons throughout the state in establishing common goals of educational accountability.

2. Translate these goals into objectives that are a broad base of important variables for assessing needs in Michigan schools.

3. Assess needs, as related to the objectives derived from the goals, to provide information for state and local decision makers to help in determining priorities for a variety of needed change efforts.

---

[1]Ernest R. House, Wendell Rivers, and Daniel L. Stufflebeam, "An Assessment of the Michigan Accountability System," Phi Delta Kappan 55 (June 1974): 663-69.

4. Test alternative delivery systems to help districts develop a research basis for their schools.

5. Foster the development of local evaluation capacity to help schools assess local needs; to help them design, implement, and assess their own innovative efforts; and to help them evaluate their personnel fairly.

6. Use feedback to guide state and local policy makers in fulfilling their leadership roles in education.

The assessment program is now based on criterion-referenced testing and is designed to show whether or not individual students have mastered the basic reading and mathematical skills identified as important by Michigan's educators.[74] During the first four years (1969-1973), the instrument was norm referenced and showed how students performed relative to one another. In the 1973-74 school year, the testing instrument was changed to a criterion-referenced test that could more adequately provide information on basic-skills mastery. The State Board of Education, the Executive Office, the legislature, local educators, teachers, students, and parents wanted information on the status and progress of Michigan's basic-skills education.[75]

It was hoped that data produced by the new criterion-referenced instrument would be more specifically related to the instructional content and purpose of Michigan's schools. It was evident that there was a need for a measuring device that would

---

[74]Michigan Educational Assessment Program, Technical Report, Vol. 1 (Lansing: Michigan State Board of Education, 1980), p. 2.

[75]Ibid., p. 21.

reflect specific learning and teaching. This new focus was uniquely adapted to the concept of minimality because the Michigan Educational Assessment Program (MEAP) officials wanted to focus on those objectives on which all students in a given grade should have had instruction; if these objectives have not been attained, they should be acquired as soon as possible.[76]

The Michigan State Department of Education holds the position that a common core of educational objectives does exist; these transcend local district boundaries, and schools are responsible for helping students attain them. In fact, these objectives can be identified through a rational process, and the effort to do so is considered worthwhile.[77] After these common core objectives were identified by officials from the State Department of Education, they were reviewed and approved by a panel of educators, citizens, and students and also by the Council on Elementary and Secondary Education. Both the Michigan Council of Teachers of Mathematics and the Michigan Reading Association were issued contracts to review both the objectives and the results of the respective tests and to make recommendations on how to improve the quality of the objectives.[78]

In his address to the National Conference on Achievement Tests and Basic Skills in 1978, John W. Porter, the Superintendent of

---

[76]Ibid., pp. 21-22.

[77]C. Philip Kearney, David L. Donovan, and Thomas H. Fisher, "In Defense of Michigan's Accountability Program," Phi Delta Kappan 56 (September 1974): 14-19.

[78]An Assessment of Michigan Accountability System (Lansing: Michigan Department of Education, May 1974), pp. 6-8.

Public Instruction for the state of Michigan at that time, discussed the purposes for which the minimal-competency testing program had been established in Michigan. He said it was a matter of public policy to identify those youngsters who had special needs, so that those needs could be met. He said, "Proposals for experimental and demonstration programs are required for systems where the test data demonstrates special needs; compensatory funds are made available."[79]

Porter stated that norm-referenced standardized tests are no longer a viable alternative. They were adequate as long as (1) the operation of public schools was relatively inexpensive; (2) public schools could screen, sort out, and select students; (3) those who paid for the schools were the parents of children in attendance; and (4) the "have nots" did not actively participate. Since these were no longer true, Michigan had an obligation to provide a more specifically outlined program of instruction and to report the results of those programs to all concerned.[80]

Some of the advantages of the Michigan program were cited by Porter: (1) the built-in necessity of studying expectations before beginning instruction; (2) a capacity to report to parents on an item-by-item basis concerning their children's performance; (3) its encouragement of teachers at the classroom level; (4) its automatic

---

[79]John W. Porter, "Achievement Testing: The Interests," in The National Conference on Achievement Tests and Basic Skills, Conference Proceedings (Washington, D.C.: National Institute of Education, U.S. Department of Health, Education, and Welfare, March 1-3, 1978), p. 14.

[80]Ibid.

check on the purchase and use of materials; and (5) its potential for focusing on needed in-service programs.[81]

The MEAP tests are given at the beginning of each school year to fourth-, seventh-, and tenth-grade pupils to gather specific information on student progress in the basic skills. These tests are given early in the school year and cover objectives that the students should have attained in the three years before taking the test.[82]

The revised edition of the fourth-grade MEAP reading test comprises 25 objectives, each of which is measured by a set of three items.[83] Answering two out of three of the items correctly is the standard by which mastery is determined. There are four additional "positive response" test objectives that are used to ascertain the extent to which pupils make reading a part of their personal lives. These objectives are affective in nature and are neither calculated nor included in the proportions data.

One of the major controversial issues surrounding the use of the MEAP is whether the objectives that are assessed are really minimal.[84] Lovell stated that minimal-competency testing is a system for measuring the acquisition of those skills assumed necessary for

---

[81]Ibid.

[82]Questions and Answers About the Michigan Educational Assessment Program (Lansing: Michigan Department of Education, September 1980), p. 4.

[83]Michigan Educational Assessment Program Handbook (Lansing: Michigan Department of Education, 1980), pp. 1-2.

[84]House, Rivers, and Stufflebeam, pp. 663-69.

minimal competency in some area of life or of learning.[85] He did not elaborate further on just how to determine for sure which competencies are necessary or who should decide this issue.

Before the MEAP test was revised, House, Rivers, and Stufflebeam were critical of the test because it was norm-referenced.[86] They felt that norm-referenced tests had content that discriminated against minorities and that norm-referenced interpretation caused unfair comparisons between districts. Responsible state and district personnel contended that it did not measure what teachers were teaching. These criticisms have been arrested by the change to a criterion-referenced instrument.

A third major criticism charged that the test was biased. The MEAP officials responded to this charge, stating that minorities should not be given separate tests since all children can and should achieve the minimal skills that are essential for living in this country.[87]

Since the onset of the testing program, there has been increasing concern by local and state officials about how to report the results fairly and adequately to local communities. As soon as the results are in, in January or February each year, the newspapers carry articles about the districts' scores. Reporting of these

---

[85]Ned B. Lovell, Rodney P. Riegle, and Clinton R. Bunke, "Minimal Competency Testing: Hopes, Fears, and Fallacies," The Educational Forum 45 (January 1981): 199-206.

[86]House, Rivers, and Stufflebeam, pp. 663-69.

[87]Edward D. Roeber, David L. Donovan, and Richard F. Cole, "Telling the Statewide Testing Story...And Living to Tell It Again!" Phi Delta Kappan 62 (December 1980): 273-74.

results was at first comparative in nature, and this was damaging when comparisons were made among races and/or between districts that differed economically. School officials felt that such comparisons were unfair since they did not consider individual progress or child-growth and development concepts.

MEAP officials attempted to alter this situation with strategies based on the following assumptions: (1) the public has a right to know, even if school officials fear to air problems; (2) the public has a right to know what action is being taken to remedy the situation if the scores indicate that students have not acquired minimal skills; (3) it is better to provide honest and complete information on the results than to have the negative aspects discovered; (4) to put these scores in the proper perspective, the public should be given information on performance measures related to other school learning, not just to minimal skills.[88]

School superintendents, other key school personnel, and the news media were briefed by state officials on how to present MEAP scores to the public so that the scores could be interpreted with respect to each school district's accomplishments and not in relation to other districts. Roeber, Donovan, and Cole maintained that this public-relations campaign helped the news media report scores, based on an understanding of the purpose of the test. They said that media coverage has changed; newspaper articles now deal with fewer comparisons, and the scores are more favorably interpreted and

---

[88]Ibid.

reported.[89] In view of many recent Michigan newspaper articles, this claim is questioned by many people who are concerned about educational practices in Michigan.

Two other perplexing concerns about the MEAP test program, which have yet to be resolved, relate to testing conditions and cost effectiveness. Both of these issues must be realistically appraised if the test is to produce meaningful results and if the program is to be economically sound.

Testing conditions and administrative procedures may not necessarily be uniform among different groups that are tested, even though they are reported as if they are. Test results can only reflect the skill level of the student, which was demonstrated on the specified day and time that the test was administered; they cannot account for adverse conditions or for administrative procedures that deviate from the manual instructions. Most educators would agree that either one or both of these factors may influence the test results.

In response to issues concerning the cost of the MEAP testing program, MEAP officials stated emphatically that the cost is reasonable compared to the cost of alternative commercial testing programs.[90] They reported that the cost to Michigan in 1979-80 was $190,000 for the MEAP every-pupil mathematics and reading tests. This was approximately $1.05 per pupil. This amount included test development,

---

[89] Ibid.

[90] Questions and Answers About the Michigan Educational Assessment Program, p. 5.

printing, and distribution costs but not administering or grading costs. These last two costs were borne by the school district. Other hidden costs, such as those for student and teacher preparation time, were not considered in the cost analysis.

In summary, Michigan's six-step accountability program centers on a criterion-referenced testing instrument that validly assesses whether students are acquiring minimal competencies in those skills deemed essential for daily living. This program is considered valid because it has been endorsed by many educational leaders, and it measures those objectives that have been defined as minimal.

At first, the MEAP test was norm referenced. Later it was changed to a criterion-referenced test so that it could show whether students were mastering the basic reading and mathematical skills identified as important by Michigan educators and endorsed by the Michigan Reading Association and the Michigan Council of Teachers of Mathematics.

Testing conditions and administrative procedures have not been kept constant from group to group, even though both of these factors may influence test results. No way has been devised to control these variables.

MEAP officials feel that the cost of the MEAP test is reasonable, even though hidden costs have not been analyzed. Individual districts must bear a portion of this cost.

Summary of the Literature Review

Although there is a need to include minimal skills in the instructional programs, little has been done to ascertain whether these skills are actually included in the instructional materials that are used. Inasmuch as they have been deemed worthy of attainment, it is important that students receive instruction in them.

Reading is a basic skill on which school success depends because reading is essential for learning most other content. Teachers use basal reading programs 85 to 90 percent of the time in the primary grades, and they follow the suggested activities in the teachers' manuals. The manuals also supply information on coordinated workbook pages that provide skill practice for basic reading objectives.

Norm-referenced survey tests of achievement are the most widely used kinds of tests. Criterion-referenced tests, however, are gaining in popularity because they provide specific information concerning whether students have mastered predetermined objectives. Criterion-referenced tests are more closely correlated with instructional programs than are norm-referenced tests.

Michigan developed and implemented a statewide criterion-referenced test designed to ascertain the state of basic-skills education. The objectives were developed on the theory that there is a common core of essential basic skills that all, or nearly all, children can attain. These function as a guideline for Michigan's schools, to ensure that students have acquired the "basics."

CHAPTER III

PROCEDURE AND METHODOLOGY

The procedure and methodology of this study are presented in this chapter. The topics included are research questions, materials and sample, research design, and the analysis treatments. Included in the research design are (1) the method used to select reading series, (2) the means of determining congruency between objectives of exercises in workbooks and teachers' manuals and the objectives of MEAP, and (3) a description of the instruments.

Research Questions

The four research questions originated with the statement of the problem, which was based on the study of the relationship between reading series used in primary classrooms and objectives tested in the MEAP at the conclusion of the primary years of schooling. They were:

1. How many objectives of the reading-skill exercises found in workbooks and teachers' manuals match MEAP objectives?

2. What reading series has the greatest percentage of skill-exercise objectives that are congruent with MEAP objectives, as determined by a panel of reading experts?

3. Will the degree of congruence between reading-exercise objectives and MEAP objectives affect MEAP test results?

4. Are methods and strategies used in teaching reading likely to influence the MEAP test results?

Materials and Sample

The materials used in answering the first two research questions were (1) all student skill exercises in the workbooks and teachers' manuals for each of the three basal reading series and (2) the Michigan Educational Assessment Program's 25 objectives for reading. (The reading series included Ginn and Co.; Holt, Rinehart, and Winston Co.; and Houghton-Mifflin Co.)

To answer the third research question, collective reading test scores were taken from all the MEAP Grade Four Classroom Listing Reports in three Western Michigan school districts (X, Y, Z). These districts had mandated the use of the three selected reading series for a minimum of three consecutive years before administering the 1980-81 MEAP test. The three treatment groups, or districts, that were compared included 11 fourth-grade classes in District X, which used the Houghton-Mifflin series; 12 fourth-grade classes in District Y, which used the Holt, Rinehart, and Winston series; and 30 fourth-grade classes in District Z, which used the Ginn series.

The three districts were not comparable in size nor in the state equalized value of property per pupil. These districts, however, had comparable expenditures per pupil for instructional programs, and median incomes were approximately the same. According to 1981 data

from the Michigan Department of Education,[1] District Z had a state equalized property value of $77,807 for each of its 8,998 pupils, and the per pupil expenditure for instructional programs was $1,006. District Y had a state equalized property value of $48,307 for each of its 4,736 pupils, and the per pupil expenditure for instructional programs was $1,103. District X had a state equalized property value of $43,509 for each of its 3,817 pupils, and the per pupil expenditure for instructional programs was $1,197. According to 1977 data from the Detroit Census Bureau,[2] the median income for District Z was $5,300. The median incomes for Districts Y and X were $5,547 and $5,191, respectively.

