INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand 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 you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. 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 any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

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

University Microfilms International

300 N. ZEEB ROAD, ANN ARBOR, MI 48106
18 BEDFORD ROW, LONDON WC1R 4EJ, ENGLAND

ERWIN, PAUL DEAN

THE MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM: A STUDY OF THE RELATIONSHIP BETWEEN MICHIGAN'S EXPERIMENTAL READING TEST AND SELECTED READING INSTRUCTIONAL PROGRAMS

Michigan State University PH.D. 1980

THE MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM: A STUDY OF THE RELATIONSHIP BETWEEN MICHIGAN'S EXPERIMENTAL READING TEST AND SELECTED READING INSTRUCTIONAL PROGRAMS

By

Paul Dean Erwin

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Administration
and Higher Education

1980

ABSTRACT

THE MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM:
A STUDY OF THE RELATIONSHIP BETWEEN MICHIGAN'S
EXPERIMENTAL READING TEST AND SELECTED READING
INSTRUCTIONAL PROGRAMS

By

Paul Dean Erwin

Purpose of the Study

This study was an attempt to establish the degree of concurrence between the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven and the K-6 reading instructional programs most commonly used in Michigan. The purpose of the study was four-fold: (1) to determine the concepts presented in the K-6 reading instructional programs, (2) to determine the concepts measured by the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven, (3) to analyze and compare the concepts presented in the K-6 reading instructional programs and the concepts measured by the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven as measured by the Reading Concepts Checklist, (RCC), and (4) to establish the degree of congruence between the K-6 reading instructional

programs as measured by the Reading Concepts Checklist, (RCC).

Procedure and Design

The Reading Concepts Checklist, (RCC), was developed as a means of describing, within a common framework, the concepts presented in the reading instructional materials and the concepts tested in the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven. This checklist was developed on the basis of conceptual consensus of agreement and based on the work of recognized authorities in the field of reading. The checklist formed the basis of two instruments: (1) A classification of K-6 instructional concepts matrix, and (2) A classification of tested concepts grades Four and Seven matrix.

The data for the reading instructional programs were collected by surveying the sixty-five teachers' manuals of the five reading instructional programs. As a concept was presented in the manual by the program, it was recorded in the matrix according to the appropriate grade level and concept.

The data from the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven were collected through a review and evaluation of the test

by a panel of reading experts. The panel determined the reading process being measured by each test item and recorded the test item in the tested concepts matrix according to the appropriate concept and test level.

The analysis leading to the comparison of the concepts presented in the reading instructional programs to the concepts tested in the Michigan Educational Assessment Program Experimental Reading Test Grade Four and Seven requires data from all levels of the K-6 reading instructional programs. The criteria which guided the selection of the reading instructional programs were as follows: (1) the reading instructional programs must be used by a majority of Michigan's K-6 students, (2) the term majority was defined as a clearly definitive number, not simply "more than half," and (3) the majority must be large enough that it represented a reasonable cross-section of Michigan's rural, urban, and large-city K-6 students. Therefore, the lower acceptable limit which defined a majority of students using the reading instructional programs to be included in the study was established as seventy-five percent. The final selection of the reading instructional programs was based on a national survey of K-8 reading specialists and reading supervisors. The reading instructional programs selected to be compared with the Michigan Educational Assessment Program Experimental Reading Test Grade Four and Seven, and which were used by at least seventy-five

percent of Michigan's K-6 students, were (1) Ginn and Company, (2) Harcourt, Brace and Jovanovich, (3) Holt, Rinehart, and Winston, (4) Houghton-Mifflin Company, and (5) Scott, Foresman Company. Usuable data were acquired from sixty-five teachers' manuals and the independent ratings of the researcher and three reading experts of the two levels of the Michigan Educational Assessment Program Experimental Reading Test Grade Four and Seven.

The two major hypotheses, developed and tested, were stated as follows:

I. There will be no difference between the five reading instructional programs in grades K-3 in the concepts they present or between the degree of concurrence between the concepts presented in each of the five reading instructional programs in grades K-3 and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 as shown in the Reading Concepts Checklist, (RCC).

II. There will be no difference between the five reading instructional programs in grades 4-6 in the concepts they present or between the degree of concurrence between the concepts presented in each of the five reading instructional programs in grades 4-6 and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 as shown in the Reading Concepts Checklist, (RCC).

A non-parametric, distribution free test, Cochran's "Q" Test, compared to a Chi-square distribution, was used to test the significance between the observed differences between the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC). The determination of the magnitude and direction of the significance of the

difference between the proportion scores was conducted by multiple comparisons of the means of the proportion scores through the use of the Dunn-Bonferroni pairwise comparison technique. The Cochran Q Test was employed to determine the level of reliability and the degree of inter-rater agreement of the panel of reading experts.

Major Findings and Conclusions

The following appraisal of the findings was reached:

1. The findings of the study indicate a lack of concurrence between the Michigan Educational Assessment Program Experimental Reading Test Grade Four and Seven, and the K-6 reading instructional programs. Total proportion scores of the matches and mismatches across the Reading Concepts Checklist, (RCC), and proprotion scores from the nine subcategories in the Reading Concepts Checklist, (RCC), indicate a lack of concurrence between the Michigan Educational Assessment Program Experimental Reading Test and each of the five reading instructional programs.

2. The findings of the study indicate the degree of concurrence present among the reading instructional programs is significantly greater between (1) Ginn and Company, (2) Harcourt, Brace and Jovanovich, and (3) Holt, Rinehart, and Winston, and is significantly greater between (4) Houghton-Mifflin Company and (5) Scott, Foresman Company; Thus forming two distinct groups.

3. The findings of the study indicate significant differences exist in the K-3 reading instructional programs in categories V, Comprehension: Vocabulary Development, VII, Inferential Comprehension, and IX, Study Skills of the Reading Concepts Checklist, (RCC), while significant differences exist in the 4-6 reading instructional programs in categories III, Phonic Analysis, IV, Structural Analysis, VI, Literal Comprehension, and VII, Inferential Comprehension of the Reading Concepts Checklist, (RCC).

In general, the findings of a significant lack of concurrence between the K-6 reading instructional programs and the Michigan Educational Assessment Program Experimental Reading Test Grade Four and Seven, should be of importance to everyone concerned with the assessment of reading concepts and skills in Michigan's K-6 grades.

DEDICATION

One of America's Best loved poets wrote:

> Two roads diverged in a yellow wood,
> And sorry I could not travel both
> And be one traveler, long I stood
> And looked down one as far as I could
> To where it bent in the undergrowth;
>
> Then took the other, as just as fair,
> And having perhaps the better claim,
> Because it was grassy and wanted wear;
> Though as for that the passing there
> Had worn them really about the same,
>
> . . . . . .
>
> Two roads diverged in a wood, and I-
> I took the one less traveled by,
> And that made all the difference.*

This work is dedicated to my wife and son,

Mary Jane and Brian

for their love, encouragement, and support. They have sacrificed so much that I might have an opportunity to travel the less traveled path.

*"The Road Not Taken" by Robert Frost

ACKNOWLEDGEMENTS

Every doctoral program is unique in its own way. Yet, every doctoral program shares a commonality; it could not have been accomplished without the assistance of many concerned and enthusiastic people. It is hoped that all who shared in the completion of this program realize their support and assistance is deeply appreciated. There are some, however, whose contribution needs special recognition.

Special recognition and appreciation is extended to Dr. Herbert C. Rudman, chairman of my guidance committee, for the many hours spent in guiding this candidate through the program and especially the work contained here. He was to become not only an advisor but a friend as well.

To Drs. William Durr, Keith Groty, and Frederick R. Ignatovich for their service on the Guidance Committee. Each has been willing to provide necessary guidance and help when and where needed.

To Drs. Gerald Duffy, William Durr, and George Sherman for their willing assistance and the hours they gave in evaluating and reviewing the Michigan Experimental Reading Test.

To Dr. Edward D. Roeber of the Michigan Department of Education for his cooperation and assistance with assessment information and materials.