Test scores for classes in District X included those of 10 special-education and 24 Title I students. District Y scores included those of one special-education and one Title I student. No scores for special-education students were reported in the District Z summaries. No scores for bilingual students were included in any of the three districts' summaries.

In response to the fourth research question, a questionnaire was used to survey first- through third-grade teachers in each of the three districts concerning teaching methods and strategies. Ninety-one percent of the 34 teachers in District X replied. In District Y, 75 percent of the 32 teachers replied, and in District Z, 93 percent of the 72 teachers replied. Most of the nonrespondents

---

[1]Shirley Waldon, Supervisor of State Aid, Michigan Department of Education, telephone interview, July 12, 1982.

[2]Timothy Jones, Information Services Specialist, Detroit Census Bureau, telephone interview, July 12, 1982.

were from District Y. This was the only group that chose to have questionnaires collected by school administrators rather than mailing responses directly to the researcher.

Research Design

Basic Design Elements and Procedure

The basic design consisted of the following elements: (1) MEAP reading objectives, collective reading test scores from MEAP Grade Four Classroom Listing Reports, and a teacher questionnaire; (2) three treatments (sets of workbooks and teachers' manuals); (3) contingency tables; and (4) the statistical model, univariate analysis of variance.

In response to the research questions, observations were based on (1) a comparison of the objectives inherent in the skill exercises of grades one through three in the workbooks and teachers' manuals for the basal series, with the MEAP objectives; (2) comparisons among the three series, according to congruence with MEAP objectives; (3) comparisons of the MEAP test results among the three districts that were using the specified reading series; and (4) an analysis, according to district, of first- through third-grade teachers' perceptions of their own methods and strategies.

Contingency tables were used both for comparisons of reading-series objectives with MEAP objectives and for comparisons among the objectives of the three reading series. Univariate analysis of variance was used to compare the three treatments (districts) according

to MEAP test results and according to categories of teacher responses on questionnaires.

The procedure used to analyze the reading series, the test results, and the questionnaires was the same for all treatment groups. The following design sequence was used:

1. Three reading experts analyzed all of the skill exercises in the teachers' manuals and workbooks for the three reading series; they identified the objective of each exercise and determined whether these exercise objectives matched the MEAP objectives.

2. Comparisons were made of the skill exercises in the three series, according to the percentage of matches and mismatches across the MEAP objectives.

3. Comparisons were made of the collective MEAP reading test scores for the fourth-grade classes across the three districts.

4. Responses to the questionnaire were analyzed by district and were compared to the MEAP test results.

Description of Materials and Instruments

Two types of materials and two instruments were used to obtain data. The materials used were (1) the skill exercises in the sets of workbooks and teachers' manuals for each of the three reading series and (2) a list of the 25 MEAP fourth-grade reading objectives from the MEAP Handbook.[3] The instruments used were

---

[3]Michigan Educational Assessment Program Handbook (Lansing: Michigan Department of Education, 1980), p. 48.

(1) the collective fourth-grade reading test scores from the Classroom Listing Reports for 1980-81 and (2) a first- through third-grade teacher questionnaire.

The criteria for selecting treatment groups were based on whether the students in the district had used one of the three reading series that were most widely adopted in Michigan and on whether they had used it for three consecutive years before taking the fourth-grade MEAP test. The most widely adopted and used basal series in Michigan were determined from reports by Nicholas P. Criscuolo[4] and Market Data Retrieval, Inc.[5] In 1973, Criscuolo reported that Ginn and Co. was the most widely used basal reading series in Michigan and that the Houghton-Mifflin Co. series and the Holt, Rinehart, and Winston series were second and third, respectively. In 1977, Market Data Retrieval, Inc., reported that Ginn and Co. was still the most widely used series but that the Holt, Rinehart, and Winston series had become second and the Houghton-Mifflin series was in third place.

The unit of analysis in each of the three reading series was the skill exercise. The purposes of the exercises in the workbooks and teachers' manuals were determined independently by a panel of certified reading experts. The experts examined the skill exercises in all three series. There were seven sets of workbooks and teachers'

---

[4]Nicholas P. Criscuolo, Improving Classroom Reading Instruction (Worthington, Ohio: Charles A. Jones Publishing Co., 1973), p. 23.

[5]Market Data Retrieval, Inc., Reading K-8 Survey, HM Co. Market Research Report No. 17 (New York: Market Data Retrieval, Inc., 1977), p. 97.

manuals that were graduated in difficulty from Levels 2 through 9 in the Ginn series. Levels 3 through 4 were grouped together in a single manual and workbook; all other levels had separate manuals and workbooks. The Holt, Rinehart, and Winston series had seven sets of workbooks and teachers' manuals; Levels 3 through 6 were grouped together in one set, whereas Levels 7 through 12 had one workbook and one teachers' manual for each level in the set. The Houghton-Mifflin series had nine sets of workbooks and manuals that were graduated in difficulty from Levels 3A, 3B, 3C, and 4 through 9. (See Table 3.1.) All of these sets of early 1970s workbooks and manuals were published for use in primary grades one through three and were not designed for one particular grade.

The 25 objectives, which all students in a given grade should have attained, were identified by the State Department of Education in Michigan. These objectives were categorized in the MEAP Handbook as follows:[6]

1. Vocabulary Meaning (6 objectives)
2. Literal Comprehension (5 objectives)
3. Inferential Comprehension (9 objectives)
4. Critical Reading Skills (1 objective)
5. Related Study Skills (4 objectives)

A list of the specific objectives in each category can be found in Table 3.1.

The 25 MEAP minimal-performance objectives provide the basis for the MEAP testing program. The test was designed to assess

---

[6]Michigan Educational Assessment Program Handbook, p. 3.

whether individual students have attained the MEAP objectives. The collective test scores for grade four from the MEAP Classroom Listing Reports for 1980-81 within each district were used to ascertain whether the district that was identified as using the materials that were most congruent with the MEAP objectives had the highest MEAP test scores.

The MEAP test is administered during the initial weeks of grade four because the test is a measure of minimal skills that should have been acquired during grades one through three.[7] The test items were based on those objectives that selected Michigan teachers, curriculum specialists, and other educators identified as important.[8] Each reading objective is assessed by a set of three test items. A student must answer two out of three of the items correctly to demonstrate attainment of the objective, according to the standards set by MEAP.[9] The Classroom Listing Reports that were used in this study are summaries of class averages from the MEAP reading test. (An example of a Classroom Listing Report for fourth-grade reading is shown in Appendix C.)

The second instrument that was used was a questionnaire. To consider a variety of variables that might have had a bearing on student learning, a questionnaire was sent to all first- through

---

[7]Michigan Educational Assessment Program, Technical Report, Vol. 1 (Lansing: Michigan State Board of Education, 1980), p. 1.

[8]Michigan Educational Assessment Program Handbook, p. 2.

[9]Ibid., p. 1.

third-grade teachers in the three districts. Part I of the questionnaire was made up of eight background-information questions; these were related to teacher training, basal series used, and years of experience. The second part of the questionnaire was made up of 55 questions concerning teaching methods and strategies. The categories of these questions were (1) classroom enrichment, (2) homework, (3) teacher-made and other materials that were not part of the basal series, (4) parental involvement, (5) time, (6) organizational pattern, (7) independent learning, (8) student interaction, (9) outside enrichment, (10) use of basal materials, (11) staff support, (12) language-development activities, (13) written work, (14) test uses, and (15) rewards. (See Appendix A for specific questions.)

Pilot Studies

Three pilot studies were conducted. First, a training session for the panel of reading experts was held. Second, the teacher questionnaire was pilot-tested; third, the questionnaire was revised and pilot-tested again.

The reading experts were recommended by a certified reading specialist who was on the faculty of Central Michigan University and was not affiliated with the three treatment groups. Each of the three reading experts held an M.A. degree in reading and was an elementary-school reading consultant in his/her local school district but not in schools involved in this study.

The researcher conducted a training session for the panel of reading experts so that they might gain consistency in the application of rules and procedures for matching skill-exercise objectives with

MEAP objectives. During this training session, three identical sets of workbooks and teachers' manuals, which were not a part of the study, were used. The experts were instructed to examine each skill exercise independently and then to match it with the appropriate objective. In a situation in which an exercise contained more than one objective, the experts were instructed to classify the exercise according to which objective made up the majority of the exercise. (A majority of the exercises were defined as simply "more than half.") Based on this determination, the experts were instructed to list the number of each exercise, according to its title and the objective it represented, in the appropriate cell of a prepared table. This process was repeated until all exercises had been examined.

After the skill exercises were matched with the appropriate objectives, the group reconvened to discuss questions, concerns, and differing opinions so that the thinking of the group would be more closely aligned. A rater reliability of .84 was established among the three certified reading experts.

After a questionnaire was developed, two pilot studies were conducted to examine other variables that might have had a bearing on the reading scores of the classes within each of the three districts. Input was sought from a certified reading specialist to ascertain whether the questionnaire included the variables that were important in beginning reading classes. This specialist recommended that four additional questions be added and that one other question be changed to a continuum form. (See Appendix A for recommended changes.)

The questionnaire was pilot-tested on 21 elementary-school teachers who were not included in the study. Eighty-five percent of the group responded, and changes were made. (See Appendix A for changes that were made.) Eight volunteers from the same group were pilot-tested a second time, and verbal input was sought from each. The final draft of the questionnaire was administered to all first- through third-grade teachers in Districts Z, Y, and X. A majority of the teachers responded.

Analysis Treatments

Descriptive and statistical treatments were used. Contingency tables were prepared (1) to ascertain how frequently the skill-exercise objectives matched the MEAP objectives and (2) to calculate the percentage of reading-exercise objectives that were congruent with MEAP objectives. Univariate analysis of variance was used to compare the reading-test scores on the five MEAP categories and also to compare the scores on questionnaire categories in each of the treatment groups.

Contingency Tables

A contingency table is a joint-frequency distribution of cases as defined by the categories of two or more variables. The chief component of the contingency-table analysis is the display of the distribution of cases in the interacting cells.

Contingency Table 3.1 was prepared to determine how frequently skill-exercise objectives matched MEAP objectives. The frequency distributions provided the raw data for calculating what percentages of

Table 3.1.--Matrix for determining correlation of MEAP objectives with the skill-exercise objectives found in workbooks and teachers' manuals of three reading series.

MEAP OBJECTIVES	Ginn & Co.							Holt, Rinehart, & Winston						Houghton-Mifflin Co.								
	My Sound and Word.. (2)	A Duck Is a Duck.. (3-4)	May I Come In? (5)	Seven Is Magic (6)	The Dog Next Door.. (7)	How It Is Nowadays (8)	With Skies and Wings (9)	Rhymes and Tales.. (3-6)	A Place for Me (7)	A Time for Friends (8)	People Need People (9)	The Way of the World (10)	Never Give Up (11)	Special Happenings (12)	Tigers (3A)	Lions (3B)	Dinosaurs (3C)	Rainbows (4)	Signposts (5)	Secrets (6)	Rewards (7)	Panorama (8)	Fiesta (9)
VOCABULARY MEANING																							
1. Word meanings affected by prefix																							
2. Word meanings affected by suffix																							
3. Meanings when word has multi-meanings																							
4. Identifying synonyms																							
5. Identifying antonyms																							
6. Determining meaning based on context																							
LITERAL COMPREHENSION																							
7. Identifying main idea of a selection																							
8. Identifying supportive details																							
9. Identifying sequence in a selection																							
10. Identifying cause/effect relationship																							
11. Identifying stated likeness/diff.																							
INFERENTIAL COMPREHENSION																							
12. Inferring main idea of selection																							
13. Inferring cause/effect																							
14. Predicting probable outcome																							
15. Inferring details that support main idea																							
16. Inferring sequence																							
17. Inferring likenesses/differences																							
18. Drawing conclusions from given info.																							
19. Identifying relationships (analogies)																							
20. Making inferences about characters																							
CRITICAL READING																							
21. Determining author's purpose																							
RELATED STUDY SKILLS																							
22. Identifying major uses of dictionary/ glossary/contents																							
23. Locating information in dictionary/ glossary/contents																							
24. Summarizing a selection																							
25. Alphabetizing to second letter																							
OTHER																							
26. Other																							

MEAP and "other" objectives were covered in the basal workbooks and teachers' manuals. ("Other" was a convenient term that would permit the inclusion of all exercises in the classification process.) There were 598 cells in a 23 x 26 design for the skill-exercise data, as it interacted with 26 objectives. The 25 MEAP objectives and the "other" objectives were grouped categorically and placed down the left side of the table. The 23 titles of basal-series levels were alphabetized by publisher across the top of the table. Reading experts independently determined whether the skill-exercise objectives in the teachers' manuals and workbooks of each reading series matched the MEAP objectives. Their determinations were classified by number in the cells formed by the intersection between the objectives and the titles of materials. If an exercise did not match any of the MEAP objectives, it was placed in the category labeled "other."

Contingency Table 3.2 was prepared to summarize the percentages and proportions of skill-exercise objectives that intersected with three groups of objectives. These included MEAP objectives, "other" objectives, or objectives not rated in each of the three reading series. "Other" was the term used for any objective of the reading-skill exercise that was not congruent with a MEAP objective. Objectives "not rated" were skill-exercise objectives on which the experts did not reach a consensus. A 3 x 3 design was used.

Analysis of Variance

One-way univariate analysis of variance (ANOVA) was used to analyze the MEAP reading test scores by categories of objectives.

Scores on categories of responses to the questionnaire administered in each district were also compared through univariate analysis of variance. All analyses of variance were tested for significance at an alpha level of .10. This alpha level was chosen because the three treatments were similar in format and structure; scores were expected to be very close. An alpha level greater than .05 would provide a wider scope from which to discuss differences in the groups.

Table 3.2.--Distribution of objectives for skill exercises across reading series.

	Ginn and Co. Skill-Exercise Objectives	Holt, Rinehart, and Winston Skill-Exercise Objectives	Houghton-Mifflin Skill-Exercise Objectives
MEAP Objectives			
Other Objectives			
Exercises Not Rated			
	(N = 1,327) Skill Exercises	(N = 1,094) Skill Exercises	(N = 1,399) Skill Exercises

One-way analysis of variance is appropriately used when only one treatment is administered to each group of subjects. Univariate ANOVA is used when there is one dependent variable in each analysis. This latter treatment was used to determine (1) whether the mean scores of the three districts differed significantly from one another on the five MEAP objectives (variables) and (2) whether the scores

on the categories of the questionnaire responses differed significantly from one another according to district.

Table 3.3 illustrates the layout of data for the 3 x 5 design, which was used to compare the mean scores of the classes in the three districts, or treatment groups, with the five categories of MEAP objectives.

Table 3.3.--Matrix for 3 x 5 design for comparing treatment groups' mean scores on categories of MEAP objectives.

Factor A (Main Effect) Independent Variable (Categories of Objectives)	Factor B (Main Effect Independent Variable (Treatments)		
	(Z) Ginn (N=30 classes)	(Y) Holt (N=12 classes)	(X) Houghton-Mifflin (N=11 classes)
1. Vocabulary Meaning			
2. Literal Comprehension			
3. Inferential Comprehension			
4. Critical Reading			
5. Related Study Skills			

Univariate analysis of variance was used to determine whether the mean scores on Factor A (categories of objectives) differed significantly from one another in each of the treatment groups (Factor B). The independent variables (categories of objectives) and treatments

were regarded as fixed factors since they were drawn from the entire target population. There was not an equal number of observations in each cell because of the gross reporting of scores. The Ginn variable had 30 observations per cell, whereas the Holt variable had 12 observations per cell and the Houghton-Mifflin variable had 11 observations per cell.

Table 3.4 illustrates the layout of data for the 3 x 15 design, which was used to compare the scores on categories of responses to the questionnaire administered in each district. Scores were assigned to the following values: 1 = always, 2 = most of the time, 3 = sometimes, 4 = seldom, and 5 = never.

Univariate analysis of variance was used to determine whether the teachers in the three districts (treatment groups) differed significantly on the strategies and methods they used in teaching reading and to determine whether these differences affected MEAP test results. Since all first- through third-grade teachers influence how and what their students learn and since the MEAP objectives tested in grade four were supposed to have been attained in grades one through three, the questionnaire was administered to all first- through third-grade teachers in the three districts, even though not all teachers had taught these particular primary students. There was not an equal number of observations in each cell of the 3 x 15 design. The treatment groups varied as to their number of first- through third-grade classroom teachers. District Z had 66 responses to the questionnaire, whereas District Y had 24 and District X had 31.

Table 3.4.--Matrix for 3 x 15 design for comparing treatment groups' mean scores on categories of questionnaire responses.

Factor A (Main Effect) Independent Variables (Questionnaire Categories)	Factor B (Main Effect) Independent Variable (Treatments)		
	(Z) Ginn (N=66 teachers)	(Y) Holt (N=24 teachers)	(X) Houghton-Mifflin (N=31 teachers)
1. Classroom enrichment			
2. Homework			
3. Teacher-made and other...			
4. Parental involvement			
5. Time			
6. Organizational pattern			
7. Independent learning			
8. Student interaction			
9. Outside enrichment			
10. Use of basal materials			
11. Staff support			
12. Language dev. act.			
13. Written work			
14. Test uses			
15. Rewards			

Summary

The objectives inherent in the skill exercises found in workbooks and teachers' manuals of three basal reading series were examined by reading experts to determine how they compared with MEAP objectives. Then percentages of these objectives were calculated according to the proportion of matches or mismatches across MEAP objectives. Frequency distributions on a contingency table provided raw data for calculating the percentages of MEAP objectives covered in each of the reading series workbooks and teachers' manuals.

Contingency tables were prepared in a 23 x 26 design for the frequency count, and the percentage data were summarized in a 3 x 3 design. The 23 titles of levels within the three reading series were alphabetized by publisher across the top of Table 3.1. The 25 MEAP objectives and the "other" objectives were placed down the left side of the table. Skill-exercise objectives were classified by number in the cells formed by the intersections between titles and objectives. Table 3.2 was designed in a 3 x 3 layout, which was used to summarize percentage data for skill-exercise objectives by publisher that intersected with MEAP objectives, other objectives, or objectives not rated.

Univariate analysis of variance was the method used to compare the mean scores of the five categories of MEAP objectives in each of the three districts. A 3 x 5 design was used.

Univariate analysis of variance was the method used to compare the mean scores of the 15 categories of questionnaire responses in each of the three districts. A 3 x 15 design was used.

The purpose of the questionnaire was to identify other variables that might have influenced test scores. The questionnaire was reviewed by a reading specialist, pilot-tested, revised, and pilot-tested again.

CHAPTER IV

DATA ANALYSIS

This chapter contains a review of the purpose of the study, descriptive data for Research Questions 1 and 2, hypotheses for Research Questions 3 and 4, and a report of the findings of the study. The findings for the first two research questions are reported descriptively according to frequency counts and percentage data. The findings for the last two research questions are reported according to rejection or acceptance of the null hypotheses.

Purpose of the Study

First, this study was conducted to determine the relationship between the skill-exercise objectives of the three most widely used reading series in Michigan and the MEAP reading objectives tested in grade four. Second, this study sought to compare the objectives of the skill exercises in the reading series to one another, in order to determine which one of the three had the highest percentage of objectives that were congruent with the MEAP objectives. Third, this study attempted to ascertain if the pupils in the school district using the basal series that was most congruent with the MEAP reading objectives had the highest reading scores on the MEAP test. Fourth, the study attempted to determine whether the methods and strategies of teachers

in each of the three districts might have influenced MEAP test results.

Research Questions

A contingency table was used in response to the following research question:

> How many objectives of the reading-skill exercises found in workbooks and teachers' manuals match MEAP objectives?

Frequency counts in contingency Table 3.1 were used when describing the relationship between the objectives inherent in the skill exercises found in the workbooks and teachers' manuals of each series and the 25 MEAP objectives and "other" objectives. (It was expected that the skill-exercise objectives of the three series would differ from the MEAP objectives.)

A contingency table was used in response to the following research question:

> What reading series has the greatest percentage of skill-exercise objectives that are congruent with MEAP objectives, as determined by a panel of reading experts?

Contingency Table 3.2 was used to describe the percentage of reading-series objectives that corresponded with the MEAP objectives. Frequency distributions across all 25 MEAP objectives and "other" objectives provided the raw data for calculating the percentages of matches and mismatches with skill-exercise objectives from the three basal series. (It was anticipated that one of the basal series would be more closely aligned with the MEAP objectives than would the other two.)

Hypothesis I was derived from the third research question, which is as follows:

> Will the degree of congruence between reading-exercise objectives and MEAP objectives affect MEAP test results?

The first testable hypothesis was based on a comparison of the MEAP test scores in the three treatment groups (Districts X, Y, and Z) to determine what effect the congruence of reading-series objectives with MEAP objectives had on the test scores of the classes within each of the districts. (Congruency was determined according to the percentages of matches and mismatches of reading-series objectives across MEAP objectives in Table 3.2.) The independent variables were the treatments, or reading series, used by the three districts and the five categories of objectives tested by the MEAP testing program. The dependent variables were the mean scores obtained on the five categories of fourth-grade MEAP reading-test objectives.

Analysis of variance was used to compare the mean scores on the five categories of MEAP objectives according to classes within the three districts using the three treatments. The mean scores on the MEAP reading tests of the classes within the three districts were expected to differ. The mean scores of the classes within the district using the basal series that was most congruent with MEAP objectives were expected to exceed the mean scores of the other two treatment groups, or districts.

Hypothesis I stated that there would be no difference on each of the dependent variables between the MEAP reading-test mean scores of the classes in the district using the most congruent reading

series and the MEAP reading-test mean scores of the classes in the districts using the other two reading series. To enable the researcher to compare the mean scores of the three treatment groups on each of the five categories of MEAP objectives, Hypothesis I was restated as five operational hypotheses:

> Operational H1a: There will be no difference between the MEAP reading-test mean scores of the classes in the district using the most congruent reading series and the MEAP reading-test mean scores of the classes in the districts using the other two reading series on vocabulary meaning.
>
> Operational H1b: There will be no difference between the MEAP reading-test mean scores of the classes in the district using the most congruent reading series and the MEAP reading-test mean scores of the classes in the districts using the other two reading series on literal comprehension.
>
> Operational H1c: There will be no difference between the MEAP reading-test mean scores of the classes in the district using the most congruent reading series and the MEAP reading-test mean scores of the classes in the districts using the other two reading series on inferential comprehension.
>
> Operational H1d: There will be no difference between the MEAP reading-test mean scores of the classes in the district using the most congruent reading series and the MEAP reading-test mean scores of the classes in the districts using the other two reading series on critical reading.
>
> Operational H1e: There will be no difference between the MEAP reading-test mean scores of the classes in the district using the most congruent reading series and the MEAP reading-test mean scores of the classes in the districts using the other two reading series on related study skills.

For the analysis of variance test of significance, the alpha level was set at .10. To reject a null hypothesis, the test required a significant difference for the dependent variables (mean scores on the five categories of MEAP objectives).

Hypothesis II was derived from the fourth research question, which is as follows:

Are methods and strategies used in teaching reading likely to influence the MEAP test results?

The second testable hypothesis was based on the relationship between the teachers' methods and strategies and the MEAP test results. It was anticipated that the teachers in the three districts would differ according to strategies and methods they used to teach reading and that these differences would influence the test scores. The independent variables were the 15 categories of responses to the questionnaire that was administered to the teachers in the three districts and the five types of possible responses. The dependent variables were the mean scores on the 15 categories of responses to the questionnaire.

Analysis of variance was used to compare the teachers' mean scores on the 15 categories of responses to the questionnaire according to districts using the three treatments, or reading series. The scores were assigned the following values: 1 = always, 2 = most of the time, 3 = sometimes, 4 = seldom, and 5 = never.

Hypothesis II stated that there would be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on each of the dependent variables, according to mean scores on teacher-questionnaire responses. To compare the scores of the teachers in the three districts, or treatment groups, on each of the 15 categories of questionnaire responses, this hypothesis was restated as 15 operational hypotheses:

Operational H2a: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on classroom enrichment, according to mean scores on teacher-questionnaire responses.

Operational H2b: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on homework, according to mean scores on teacher-questionnaire responses.

Operational H2c: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on teacher-made and other materials, according to mean scores on teacher-questionnaire responses.

Operational H2d: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on parental involvement, according to mean scores on teacher-questionnaire responses.

Operational H2e: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on time, according to mean scores on teacher-questionnaire responses.

Operational H2f: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on organizational patterns, according to mean scores on teacher-questionnaire responses.

Operational H2g: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on independent learning, according to mean scores on teacher-questionnaire responses.

Operational H2h: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on student interaction, according to mean scores on teacher-questionnaire responses.

Operational H2i: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on outside enrichment, according to mean scores on teacher-questionnaire responses.

Operational H2j: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on basal materials, according to mean scores on teacher-questionnaire responses.

Operational H2k: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on staff support, according to mean scores on teacher-questionnaire responses.

Operational H2l: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on language development, according to mean scores on teacher-questionnaire responses.

Operational H2m: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on written work, according to mean scores on teacher-questionnaire responses.

Operational H2n: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on test uses, according to mean scores on teacher-questionnaire responses.

Operational H2o: There will be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on rewards, according to mean scores on teacher-questionnaire responses.

For the analysis of variance test of significance, the alpha level was set at .10. To reject a null hypothesis, the test required a significant difference for the dependent variables (mean scores on the teachers' responses to the categories of questions on the questionnaire).

Findings of the Study

Introduction

The findings for the first two research questions were based on (1) a comparison of skill-exercise objectives in the three reading series to the objectives tested by MEAP and (2) a comparison of the percentages of reading-series objectives that matched or mismatched the 25 MEAP objectives. The Ginn series, used in District Z, contained 1,327 skill exercises; the Holt, Rinehart, and Winston series, used in District Y, contained 1,094 skill exercises; and the Houghton-Mifflin series, used in District X, contained 1,399 skill exercises.

The findings for Hypothesis I were based on the mean scores of the MEAP reading test for 53 fourth-grade classes within the three districts, or treatment groups. There were 30 classes that used the Ginn treatment; 12 classes that used the Holt, Rinehart, and Winston treatment; and 11 classes that used the Houghton-Mifflin treatment. Classes were not randomly assigned to treatments but were selected according to the district that had mandated the use of the specified series, or treatment, for a minimum of three consecutive years before administering the 1980-81 MEAP Grade Four Reading Test.

The findings for Hypothesis II were based on the mean scores, according to treatment group, of the teacher-questionnaire responses. There were 66 responses from teachers in the Ginn treatment group; 31 responses from teachers in the Holt, Rinehart, and Winston treatment group; and 34 responses from teachers in the Houghton-Mifflin treatment group.

Contingency Tables

Contingency tables were used to answer the first two research questions. Shown in contingency Table 4.1 is the specific number of objectives in each level of the reading series that match MEAP or "other" objectives. (Each of these first- through third-grade levels was designated by title according to publishing company.) In this table we can see that (1) 430 of the Ginn series skill objectives matched MEAP objectives, whereas 830 matched "other" objectives; (2) 418 of the Holt, Rinehart, and Winston series skill objectives matched MEAP objectives, whereas 623 matched "other" objectives; and (3) 828 of the Houghton-Mifflin series skill objectives matched MEAP objectives, whereas 439 matched "other" objectives.

Table 4.2 reveals that, for each reading series, less than 60 percent of all the skill-exercise objectives matched the MEAP objectives. Thirty-two percent of the skill-exercise objectives in the Ginn series matched the MEAP objectives; 38 percent of the skill-exercise objectives in the Holt, Rinehart, and Winston series matched the MEAP objectives; and 59 percent of the skill-exercise objectives in the Houghton-Mifflin series matched the MEAP objectives. In the Ginn series, 63 percent of the skill-exercise objectives matched "other" objectives, whereas 57 percent of the skill-exercise objectives in the Holt, Rinehart, and Winston series and 31 percent of the skill-exercise objectives in the Houghton-Mifflin series matched "other" objectives.

Comparing each of the skill-exercise objectives of the three reading series to the MEAP objectives, the researcher found that the

Table 4.1.--Correlation of MEAP and "other" objectives with objectives of skill exercises found in workbooks and teachers' manuals of three reading series.

MEAP OBJECTIVES	Ginn & Co.								Holt, Rinehart, & Winston								Houghton-Mifflin Co.								
	My Sound and Word.. (2)	A Duck Is a Duck.. (3-4)	May I Come In? (5)	Seven Is Magic (6)	The Dog Next Door.. (7)	How It Is Nowadays (8)	With Skies and Wings (9)		Rhymes and Tales.. (3-6)	A Place for Me (7)	A Time for Friends (8)	People Need People (9)	The Way of the World (10)	Never Give Up (11)	Special Happenings (12)		Tigers (3A)	Lions (3B)	Dinosaurs (3C)	Rainbows (4)	Signposts (5)	Secrets (6)	Rewards (7)	Panorama (8)	Fiesta (9)
VOCABULARY MEANING																									
1. Word meanings affected by prefix							5						4	2	4						1	1	3	2	4
2. Word meanings affected by suffix		1	4		12	4	12				1	2	3	3	3					1	1	3	2	3	3
3. Meanings when word has multi-meanings		2		2	3	2	6		1	2	2	1		1	3					3	5	4	7	6	1
4. Identifying synonyms			1	1	7	8	1					5	1	4	2									7	1
5. Identifying antonyms				4		1	2		1	1		1		2	1					1		1	2	2	
6. Determining meaning based on context	8	10	14	24	19	21	24		14	21	26	15	25	34	36		19	24	24	60	66	64	52	71	47
LITERAL COMPREHENSION																									
7. Identifying main idea of a selection				1	4	2	3		1	1	3	8	3	2	5		11	8	8	1	20	3	1		1
8. Identifying supportive details		3	4	17	10	16	14		4	10	5	9	7	17	13				2	11	3	5	12	6	9
9. Identifying sequence in a selection		2	5	7	4	8	5		7	1		5	2	2	3		1	3	3	9	7	9	4	1	4
10. Identifying cause/effect relationship				1	1	1	2					1	2	3											
11. Identifying stated likeness/diff.			1	1			2						1	3	2						4		2	1	
INFERENTIAL COMPREHENSION																									
12. Inferring main idea of selection				4	6	3	2			1	1	1	4	2	2		1			19	2	3	2	2	2
13. Inferring cause/effect			1			3	4		1																
14. Predicting probable outcome				2	4	2	1				1	1	2	2	2		1		1	2				3	2
15. Inferring details that support main idea		1	1	1	2	2	3						1	4	5		1	3	3				1	4	
16. Inferring sequence					1	2							1										1	1	1
17. Inferring likenesses/differences					1		1																2	5	
18. Drawing conclusions from given info.		2	3	7	8	10	7			1	1	4	2	1	5		4	8	5	15	16	8	12	7	2
19. Identifying relationships (analogies)		1			2								2												
20. Making inferences about characters				2	5	3	3							3	1			1	1				1	1	1
CRITICAL READING																									
21. Determining author's purpose									2				1	1											
RELATED STUDY SKILLS																									
22. Identifying major uses of dictionary/ glossary/contents		1			1		2						1		4										1
23. Locating information in dictionary/ glossary/contents			1	1	1	1	7						2	3	3								1	8	19
24. Summarizing a selection			1	1											1										
25. Alphabetizing to second letter					1	1	1					3	2	3								4	4	4	3
OTHER																									
26. Other	87	134	128	134	164	111	76	113	67	83	89	104	95	72	21		23	23	45	59	79	71	61	57	

match between the Houghton-Mifflin series objectives and the MEAP objectives was 27 percent greater than the one between the Ginn series objectives and the MEAP objectives. The match between the Houghton-Mifflin series objectives and the MEAP objectives was 21 percent greater than the one between the Holt, Rinehart, and Winston series objectives and the MEAP objectives. The Houghton-Mifflin series objectives, therefore, were most congruent with the MEAP objectives.

Table 4.2.--Distribution of objectives for skill exercises across reading series.

Objectives	Ginn and Co. Skill-Exercise Objectives	Holt, Rinehart, and Winston Skill-Exercise Objectives	Houghton-Mifflin Skill-Exercise Objectives
MEAP objectives	430 (32%)	418 (38%)	828 (59%)
"Other" objectives	834 (63%)	623 (57%)	439 (31%)
Exercises not rated	63	53	139
	(N=1,327)	(N=1,094)	(N=1,399)

One-Way Univariate Analyses of Variance

One-way univariate analyses of variance were used to answer Research Questions 3 and 4. This statistical treatment was used (1) to compare mean reading-test scores on the various categories of MEAP objectives according to treatment groups and (2) to compare mean

scores on the various categories of questionnaire responses according to treatment group.

As shown in Table 4.3, Operational Hypotheses 1a, 1b, 1c, and 1e were accepted. There was no difference between the MEAP reading-test mean scores of the classes in the district using the most congruent reading series and the MEAP reading-test mean scores of the classes in the districts using the other two reading series on vocabulary meaning, literal comprehension, inferential comprehension, and related study skills.

Table 4.3.--Analysis of variance for MEAP test scores by category of MEAP objectives.

Category	MS Between	MS Within	F	Sig.
1 Vocabulary meaning	17.9361	32.6575	.5492	.58
2 Literal comprehension	12.8852	30.2280	.4263	.66
3 Inferential comprehension	27.1254	36.5209	.7427	.48
4 Critical reading	237.1645	86.2719	2.7490	.07*
5 Related study skills	26.7338	65.1603	.4103	.67

Between-groups df = 2; within-groups df = 50.

As indicated in Table 4.3, the only category of MEAP objectives that was statistically significant was Category 4, critical reading. Operational Hypothesis 1d was rejected; there was a difference

in the MEAP reading-test mean scores of the classes in the district using the most congruent reading series and the MEAP reading-test mean scores of the classes in the districts using the other two reading series on critical reading. The significance level for Category 4, critical reading, was .07.

As shown in Table 4.4, the Houghton-Mifflin treatment group (of the three, the one most congruent with MEAP) did not have the highest mean score on Category 4, critical reading. The Holt, Rinehart, and Winston treatment group had the highest mean score on MEAP Category 4; the Ginn treatment group and the Houghton-Mifflin treatment group had the second and third highest scores, respectively.

Table 4.4.--MEAP test mean scores on objective Category 4 for the three treatment groups.

MEAP Category 4	Ginn and Co. (Z) $\overline{X}$	Holt, Rinehart, and Winston (Y) $\overline{X}$	Houghton-Mifflin (X) $\overline{X}$
Critical reading	74.8	77.36	68.75

Population means: 73.96.

To examine the number of skill-exercise objectives that matched MEAP objectives, the researcher used Contingency Table 4.5. This table provided a summary of the number of skill-exercise objectives that matched MEAP Category 4.

Table 4.5.--Correlation of MEAP objective Category 4, critical reading, with number of matching reading-series objectives.

MEAP Category 4	Ginn and Co. (Z)	Holt, Rinehart, and Winston (Y)	Houghton-Mifflin (X)
Critical reading	0	4	0

Correlating the mean scores in Table 4.4 with the number of skill-exercise objectives in the three reading series that matched MEAP Category 4 in Table 4.5, the researcher found that only the Holt, Rinehart, and Winston series had skill-exercise objectives that matched the MEAP objective on Category 4. The Holt, Rinehart, and Winston treatment group was expected to score highest on this category of MEAP objectives, but why the Ginn treatment group scored higher than the Houghton-Mifflin treatment group on Category 4 is unclear because neither used reading series with skill-exercise objectives that matched the category of MEAP objectives that was tested.

To enable the researcher to examine variables other than the reading series that might have influenced the MEAP test scores, a questionnaire was administered to the first- through third-grade teachers in the three treatment groups, or districts. (See Appendix A.) Questionnaire responses ranged in value from 1 to 5, with a response of 1 representing "always" and a response of 5 representing "never." Lower numbers indicated greater emphasis placed on the subject matter.

Table 4.6.--Analysis of variance for teachers' scores on questionnaire categories in Districts Z, Y, and X.

Category	MS Between	MS Within	F	Sig.
1 Classroom enrichment	.32	.19	1.60	.20
2 Homework	.58	.64	.91	.40
3 Teacher-made and other materials	.63	.34	1.85	.16
4 Parental involvement	.74	.33	2.24	.11
5 Time	3.01	.37	8.18	.0005*
6 Organizational pattern	.67	.17	4.02	.02*
7 Independent learning	4.25	.47	9.02	.0002*
8 Student interaction	1.71	.61	2.81	.06*
9 Outside enrichment	.30	.26	1.16	.32
10 Use of basal materials	.1032	.1116	.72	.40
11 Staff support	49.03	1.27	38.62	.0000*
12 Language development	2.47	.27	9.20	.0002*
13 Written work	1.59	.42	3.78	.026*
14 Test uses	.78	.59	1.31	.27
15 Rewards	4.48	1.33	3.37	.038*

Between-groups df = 2; within groups df = 118.

As shown in Table 4.6, Operational Hypotheses 2a, 2b, 2c, 2d, 2i, 2j, and 2n were accepted. There was no difference in the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the

districts using the other two reading series on the following: classroom enrichment, homework, teacher-made and other materials, parental involvement, outside enrichment, use of basals, and test uses.

As shown in Table 4.6, eight categories of questionnaire responses were statistically significant: Categories 5, 6, 7, 8, 11, 12, 13, and 15. Operational Hypotheses 2e, 2f, 2g, 2h, 2k, 2l, 2m, and 2o were rejected. These categories of questionnaire responses were statistically significant according to the analysis of variance test of significance. The alpha level was set at .10.

The eight categories of questionnaire responses that were statistically significant are listed in Table 4.7. Discussion of the findings for each of these categories follows.

As shown in Table 4.7, District Y teachers devoted more time to reading than did District X and Z teachers. This means they spent more time for preparation and instruction, as well as for teaching and discussing basic skills with their students.

Teachers in District Z emphasized organizational patterns more than did teachers in Districts X and Y. Organizational patterns include structured programs, team teaching, and individualization; they also include grouping by ability, seating students in rows, and having students remain in the same reading group all year. (For teachers to have indicated that they individualized and immobilized students of the same group was contradictory, so these two questions in the category termed organizational patterns were considered invalid as answered.)

Table 4.7.--Teachers' mean scores on the statistically significant questionnaire categories for Districts Z, Y, and X.

Questionnaire Category	District Z	District Y	District X
5 Time	2.23	2.14	2.70
6 Organizational pattern	2.67	2.78	2.92
7 Independent learning	2.28	2.64	2.89
8 Student interaction	3.40	3.75	3.26
11 Staff support	2.21	1.58	4.03
12 Language development	2.33	2.35	2.79
13 Written work	2.30	1.94	2.15
15 Rewards	2.76	2.38	2.13

Independent learning, Category 7, was emphasized more by teachers in District Z than by those in Districts X and Y. This category includes learning that is self-initiated, reading in the content areas, using library resources in a variety of ways, requesting explicit directions, choosing reading matter, and choosing the assignment to complete first.

Student interaction received greater emphasis in District X than in the other two districts. There was more peer tutoring and older-student tutoring in this treatment group than in the others.

District Y had more staff support for both students and teachers. This district provided more aid from specialists, more help for students by the classroom teacher, and more inservice training programs for teachers when a new reading program was adopted.

Teachers in District Z placed more emphasis on language development than teachers in the other two districts. Language-development questions referred mainly to telling stories and to discussing information and ideas. (See Question 47-49, p. 126.)

Teachers in District Y more often required their students to examine and correct their own written assignments than did teachers in the other two districts.

Teachers in District X more often provided rewards, such as treats or prizes, for doing good work than did teachers in Districts Z and Y.

In summary, the MEAP test results revealed that the Holt, Rinehart, and Winston treatment group had the highest mean score on the only variable that was statistically significant: Category 4, critical reading. This might have been attributable to two factors. First, this group's treatment, or reading series, was the only treatment that had any skill-exercise objectives that matched the MEAP objective in the critical-reading category. Second, the teachers from this group perceived themselves as devoting more time to all aspects of reading than did the teachers from the other groups; they also perceived themselves as stressing written work more and as having more staff support for both teachers and students.

The Ginn treatment group had the second highest mean score on Category 4, critical reading. Their treatment, or reading series, did not contain any skill-exercise objectives that correlated with the

MEAP objective in Category 4. Their teachers did, however, stress independent learning, language development, and a more structured organizational pattern than did teachers in the other two groups.

The Houghton-Mifflin treatment group had the lowest mean score on Category 4, critical reading. Their reading series did not contain any skill-exercise objectives that correlated with the MEAP objective in Category 4. The teachers in this group stressed only two of the statistically significant categories of questionnaire responses. These teachers perceived themselves as emphasizing student interaction and student rewards more than did the teachers in the other two treatment groups.

Summary

Contingency tables were used (1) to illustrate the relationship between the objectives of the skill exercises found in the workbooks and teachers' manuals of the three reading series and the objectives of the MEAP test and (2) to illustrate what percentage of the skill-exercise objectives of each reading series corresponded with MEAP objectives and to facilitate the comparison of these percentages.

The data showed that 430 of the Ginn series skill objectives; 418 of the Holt, Rinehart, and Winston series skill objectives; and 828 of the Houghton-Mifflin series skill objectives matched the MEAP objectives. Thirty-two percent of the skill-exercise objectives in the Ginn series matched the MEAP objectives, whereas 38 percent of the skill-exercise objectives in the Holt, Rinehart, and Winston series and 59 percent of those in the Houghton-Mifflin series matched the

MEAP objectives. The Houghton-Mifflin series skill-exercise objectives most nearly matched the MEAP objectives; they were, therefore, most congruent with MEAP objectives.

Analysis of variance tests of significance were used (1) to compare, by categories of objectives, mean scores of the treatment groups on the MEAP Grade Four Reading Test and (2) to compare, by treatment group, mean scores on categories of teacher-questionnaire responses. The alpha level for all ANOVA tests was set at .10.

Operational Hypotheses 1a, 1b, 1c, and 1e were accepted. There was no difference between the MEAP reading test mean scores of the classes in the district using the most congruent reading series and the MEAP reading test mean scores of the classes in the districts using the other two reading series on vocabulary meaning, literal comprehension, inferential comprehension, and related study skills.

Operational Hypothesis 1d was rejected. There was a significant difference among the three groups on critical reading at the .07 significance level.

Operational Hypotheses 2a, 2b, 2c, 2d, 2i, 2j, and 2n were accepted. There was no difference in the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on classroom enrichment, homework, teacher-made and other materials, parental involvement, outside enrichment, use of basals, and test uses.

Operational Hypotheses 2e, 2f, 2g, 2h, 2k, 2l, 2m, and 2o were rejected. There was a difference in the methods and strategies

of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on time, organizational pattern, independent learning, student interaction, staff support, language development, written work, and rewards. The mean scores on these categories of questions were significantly different, according to the analysis of variance test of statistical significance. The alpha level was set at .10.

It was anticipated that the Houghton-Mifflin treatment group would have the highest mean scores on the five categories of MEAP objectives because their treatment, or reading series, was the most congruent with MEAP objectives. However, the Holt, Rinehart, and Winston treatment group had the highest mean score on the only statistically significant category of MEAP objectives; and the Holt, Rinehart, and Winston series was the only one that had skill-exercise objectives that matched MEAP objectives in Category 4, critical reading. The teachers from the Holt, Rinehart, and Winston group perceived themselves as devoting more time to all aspects of reading than did teachers from the other groups. These teachers also perceived themselves as stressing written work more and as having more staff support than did the teachers from the other two groups.

CHAPTER V

CONCLUSIONS AND NEED FOR FURTHER RESEARCH

This chapter begins with a summary of this study. Then the following topics are presented: Conclusions of the Study, Discussion of the Findings, and Need for Further Research.

Summary

Statement of the Problem

This study compared the 25 MEAP minimal reading objectives with the objectives represented in grades one through three basal reading exercises of the three most widely used series in Michigan, as determined by Market Data Retrieval, Inc.,[1] and Nicholas P. Criscuolo.[2] The researcher attempted to ascertain the following:

1. How many objectives of the reading-skill exercises found in workbooks and teachers' manuals match MEAP objectives?

2. What reading series has the greatest percentage of skill-exercise objectives that are congruent with MEAP objectives, as determined by a panel of reading experts?

---

[1]Market Data Retrieval, Inc., Reading K-8 Survey, HM Co. Market Research Report No. 17 (New York: Market Data Retrieval, Inc., 1977), p. 97.

[2]Nicholas P. Criscuolo, Improving Classroom Reading Instruction (Worthington, Ohio: Charles A. Jones Publishing Co., 1973), p. 23.

3. Will the degree of congruence between reading-exercise objectives and MEAP objectives affect MEAP test results?

4. Are methods and strategies used in teaching reading likely to influence the MEAP test results?

Need for the Study

The need for this study was based on the premise that the curriculum, or course of study, affects what students learn. Educators need to address the increasing concern about the quality of education by demonstrating that certain skills--those that meet the objectives of the school system as well as of the community--are being developed through the use of selected curricular materials. Of course, achievement cannot be expected on topics that are tested but not taught; a high correlation between what is taught and what is tested is essential if tests are to measure adequately what is taught. If the correlation between what is taught and what is tested is low, the test is insensitive to achievement gains resulting from the curriculum. There was a need, then, to establish the relationship between the MEAP objectives and the objectives of reading-skill exercises found in workbooks and teachers' manuals of the reading series used in teaching minimal concepts and skills.

Review of the Literature

A review of the literature revealed that the back-to-basics movement and the call for accountability have given rise to concerns about the capacity of testing instruments to assess accurately what is taught in schools. The Michigan State Department of Education

established an accountability model that was designed to insure that students receive instruction in basic skills.

Reading is a basic skill on which the success of schools depends because reading is essential for learning most other content. Teachers of the primary grades in schools throughout the United States use basal-reading-series textbooks, teachers' manuals, and workbooks 85 to 90 percent of the time.

Criterion-referenced tests are gaining in popularity as a tool for assessing the achievement of specified objectives of instructional programs. The criterion-referenced test developed by MEAP officials was intended to be consistent with the instructional content and purposes of basic education in Michigan schools. This test, however, focuses on specific objectives that are independent of any particular reading series that is used.

Research has shown that testing and test results are dependent upon the type of measurement instrument used. Some researchers have concluded that the only valid way to assess the effects of a treatment is to use tests that parallel what is taught. Such tests are criterion referenced and focus on specific objectives that permit educators to ascertain whether a student possesses a particular competency. For the criterion-referenced test to be well constructed, the test items must represent the same domain as the instructional exercises. Criterion-referenced tests should measure an individual's level of proficiency, providing information on mastery or nonmastery that can be used to diagnose deficits or ascertain strengths.

Method and Design

The three reading series used in the three districts were the Ginn and Company reading series; the Holt, Rinehart, and Winston reading series; and the Houghton-Mifflin reading series. The objectives of the skill exercises found in the workbooks and teachers' manuals of the reading series were determined by a panel of three reading experts. Percentages of skill-exercise objectives that matched MEAP objectives were examined to determine the congruence of the reading-series objectives with the MEAP objectives.

The sample for the study consisted of 53 fourth-grade classes in three districts that used three different reading series for primary-level reading. District Z had 30 fourth-grade classes that used the Ginn reading series; District Y had 12 fourth-grade classes that used the Holt, Rinehart, and Winston reading series; and District X had 11 fourth-grade classes that used the Houghton-Mifflin reading series.

The three instruments used in this study were (1) a list of MEAP objectives, (2) Classroom Listing Reports, and (3) a primary-school-teachers questionnaire.

A 23 x 26 contingency table was designed to determine how frequently skill-exercise objectives from workbooks and teachers' manuals of the three reading series intersected with MEAP and "other" objectives. A second contingency table in a 3 x 3 design was prepared to illustrate the percentage of skill-exercise objectives that intersected with MEAP objectives, "other" objectives, or exercises "not rated" for lack of expert consensus.

A 3 x 5 design was used for comparing the five categories of the MEAP test scores for the three treatment groups, or districts. One-way univariate analysis of variance was used for comparing the three treatment groups' mean scores on the five categories of the MEAP test. In each of the univariate analysis of variance tests, the alpha level of significance was set at .10.

To determine whether teachers' methods and strategies might have influenced test results in each of the three districts, a questionnaire was administered. A 3 x 15 design was used for comparing teachers' responses in the three districts according to the mean scores on 15 categories of responses to the teachers' questionnaire. The primary statistic for comparing the categories of responses was one-way univariate analysis of variance. In each of the univariate analysis of variance tests, the alpha level of significance was set at .10.

The first hypothesis was that there would be no difference between the MEAP reading-test mean scores of classes in the district using the most congruent reading series and the mean scores of the classes in the districts using the other two reading series on each of the dependent variables (mean scores on categories of objectives). This hypothesis was divided into five operational hypotheses to make comparisons among the three districts on the five categories of the MEAP test.

The second hypothesis was that there would be no difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of

teachers in the districts using the other two reading series on each of the dependent variables (mean scores on categories of teacher-questionnaire responses). This hypothesis was divided into 15 operational hypotheses to make comparisons among the three districts on the 15 categories of the questionnaire.

Analysis of the Data

The contingency tables showed (1) the frequency with which reading-series objectives matched MEAP objectives and (2) the percentage of objectives in each reading series that matched MEAP objectives. Out of 1,327 exercises, the Ginn series had 430 skill-exercise objectives that matched MEAP objectives, whereas the Holt, Rinehart, and Winston series had 418 out of 1,094 exercises and the Houghton-Mifflin series had 828 out of 1,399 exercises. Thirty-two percent of the Ginn series skill-exercise objectives matched MEAP objectives; 38 percent of the Holt, Rinehart, and Winston series skill-exercise objectives matched MEAP objectives; and 59 percent of the Houghton-Mifflin series skill-exercise objectives matched MEAP objectives. The Houghton-Mifflin series objectives were found to be the most congruent with MEAP objectives.

Operational Hypotheses 1a, 1b, 1c, and 1e were accepted. There was no difference between the MEAP reading-test mean scores of the classes in the district using the most congruent reading series and the MEAP reading-test mean scores of the classes in the districts using the other two reading series on vocabulary meaning, literal comprehension, inferential comprehension, and related study skills.

Operational Hypothesis 1d was rejected. There was a difference between the MEAP reading-test mean scores of the classes in the district using the most congruent reading series and the MEAP reading-test mean scores of the classes in the districts using the other two reading series on critical reading. (The significance level for this hypothesis was .07.)

Operational Hypotheses 2a, 2b, 2c, 2d, 2i, 2j, and 2n were accepted. There was no significant difference between the methods and strategies of teachers in the district using the most congruent reading series and the methods and strategies of teachers in the districts using the other two reading series on classroom enrichment, homework, teacher-made and other materials, parental involvement, outside enrichment, use of basals, and test uses, according to mean scores on teacher-questionnaire responses.

Operational Hypotheses 2e, 2f, 2g, 2h, 2k, 2l, 2m, and 2o were rejected. There was a difference in time at the .0005 significance level, a difference in organizational pattern at the .02 significance level, a difference in independent learning at the .002 significance level, a difference in student interaction at the .06 significance level, a difference in staff support at the .0000 significance level, a difference in language development at the .0002 significance level, a difference in written work at the .026 significance level, and a difference in rewards at the .038 significance level. Teachers from the Ginn treatment group in District Z emphasized organizational patterns, independent learning, and language development more than did teachers in the other districts. Teachers from the Holt, Rinehart,

and Winston treatment group in District Y were strongest in time, written work, and staff support. Teachers from the Houghton-Mifflin treatment group in District X stressed student interaction and rewards more than did the teachers in the other two districts.

Discussion of Findings

Less than 60 percent of all the skill-exercise objectives in the three basal series correlated with the MEAP objectives. This does not mean that the objectives of either are inappropriate. The MEAP objectives are minimal, whereas many other objectives of reading are not. For example, skimming, rate of reading, higher-level reasoning, and creativity are important outcomes of reading that are generally considered to be broader objectives than those identified as essential components for further learning.

The Houghton-Mifflin series had the most skill exercises and had nearly twice as many skill exercises with objectives that correlated with MEAP objectives as did the other two groups. However, the percentage of reading-series skill-exercise objectives that correlated with MEAP objectives did not positively influence the test results 80 percent of the time, nor did the Houghton-Mifflin treatment group have the highest mean score on any of the five categories of MEAP objectives tested. Four out of five of the mean scores on the categories of MEAP objectives did not prove to be statistically significant. This indicated that the greatest amount of exposure, through the reading series, did not most positively affect test results.

The disproportionate size of the treatment groups might have made a difference in test scores: there were 11 classes in the Houghton-Mifflin treatment group, 12 classes in the Holt, Rinehart, and Winston treatment group, and 30 classes in the Ginn treatment group. The Ginn treatment group was much larger than the other two groups; the larger the group tested, the more likely it is that testing errors will be balanced--with some children scoring below their "true" scores and others scoring above. When a greater number of scores is used, this provides a wider range from which to generate mean scores. This may explain why the Ginn treatment group scored higher than did the Houghton-Mifflin treatment group. This, however, does not explain why the Holt, Rinehart, and Winston treatment group scored the highest, since this was the second smallest group of classes.

The Primary-School Teachers' Questionnaire was administered to ascertain some of the other variables that might have influenced the MEAP test scores. It was noted that the Houghton-Mifflin treatment group had the lowest mean score on critical reading, the only statistically significant category of MEAP objectives; in keeping with these findings, the teachers in this group had the weakest scores on six out of eight of the statistically significant categories of responses to the questionnaire. The teachers in this group indicated that they spent less time planning and teaching reading, had less-structured organizational patterns, had fewer independent learning activities, had considerably less staff support, and had fewer language-development lessons than did teachers in the other two

groups. Also, these teachers did not require students to examine and correct their written assignments as often as did the Holt, Rinehart, and Winston group's teachers. Inasmuch as these variables did not seem to affect vocabulary meaning, literal and inferential comprehension, and related study skills, they might have had more effect on the critical reading category than on the other four MEAP categories.

The two statistically significant categories of the questionnaire on which the teachers from the Houghton-Mifflin treatment group scored the strongest were student interaction and rewards. A greater percentage of teachers from this group perceived that their students received more tutoring from peers and other students and were given more treats and prizes when they performed well.

The Ginn treatment group had the second highest mean score on critical reading, although this group did not have any skill exercises with objectives that correlated with the MEAP objectives for Category 4, critical reading. Their teachers, however, stressed independent learning, language development, and a more structured organizational pattern than did teachers in the other groups. These variables might have been more important than student interaction and rewards in learning to read critically, thereby affecting the mean scores more in the critical-reading category than in the other four categories.

The MEAP test results revealed that the Holt, Rinehart, and Winston treatment group had the highest mean score on the only variable that was statistically significant: Category 4, critical reading. This might have been attributable to the reading series used, since

this group used the only reading series that had any skill-exercise objectives that matched the MEAP objectives for the critical-reading category. Too, there were three questionnaire variables that might have influenced the test results in this category. The mean scores on the questionnaire showed that teachers from this group perceived that they devoted more time to all aspects of reading than did teachers from the other groups. These teachers also perceived that they stressed written work more and had more staff support than did the teachers from the other two treatment groups. Written work, time, and staff support might have had greater effect on critical-reading scores than on the scores for the other MEAP reading-test categories.

In summary, although there was no difference in the MEAP test results on four out of the five categories of MEAP objectives, one reading series differed considerably on the number of objectives that matched MEAP objectives. The Houghton-Mifflin series had more skill exercises than did the other two series, and it also had the highest percentage of skill-exercise objectives that matched MEAP objectives, but this treatment group did not have the highest scores on the MEAP test.

Although the Houghton-Mifflin treatment group's reading-series objectives were the most congruent with MEAP objectives, its mean scores on the MEAP test were not the highest of the three groups. It is assumed that either the variables that were considered or other confounding variables affected the test results. Some of the variables that were discussed in this chapter were quality of instruction,

teachers' methods and strategies, students' abilities, testing conditions, mechanical errors, and disproportionately sized treatment groups. Also, the gross reporting of scores might have made differences in MEAP test results indiscernible.

The results of this study can be generalized only to the three populations or treatment groups. This study suggested that exposure to the objectives of reading and/or to some methods and strategies of teachers might have influenced the test scores for critical reading. There was not enough evidence to determine which variables had the greatest effect on the overall reading process or on specific aspects of reading such as word meaning, comprehension, and related study skills. If one knew which variables contributed most to beginning reading, it would be possible to delineate other variables and to ascertain their causes and effects in various student populations.

Conclusions of the Study

Conclusions were derived from the analysis of the three types of data generated in the study. Contingency tables were used to compare the reading-series objectives found in workbooks and teachers' manuals with the MEAP objectives; they were also used to make comparisons among the skill-exercise objectives of the three reading series according to the proportion of matches and mismatches across the 25 MEAP objectives. One-way univariate analysis of variance was used to determine (1) the significance levels of mean scores on the MEAP reading test by categories of MEAP objectives and (2) the significant differences among the mean scores of responses to the 15 categories of questions on the teachers' questionnaire.

There was a difference between the skill-exercise objectives of each of the three reading series and the MEAP objectives. Although the Ginn and Houghton-Mifflin series had approximately 300 more exercises than the Holt, Rinehart, and Winston series, the two series had nearly the same number of skill exercises with objectives that matched MEAP objectives. The Houghton-Mifflin series had almost twice as many skill exercises with objectives that matched MEAP objectives as did the other two series. Four hundred thirty out of a total of 1,327 Ginn-series skill exercises had objectives that matched MEAP objectives; 418 out of a total of 1,094 Holt, Rinehart, and Winston series skill exercises had objectives that matched MEAP objectives; and 828 out of a total of 1,399 Houghton-Mifflin series skill exercises had objectives that matched MEAP objectives.

Comparing the three reading series across the MEAP objectives revealed that 59 percent of the Houghton-Mifflin series objectives matched MEAP objectives, but only 32 percent of the Ginn series objectives and 38 percent of the Holt, Rinehart, and Winston series objectives matched MEAP objectives. The Houghton-Mifflin series objectives were the most congruent with MEAP objectives overall. However, the Holt, Rinehart, and Winston series was the most congruent with the MEAP objectives in Category 4, critical reading, which yielded the only statistically significant variable in the ANOVA tests.

Since the Houghton-Mifflin series objectives were the most congruent with MEAP objectives overall, it was expected that this treatment group would have the highest MEAP reading-test scores. However, this was not the case. The mean scores on the MEAP test

revealed a significant difference among the treatment groups on only one category of MEAP objectives--Category 4, critical reading, on which the Holt, Rinehart, and Winston treatment group scored highest. The Ginn treatment group scored second highest, and the Houghton-Mifflin treatment group scored third.

Other variables that might have made a difference in the MEAP results were ascertained from a teachers' questionnaire. The mean scores on the 15 categories of responses to this questionnaire were statistically significant on eight of the categories. The teachers from the Ginn treatment group in District Z had the strongest mean scores on Categories 6, 7, and 12: organizational pattern, independent learning, and language development. The teachers from the Holt, Rinehart, and Winston treatment group in District Y had the strongest mean scores on Categories 5, 11, and 13: time, staff support, and written work. The teachers from the Houghton-Mifflin treatment group in District X had the strongest mean scores on Categories 8 and 15: student interaction and rewards.

The mean scores on the 15 categories of responses to the questionnaire were not statistically significant in seven of the categories. These categories included classroom enrichment, homework, teacher-made and other materials, parental involvement, outside enrichment, use of basals, and test uses.

Reflections From the Study

Although the scope of this study was limited to specific reading series and the MEAP test, it generated many questions that have not been answered or that are debatable. Because of the

complexity of the beginning reading process, it is not an easy field to study.

There is a need for sorting out and carefully weighing educational practices that should promote further learning. Based on this study, some of the questions that require careful consideration are as follows:

1. Are criterion-referenced tests the best measures of success in school?

2. What are some ways to adequately assess transfer of learning?

3. How can adequacy of exposure to a learning process be determined?

4. Are there other evaluative procedures that are better measures of learning than are criterion-referenced tests?

5. Should teacher-training programs address ramifications of competency testing?

6. Should objectives that will be tested in various commercial testing programs be identified and/or stressed in basal reading materials?

7. Should teacher accountability be dependent on student test results?

8. What are some of the major effects of tests on teaching?

9. Is it possible, through criterion-referenced testing, to diagnose specific problems that children have in reading even though these problems have come from a number of different sources?

10. Can the effectiveness of a program be determined without measures of the affective domain? What are the implications and justifications of testing in the affective areas?

Need for Further Research

Analyses of the relationship between reading theories and MEAP objectives were conducted using the five categories of objectives on the MEAP reading test and all of the skill exercises found in workbooks and teachers' manuals of each of the reading series for grades one through three. Based on the fact that congruence between basal reading materials and test content did not affect test results, other concepts concerning appropriate learning procedures and teaching techniques need to be examined. A number of possibilities for further research have grown out of these findings and from reflections on this study. These include:

1. A random-block design could be used for each district to analyze data for more homogeneous groups--that is, for students with similar abilities, interests, strengths, or weaknesses. These variables might influence how and what students learn.

2. A stratified sample would insure that certain subgroups in the population would be included in each sample. In this method, students could be grouped in three levels on the basis of general intelligence; this would permit comparisons among subgroups of similar abilities.

3. Randomly selecting a sample from each of the classes within the districts would insure a balanced sample size that could be small enough to eliminate gross reporting of scores.

4. Demographic data could be gathered to determine the degree of similarity or difference among students in the three districts. These data could include the I.Q.; levels of students with high, medium, and low abilities; the number of special-education students; and/or the number of Title I students in each group.

5. Teacher interviews could be used to ascertain (1) the number of MEAP objectives taught in classes other than reading, (2) the amount of time spent on MEAP objectives, (3) what evaluative procedures other than criterion-referenced tests are used, (4) teacher and student attitudes toward testing, and (5) affective reading goals of the teacher.

6. This study could be replicated with the use of more recent literature and basal-reading series that would reflect current knowledge concerning the reading process.

7. Classroom observations might be used to determine the effect of teacher methods and strategies on the beginning reading process today, inasmuch as great advances have been made in determining how students learn to read.

In conclusion, to assure that evaluation does not become the major concern in teaching, it is vital to focus on a variety of variables that affect the reading process. Many variables need to be considered, and the most important ones need to be closely delineated so that the effects of instructional programs can more easily be ascertained. Inasmuch as reading authorities do not agree on which variables have the greatest influence on the beginning reading process, it is important continually to explore more and different possibilities.

APPENDICES

APPENDIX A

QUESTIONNAIRE DATA

September 1981

Dear pilot-test volunteer,

To prepare for my dissertation survey, I need your response to the enclosed questionnaire. The purpose of this pilot-test is not to acquire data, but to enable me critically to evaluate the questionnaire itself, which will be used later to gather data reflecting various sources that may contribute to student achievement in reading.

The questionnaire assumes that you are a primary-school teacher who is responsible for teaching reading on a regular basis. Please follow these directions:

1. Time yourself to ascertain how long it takes you to complete the questionnaire; answer the questions at one sitting.

2. Correct any errors in spelling.

3. Comment on confusing directions and/or questions. Comment on the organization of the questionnaire, its wording, etc.--on anything that might confuse teachers.

4. Note anything you feel would improve the questionnaire.

Your input is critical in my determining needed changes in the survey instrument. Please be honest; your evaluation is important!

You will find an additional sheet of paper for your comments. Please complete both the questionnaire and the comment pages and return them to me as soon as possible.

Thank you for your help. It is greatly appreciated.

Sincerely,

Phyllis H. Music

PRIMARY-TEACHERS' READING QUESTIONNAIRE

GENERAL DIRECTIONS:

For research purposes, it is important that you circle one response for each question (of this questionnaire). Mark your choices clearly in pencil. Erase completely when necessary.

PART I

This section is regarding background information. Circle only 1 answer for each.

1. Grade teaching (1st, 2nd, 3rd, combination of these)
2. Years of teaching (0-3, 4-7, 8-11, 12-more)
3. Years of teaching at the present grade level (0-3, 4-7, 8-11, 12-more)
4. Number of grades in the building (K-3, K-4, K-5, K-6, other _____)
5. Basal reading series used in your building through 1979-80 school year) (Houghton-Mifflin, Scott Foresman, Holt, Ginn, other _____)
6. Copyright date of series indicated in number 5 (1969-70, 1971-2, 1973-4, 1975-6, 1977-8, other _____)
7. My students have (many, few, no) opportunities for educational experiences or enrichment activities outside of the school environment.

PART II.

Primary Reading Instruction

Below are a variety of statements related to various strategies and methods used in teaching reading. Consider your own methods. Circle one choice for each of the following statements that would identify your own particular methods and/or classroom situation.

1 = Always 2 = Most of the time 3 = Sometimes 4 = Seldom 5 = Never

1.	Instructional bulletin boards are used as a part of my teaching method.	1 2 3 4 5
2.	Reading centers are used to stimulate and motivate students to do reading activities.	1 2 3 4 5
3.	I read aloud to my students.	1 2 3 4 5
4.	I share my own personal stories with children.	1 2 3 4 5
5.	A variety of hands-on activities are provided in my room.	1 2 3 4 5
6.	I use games and simulations to teach reading.	1 2 3 4 5
7.	Students are required to do reading homework for my class.	1 2 3 4 5
8.	Homework is used as a means for students to finish assignments that were begun in class.	1 2 3 4 5
9.	I use teacher-made materials to teach reading.	1 2 3 4 5
10.	I use a variety of other resources as well as the basal reading program (SRA, Reader's Digest, commercially prepared materials, etc.).	1 2 3 4 5
11.	I use enrichment activities that are not suggested in the teachers' guides.	1 2 3 4 5
12.	I work on reading skills that are not included in the instructional reading program.	1 2 3 4 5
13.	The students' parents are willing resources for my class.	1 2 3 4 5
14.	The parents evidence concern for their children's progress.	1 2 3 4 5
15.	The parents contribute to their children's foundations in basic skills.	1 2 3 4 5

16.	At least 50 minutes is allotted for each reading instruction period.	1 2 3 4 5
17.	More time is spent on teaching basic skills than on the application of newly acquired skills.	1 2 3 4 5
18.	I spend more time discussing ideas with students than instructing students.	1 2 3 4 5
19.	I spend more than 30 minutes daily in preparation for the teaching of reading.	1 2 3 4 5
20.	Structured programs work better than less-structured ones for me.	1 2 3 4 5
21.	In my room, students are grouped according to ability for instructional purposes.	1 2 3 4 5
22.	The seating arrangement is by rows.	1 2 3 4 5
23.	There are more than 27 students in my group.	1 2 3 4 5
24.	We use team teaching to teach reading.	1 2 3 4 5
25.	My students remain the same reading group for the whole school year.	1 2 3 4 5
26.	I use an individualized approach to reading.	1 2 3 4 5
27.	Older students come into my room to tutor the younger ones.	1 2 3 4 5
28.	Classmates are encouraged to tutor each other.	1 2 3 4 5
29.	My students spend their free time outside of school watching television.	1 2 3 4 5
30.	There is evidence that my students read at home as well as at school.	1 2 3 4 5
31.	My students participate in extra-curricular activities such as scouts, after-school sports, music, etc.	1 2 3 4 5
32.	I emphasize and encourage self-initiated learning.	1 2 3 4 5
33.	My students read in the content areas during reading class time.	1 2 3 4 5
34.	My students use the library resources for a variety of activities.	1 2 3 4 5

35.	My students use the library as an informational resource.	1 2 3 4 5
36.	My students require detailed directions before they can begin assignments.	1 2 3 4 5
37.	The students are permitted to choose the books that they wish to read.	1 2 3 4 5
38.	Students are permitted to choose the order in which they prefer to cover classroom assignments.	1 2 3 4 5
39.	I use the basal reading program that has been adopted by my school system.	1 2 3 4 5
40.	I rely on the teacher's manual for instructional ideas and information.	1 2 3 4 5
41.	Workbooks are used as a part of my instructional program.	1 2 3 4 5
42.	Workbooks are used with all reading groups that I teach.	1 2 3 4 5
43.	I use the enrichment activities that are suggested in the teachers' guides.	1 2 3 4 5
44.	I provide extra help for students who need it during breads, after school, at recess time, etc.	1 2 3 4 5
45.	A reading specialist works with some of my students.	1 2 3 4 5
46.	My students are encouraged to share stories that they have read.	1 2 3 4 5
47.	My students are encouraged to share information that they have read.	1 2 3 4 5
48.	More time is spent discussing reading ideas and stories than doing written activities.	1 2 3 4 5
49.	Students are required to examine their written work to find and correct their mistakes.	1 2 3 4 5
50.	I use standardized test scores to establish reading groups.	1 2 3 4 5
51.	I use test information to alter my instructional program.	1 2 3 4 5
52.	My students are rewarded for doing good work (prizes, treats, etc.).	1 2 3 4 5

COMMENTS

Time required to complete the questionnaire: _____

Other comments....

Please be sure to complete and return to me as soon as possible. Thank you!

Phyllis Music

Survey Instrument Changes

A reading specialist was consulted for recommendations on how to improve the questionnaire. She recommended, for Part I, adding Question 8 and changing Question 7 to a continuum form; too, she recommended adding Questions 44, 54, and 55 in Part II.

As a result of the pilot-test, the following changes were made:

1. "Stories" was changed to "experiences" in Question 4.
2. "Evidence" was changed to "show" in Question 14.
3. "Contribute to" was changed to "help" in Question 15.
4. "Basic" was inserted before the last word in Question 17.
5. "Yearly" was added to indicate the usual number, instead of the specific year, in Question 23.
6. The typographical error "breads" was changed to "breaks" in Question 46.

Table A1.--Questions included in each questionnaire category.

	Questionnaire Categories	Question Numbers
1	Classroom enrichment	1-6
2	Homework	7-8
3	Teacher-made and other materials	9-12
4	Parent involvement	13-15
5	Time	16-19
6	Organizational pattern	20-21
7	Independent learning	32-38
8	Student interaction	27-28
9	Outside enrichment	29-31
10	Use of basal materials	39-44, 55
11	Staff support	45-46, 54
12	Language development	47-49
13	Written work	50
14	Tests	51-52
15	Rewards	53

Total number of questions = 55.

October 5, 1981

Dear Primary-School Teachers,

I need your help! I am conducting a study of primary-school teachers to ascertain what methods, materials, and strategies they use to teach reading. The data will be used in my dissertation research to determine what variables other than curriculum materials influence how and what students learn in reading.

Input is needed from school districts that use specific basal reading curriculums. I have spoken with ... and ... about the reading program used in ..., and they have permitted me to conduct a survey of the teachers in your district.

Won't you please take 15 minutes of your valuable time to complete the enclosed questionnaire? I hope the data gathered will provide additional insights into effective reading practices and so help all of us who strive to become better teachers.

Confidentiality in research is important. At no time will the respondents to questionnaires be identified. You will, however, notice a code number on your questionnaire; this number will be used only to facilitate follow-up procedures and to prevent you from receiving bothersome reminders.

Please complete the questionnaire and return it to your principal's office no later than October 12. It will be forwarded to me.

Thank you in advance for your time and cooperation. They are both greatly needed and appreciated.

Sincerely,

Phyllis H. Music
658 Bay East Drive
Traverse City, Michigan 49684

PRIMARY-TEACHERS' READING QUESTIONNAIRE

GENERAL DIRECTIONS:

For research purposes, it is important that you circle one response for each question on this questionnaire. Mark your choices clearly in pencil. Erase completely when necessary.

PART I

This section is regarding background information. Circle only 1 answer for each.

1. Grade teaching (1st, 2nd, 3rd, combination of these)

2. Years of experience (0-3, 4-7, 8-11, 12-more)

3. Years of teaching at the present grade level (0-3, 4-7, 8-11, 12-more)

4. Number of grades in the building (K-3, K-4, K-5, K-6, other _____)

5. Basal reading series used in your building through 1979-80 school year (Houghton-Mifflin, Scott Foresman, Holt, Ginn, other _____)

6. Copyright date of series indicated in number 5 (1969-70, 1971-2, 1973-4, 1975-6, 1977-8, other _____)

7. Where do your students fit on the continuum below? (Place an X to indicate the number of opportunities that your students have for educational and enrichment experiences outside the school environment.)

. _____ . _____ . _____ . _____ .
many none

8. I have had (0-1, 2-3, 4-5, 6-more) courses in how to teach reading to elementary students.

PART II

Primary Reading Instruction

Below are a variety of statements related to various strategies and methods used in teaching reading. Consider your own methods. Circle one choice for each of the following statements that would identify your own particular methods and/or classroom situation. The choices are:

1 = Always 2 = Most of the time 3 = Sometimes 4 = Seldom 5 = Never

1.	Instructional bulletin boards are used as a part of my teaching method.	1 2 3 4 5
2.	Reading centers are used to stimulate and motivate students to do reading activities.	1 2 3 4 5
3.	I read aloud to my students.	1 2 3 4 5
4.	I share my own personal experiences with children.	1 2 3 4 5
5.	A variety of hands-on activities are provided in my room.	1 2 3 4 5
6.	I use games and simulations to teach reading.	1 2 3 4 5
7.	Students are required to do reading homework for my class.	1 2 3 4 5
8.	Homework is used as a means for students to finish assignments that were begun in class.	1 2 3 4 5
9.	I use teacher-made materials to teach reading.	1 2 3 4 5
10.	I use a variety of other resources as well as the basal reading program (SRA, Reader's Digest, commercially prepared materials, etc.).	1 2 3 4 5
11.	I use enrichment activities that are not suggested in the teachers' guides.	1 2 3 4 5
12.	I work on reading skills that are not included in the instructional reading program.	1 2 3 4 5
13.	The students' parents are willing resources for my class.	1 2 3 4 5
14.	The parents show concern for their children's progress.	1 2 3 4 5

15.	The parents help their children to acquire basic reading skills.	1 2 3 4 5
16.	At least 50 minutes is allotted for each reading instruction period.	1 2 3 4 5
17.	More time is spent on teaching basic skills than on the application of newly acquired basic skills.	1 2 3 4 5
18.	I spend more time discussing ideas with students than instructing students.	1 2 3 4 5
19.	I spend more than 30 minutes daily in preparation for the teaching of reading.	1 2 3 4 5
20.	Structured programs work better than less-structured ones for me.	1 2 3 4 5
21.	Students are grouped according to ability for instructional purposes in my room.	1 2 3 4 5
22.	The seating arrangement is by rows.	1 2 3 4 5
23.	There are more than 27 students assigned to my classroom yearly.	1 2 3 4 5
24.	We use team teaching to teach reading.	1 2 3 4 5
25.	My students remain in the same reading group throughout the school year.	1 2 3 4 5
26.	I use an individualized approach to reading.	1 2 3 4 5
27.	Older students come into my room to tutor the younger ones.	1 2 3 4 5
28.	Classmates are encouraged to tutor each other.	1 2 3 4 5
29.	My students spend their free time outside of school watching television.	1 2 3 4 5
30.	There is evidence that my students read at home as well as at school.	1 2 3 4 5
31.	My students participate in extra-curricular activities such as scouts, after-school sports, music, etc.	1 2 3 4 5
32.	I emphasize and encourage self-initiated learning.	1 2 3 4 5

33.	My students read in the content areas during reading class time.	1 2 3 4 5
34.	My students use the library resources for a variety of activities.	1 2 3 4 5
35.	My students use the library as an informational resource.	1 2 3 4 5
36.	My students require detailed directions before they can begin assignments.	1 2 3 4 5
37.	The students are permitted to choose the books that they wish to read during silent reading periods.	1 2 3 4 5
38.	Students are permitted to choose the order in which they prefer to cover classroom assignments.	1 2 3 4 5
39.	I use the basal reading program that has been adopted by my school system.	1 2 3 4 5
40.	I rely on the teacher's manual for instructional ideas and information.	1 2 3 4 5
41.	Workbooks are used as a part of my instructional program.	1 2 3 4 5
42.	Workbooks are used with the reading groups that I teach.	1 2 3 4 5
43.	I use the enrichment activities that are suggested in the teachers' guides.	1 2 3 4 5
44.	My students must complete a textbook level before going on to the next higher level.	1 2 3 4 5
45.	I provide extra help for students who need it during breaks, after school, at recess time, etc.	1 2 3 4 5
46.	A reading specialist works with some of my students.	1 2 3 4 5
47.	My students are encouraged to share stories that they have read.	1 2 3 4 5
48.	My students are encouraged to share information that they have read.	1 2 3 4 5
49.	More time is spent discussing reading ideas and stories than doing written activities.	1 2 3 4 5

50.	Students are required to examine their written work to find and correct their mistakes.	1 2 3 4 5
51.	I use standardized test scores to establish reading groups.	1 2 3 4 5
52.	I use test information to alter my instructional program.	1 2 3 4 5
53.	My students are rewarded for doing good work (prizes, treats, etc.).	1 2 3 4 5
54.	Inservices are provided when my school system adopts a new basal reading program.	1 2 3 4 5
55.	Once a reading textbook is begun, I am expected to have students complete it.	1 2 3 4 5

APPENDIX B

MEAP CLASSROOM-LISTING REPORT

MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM 1980-81

CLASSROOM LISTING REPORT

TEACHER:

SCHOOL:

SECTION: 0

DISTRICT:

GRADE: 4

See reverse side for SKILL AREA and OBJECTIVE DESCRIPTION

STUDENT NAME	READING OBJECTIVE CODE																																		Total Objectives Attained
	I A	I B	I C	I D	I E	I F			II B	II C	II E	II F	II G			III A	III B	III C	III D	III E	III F	III G	III H	III I			IV A			V A	V B	V D	V F		
	N				N												N						N									N N	N		21 23 23
											N													N											24 24
					N	N					N					N			N			N		N			N N N			N	N	N N N	N		19 22 17
						N												N					N				N								21 25
					N	N					N					N	N			N	N	N	N				N								25 15 25
																											N								24 25
	N		N		N	N				H		H				N		N	N		N	N	N	N			N			N			N N		25 10 23
																																			25 25
																				N	N									N		N	N N		25 21 23
	N	N		N	N					H	N	N					N		N			N									N	N			12 25
																							N												24 25
PERCENT OF STUDENTS ATTAINING OBJECTIVE:	89	96	96	96	81	85			A	93	85	93	A			89	89	93	89	93	89	85	81	85			70			89	93	74	74		

BIBLIOGRAPHY

BIBLIOGRAPHY

Airasian, P., and Madaus, G. Measurement in Education, May 1972. Citing E. L. Thorndike. Seventeenth Yearbook of the National Society for the Study of Education, 1918.

An Assessment of Michigan Accountability System. Lansing: Michigan Department of Education, May 1974.

Bagford, Jack. Instructional Competence in Reading. Columbus, Ohio: Charles E. Merrill Publishing Co., 1975.

Beck, Michael, and Stetz, Frank P. "Standardized Testing as Viewed by Test Specialists and Users." Paper presented at the National Council on Measurement in Education, Washington, D.C., 1980.

Berdie, Douglas, and Anderson, John F. Questionnaires: Design and Use. Metuchen, N.J.: The Scarecrow Press, Inc., 1974.

Biemiller, A. "The Development of the Use of Graphic and Contextual Information as Children Learn to Read." Reading Research Quarterly 6 (Fall 1970): 75-96.

Blachowicz, Camille L. "Reading Objectives: A Competency-Based Accountability Model." The Reading Teacher 28 (April 1975): 645-61.

Bloom, B. S. Human Characteristics and School Learning. Quoted by Jeanne Chall. Reading 1967-1977: A Decade of Change and Promise. Bloomington, Indiana: The Phi Delta Kappa Educational Foundation, 1977.

Canney, George. "Organizing and Applying Test Results." In Reading Tests and Teachers: A Practical Guide. Edited by Robert Schreiner. Newark, Delaware: International Reading Association, Inc., 1979.

Chall, Jeanne. Reading 1967-1977: A Decade of Change and Promise. Bloomington, Indiana: The Phi Delta Kappa Educational Foundation, 1977.

Clay, Marie M. "Reading Errors and Self-Correction Behavior." British Journal of Educational Psychology 39 (1969): 47-56.

Clymer, Theodore, and Barrett, Thomas C. My Sound and Word Book. Lexington, Mass.: Ginn and Co., 1973.

Clymer, Theodore, and Jones, Virginia W. Seven Is Magic. Lexington, Mass.: Ginn and Co., 1973.

Clymer, Theodore, and Parr, Billie. A Duck Is a Duck. Lexington, Mass.: Ginn and Co., 1973.

_____. Helicopters and Gingerbread. Lexington, Mass.: Ginn and Co., 1973.

Clymer, Theodore, and Ruddell, Robert B. How It Is Nowadays. Lexington, Mass.: Ginn and Co., 1973.

Clymer, Theodore, and Wolfe, Elaine Vilscek. The Dog Next Door and Other Stories. Lexington, Mass.: Ginn and Co., 1973.

Clymer, Theodore, and Wolfing, Gretchen. With Skies and Wings. Lexington, Mass.: Ginn and Co., 1973.

Clymer, Theodore, and Wyatt, Nita M. May I Come In? Lexington, Mass.: Ginn and Co., 1973.

Cooper, Charles, and Petrosky, Anthony. "A Psycholinguistic View of the Fluent Reading Process." Journal of Reading 20 (1976): 184-207.

Crawley, Sharon. "SOB: Selection Basals." New England Reading Association Journal 12,2 (1977), p. 3. ED 155 588.

Criscuolo, Nicholas P. Improving Classroom Reading Instruction. Worthington, Ohio: Charles A. Jones Publishing Co., 1973.

Duffy, Gerald G. "Maintaining a Balance in Objective-Based Reading Instruction." Reading Teacher 31 (February 1978): 519-23.

_____, and Sherman, George B. How to Teach Reading Systematically. New York: Harper & Row, 1973.

Durkin, Dolores. Teaching Young Children to Read. Boston: Allyn & Bacon, Inc., 1980.

Durr, William; LePere, Jean M.; and Alsin, Mary Lou. Dinosaurs. Boston: Houghton-Mifflin Co., 1971, 1974.

_____. Lions. Boston: Houghton-Mifflin Co., 1971, 1974.

_____. Rainbows. Boston: Houghton-Mifflin Co., 1971, 1974.

_____. Secrets. Boston: Houghton-Mifflin Co., 1971, 1974.

_____. Signposts. Boston: Houghton-Mifflin Co., 1971, 1974.

_____. Tigers. Boston: Houghton-Mifflin Co., 1971, 1974.

Durr, William; LePere, Jean M.; and Brown, Ruth. Fiesta. Boston: Houghton-Mifflin Co., 1971, 1974.

_____. Panorama. Boston: Houghton-Mifflin Co., 1971, 1974.

Durr, William; LePere, Jean M.; and Niehaus, Bess. Rewards. Boston: Houghton-Mifflin Co., 1971, 1974.

Ebel, R. L. "Blind Guessing on Objective Achievement Tests." Journal of Educational Measurement 5 (1968): 321-25.

_____. "The Case for Minimum Competency Testing." Phi Delta Kappan (April 1978): 546.

_____. "One of the Troubles With Testing." College of Education Newsletter (Michigan State University) 3 (Spring 1981).

_____. In Readings in Measurement and Evaluation in Education and Psychology. Edited by William A. Mehrens. New York: Holt, Rinehart, and Winston, 1976.

_____. The Uses of Standardized Testing. Fastback 93. Bloomington, Indiana: The Phi Delta Kappa Educational Foundation, 1977.

Farr, Roger. Reading: What Can Be Measured? IRA/ENA Knight Research Fund Monograph. Newark, Delaware: Reading Association, Inc., 1969.

_____. "Standardized Reading Tests." In Reading for All. Edited by Robert Karlin. Newark, Delaware: International Reading Association, 1972.

Farr, Roger, and Brown, Virginia L. "Evaluation and Decision Making." In Elementary Reading. Edited by Althea Beery, Thomas C. Barrett, and William R. Powell. Boston: Allyn & Bacon, Inc., 1974.

Floden, Robert E.; Porter, Andrew C.; Schmidt, William H.; and Freeman, Donald J. Don't They All Measure the Same Thing? Consequences of Selecting Standardized Tests. East Lansing: Institute for Research on Teaching, Michigan State University, July 1978.

Freeman, Donald; Kuhs, Therese; Knappen, Lucy, and Porter, Andrew. A Closer Look at Standardized Tests. East Lansing: Institute for Research on Teaching, Michigan State University, November 1978.

Guthrie, John. "Research Testing: Uses and Visibility." Journal of Reading 23 (March 1980): 542-44.

_____. "Vocabulary Control." The Reading Teacher 33 (November 1979): 240-42.

Hall, Mary Anne; Ribovich, Jerilyn K.; and Rameg, Christopher J. Reading and the Elementary School Child. New York: D. Van Nostrand Co., 1979.

Harmer, William R. "The Selection and Use of Survey Reading Achievement Tests." In The Evaluation of Children's Reading Achievement. Edited by Thomas C. Barrett. Newark, Delaware: International Reading Association, Inc., 1967.

Harris, Albert J., and Sipay, Edward R. How to Increase Reading Ability. New York: David McKay Co., 1976.

Hood, Joyce. "Sight Words Are Not Going Out of Style." The Reading Teacher 33 (January 1977): 379-82.

House, Ernest R.; Rivers, Wendell, and Stufflebeam, Daniel L. "An Assessment of the Michigan Accountability System." Phi Delta Kappan 55 (June 1974): 663-69.

Huey, Edmund Burke. The Psychology and Pedagogy of Reading. Cambridge, Mass.: The M.I.T. Press, 1977.

Jenkins, Joseph R., and Pany, Darlene. "Curriculum Biases in Reading Achievement Tests." Journal of Reading Behavior 10 (Winter 1978): 345-57.

_____. Teaching Reading Comprehension in the Middle Grades. Reading Education Report No. 4. Urbana-Champaign: University of Illinois, January 1978. ED 151 756.

Jones, Timothy. Information Services Specialist, Detroit Census Bureau. Telephone interview, July 12, 1982.

Kappan Interview. "The Father of Behavioral Objectives Criticizes Them: An Interview With Ralph Tyler." Phi Delta Kappan 55 (September 1973): 55-56.

Kavali, Kenneth. "Selecting and Evaluating Reading Tests." In Reading Tests and Teachers: A Practical Guide. Edited by Robert Schreiner. Newark, Delaware: International Reading Association, Inc., 1979.

Kearney, C. Philip; Donovan, David L.; and Fisher, Thomas H. "In Defense of Michigan's Accountability Program." Phi Delta Kappan 56 (September 1974): 14-19.

Klein, Stephen P., and Kosecoff, Jacquiline. "Issues and Procedures." In Readings in Measurement and Evaluation in Education and Psychology. Edited by William A. Mehrens. New York: Holt, Rinehart and Winston, 1976.

Lovell, Ned B.; Riegle, Rodney P.; and Bunke, Clinton R. "Minimal Competency Testing: Hopes, Fears, and Fallacies." The Educational Forum 45 (January 1981): 199-206.

Lundsteen, Sara W. "Levels of Meaning in Reading." The Reading Teacher 28 (December 1974): 268-72.

Market Data Retrieval, Inc., HM Co. Reading K-8 Survey. Market Research Report. New York: Market Data Retrieval, Inc., 1977.

McClutcheon, Gail. "How Do Elementary School Teachers Plan? The Nature of Planning and Influences on It." The Elementary School Journal 81 (September 1980): 4-22.

McDonald, A. S. "Using Standardized Tests to Determine Reading Proficiency." Journal of Reading 8 (1964): 58-61.

Mehrens, William A., and Ebel, Robert L. "Some Comments on Criterion-Referenced and Norm-Referenced Achievement Tests." NCWE Measurement in Education, Vol. 10, No. 1. Washington, D.C.: National Council on Measurement in Education, Winter 1979.

Mehrens, William A., and Lehmann, Irvin J. Standardized Tests in Education. New York: Holt, Rinehart and Winston, 1980.

Michigan Educational Assessment Program. Technical Report, Vol. 1. Lansing: Michigan State Board of Education, 1980.

Michigan Educational Assessment Program Handbook. Lansing: Michigan Department of Education, 1980.

The National Conference on Achievement Tests and Basic Skills. Washington, D.C.: The National Institute of Education, U.S. Department of Health, Education, and Welfare, February 1979. ED 171 784.

Otto, Wayne, and Chester, Robert D. Objective-Based Reading. Reading, Mass.: Addison-Wesley Publishing Co., 1976.

Perrone, Vito. The Abuses of Standardized Testing. Fastback 92. Bloomington, Indiana: The Phi Delta Kappa Educational Foundation, 1977.

_____. "Achievement Testing: An Overview of the Issues." The National Conference on Achievement Tests and Basic Skills. Washington, D.C.: The National Institute of Education, U.S. Department of Health, Education, and Welfare, February 1979.

Pipho, Chris. State Activity Minimal Competency Testing. Denver, Colorado: Education Commission of the States, October 5, 1978.

Popham, James W. "Normative Data for Criterion Referenced Tests?" Phi Delta Kappan 57 (May 1976): 593-94.

Porter, John W. "Achievement Testing: The Interests." The National Conference on Achievement Tests and Basic Skills. Conference Proceedings. Washington, D.C.: National Institute of Education, U.S. Department of Health, Education, and Welfare, March 1978.

"A Position of Minimum Competencies in Reading." Journal of Reading 23 (October 1979): 50-51.

Prescott, George. "Criterion-Referenced Test Interpretation." In Elementary Reading Instruction: Selected Material. Edited by Althea Beery, Thomas C. Barrett, and William R. Powell. Boston: Allyn & Bacon, Inc., 1974.

Price, Gary. "Standardized Achievement Tests for Young Children: An Analysis." Paper presented at the Big Ten Symposium on Early Childhood Education, Madison, Wisconsin, October 1978.

Questions and Answers About the Michigan Educational Assessment Program. Lansing: Michigan Department of Education, September 1980.

Quinto, Frances, and McKenna, B. Alternatives to Standardized Testing. Washington, D.C.: National Education Association, 1977.

Resnick, Daniel P. "Testing in America: A Supportive Environment." Phi Delta Kappan 62 (May 1981): 625-28.

Resnick, Lauren B. "Achievement Testing: An Overview of the Issues." The National Conference on Achievement Tests and Basic Skills. Washington, D.C.: The National Institute of Education, U.S. Department of Health, Education, and Welfare, February 1979. ED 171 784.

_____. "Introduction: Research to Inform a Debate." Phi Delta Kappan 62 (May 1981): 623-24.

Roeber, Edward D.; Donnovan, David L.; and Cole, Richard F. "Telling the Statewide T Sting Story . . . And Living to Tell It Again!" Phi Delta Kappan 62 (December 1980): 273-74.

Rosecky, Marion. "Are Teachers Selective When Using Basal Guidebooks?" Reading Teacher 31 (January 1978): 381-84.

Rosenshine, B. "Skill Hierarchies in Reading Comprehension." In Theoretical Issues in Reading Comprehension. Edited by R. J. Spiro, B. C. Bruce, and W. F. Brewer. Hillsdale, N.J.: Erlbaum, 1978.

Spache, George D. The Teaching of Reading. Bloomington, Ind.: Phi Delta Kappan, Inc., 1972.

_____, and Spache, Evelyn B. Reading in Elementary Schools. Boston: Allyn & Bacon, Inc., 1973.

Tallmadge, Kasten G., and Horst, Donald P. "The Use of Different Achievement Tests in the ESEA Title I Evaluation System." Paper presented at the 62nd Annual Meeting of the American Educational Research Association, Toronto, Ontario, Canada, 1978.

Turnbull, William. Quoted by John Guthrie. "Research Testing: Uses and Visibility." Journal of Reading 23 (March 1980): 542-44.

Tyler, Ralph W., and White, Sheldon H. Testing, Teaching, and Learning. Washington, D.C.: The National Institute of Education, U.S. Department of Health, Education, and Welfare, August 1979.

Waldon, Shirley. Supervisor of State Aid, Michigan Department of Education. Telephone interview, July 12, 1982.

Weiss, Bernard J.; Everttts, Eldonna L.; and Stever, Loreli Olson. Never Give Up. New York: Holt, Rinehart, and Winston, 1977.

_____. People Need People. New York: Holt, Rinehart, and Winston, 1977.

_____. A Place for Me. New York: Holt, Rinehart, and Winston, 1977.

_____. Rhymes and Tales, Books and Games, Pets and People, Can You Imagine? New York: Holt, Rinehart, and Winston, 1977.

_____. Special Happenings. New York: Holt, Rinehart, and Winston, 1977.

_____. A Time for Friends. New York: Holt, Rinehart, and Winston, 1977.

_____. The Way of the World. New York: Holt, Rinehart, and Winston, 1977.

Woodbury, Marda. Selecting Instructional Materials. Fastback 110. Bloomington, Indiana: Phi Delta Kappa Educational Foundation, 1978.

Wrightstone, J. Wayne; Hogan, Thomas P.; and Abbott, Muriel M. "Accountability in Education and Associated Measurement Problems." In Readings in Measurement and Evaluation in Education and Psychology. Edited by William A. Mehrens. New York: Holt, Rinehart, and Winston, 1976.