To Necia Black who served as a source of encouragement and information through the statistical aspects of this endeavor.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF APPENDICES

Appendix Page

CHAPTER I

INTRODUCTION

Statement of the Problem

Nearly every element of the mass media has published or broadcast a news item discussing the downward trend of the achievement levels in America's schools. The increased publicity about the quality of American educational programs has caused taxpayers to question what they are receiving for the money they are spending. The public believes the schools are certifying incompetent students as competent by passing them along, graduating them, granting them diplomas.[1]

Increased concern about the quality of American education has led to a renewed interest in competency based education. The concept of competency based education suggests the existence of standards or a desired level of preformance. A result of citizen interest in competency based education has been to place the pressure of accountability upon all levels of the educational system.

---

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

The pressure of accountability is evidenced in the document, State Activity: Minimal Competency Testing, prepared by Pipho in October, 1978.[2] Thirty-six states were involved in some phase of an accountability program. Michigan is one of those thirty-six states. It has a comprehensive six step accountability model: 1) identify goals, 2) develop performance objectives, 3) assess needs, 4) analysis of delivery systems, 5) testing and evaluation, 6) final recommendation for change or recycle to step one.[3] A portion of Michigan's Model is known as the Michigan Educational Assessment Program. The Michigan Educational Assessment Program (MEAP) was initiated by the State Board of Education, supported by the Governor, and first funded by Act 307 of the Public Acts of 1969 and then under Act 38 of the Public Acts of 1970.[4] Initially, it took the form of a norm-referenced test. It was changed to an objective-referenced test in 1973-1974 because 1) the accountability model specifically called for objective-

---

[2]Chris Pipho, State Activity Minimal Competency Testing, Education Commission of the States, Denver, Colorado, October 5, 1978, p. 1-12.

[3]Michigan Department of Education, Michigan Accountability 1976-77 (Lansing, Michigan: Undated) p.3.

[4]Michigan Educational Assessment Program, First Report of the 1977-78 Michigan Educational Assessment Program, Interpretive Manual (Lansing, Michigan: Michigan Department of Education, 1978), Foreward.

referenced assessment, 2) the development of performance objectives and tests tied directly to them is a useful process for educators for the classification of instructional intentions, and 3) the objective-referenced test data are much more specific and more useful in assisting teachers to respond to individual student needs.[5]

The statement, the development of performance objectives and tests tied directly to them as a useful process for educators for the classifications of instructional intentions, is accurate only to the extent of the relationship between the test and the field of study. The extent of that relationship has been debatable. The debate centers on issues such as whether there is a consensus of opinion among educators that the objectives constitute the worthwhile objectives local districts should be striving to attain and who was involved in writing the objectives. The claim that hundreds of Michigan teachers, curriculum specialists, and administrators were involved in the writing of the objectives[6] was countered by the claim that only a few persons were involved and that the objectives chosen do not represent consensusal choices of even the small group

---

[5]Philip Kearney, David L. Donovan, and Thomas H. Fisher, "In Defense of Michigan's Accountability Program," Phi Delta Kappan 56 (September, 1974), p. 16.

[6]William Mehrens. Technical Report: The Fifth Report of the 1973-74 Michigan Educational Assessment Program. (New York: ERIC Document Reproduction Services, Ed 120218, July, 1976), p. 18.

who were involved in developing the objectives.[7] Both issues are important because they underscore the problem this research seeks to address.

The rationale that objective-referenced test data are more specific and more useful in assisting teachers to respond to individual student needs has legitimate bases. Objectives are specific. Objectives provide direction for the teacher. They assist the teacher in planning instruction, guiding student learning, and provide the criteria to evaluate student outcomes.[8] The debate concerning the degree of concurrance between the test content and the instructional materials tends to raise questions concerning the usefulness of the Michigan Educational Assessment Program in assisting teachers to respond to individual student needs.

The general problem this research project seeks to address is the insufficiency of available data concerning the relationship of the Michigan Educational Assessment Program's content to the instructional programs used throughout the State of Michigan.

---

[7]Ernest R. House, Wendell Rivers, and Daniel L. Stufflebeam, "A Counter-Point to Kearney, Donovan, and Fisher," Phi Delta Kappan 56 (September, 1974), p. 19.

[8]William A. Mehrens and Irvin J. Lehmann, Measurement and Evaluation in Education and Psychology 2nd ed., New York: Holt, Rinehart and Winston, 1978, p. 19.

Within this general framework, of specific concern is the need to identify the relationship between the concepts being measured by the Michigan Educational Assessment Program Experimental Reading Test for grades four and seven and the concepts presented through local instructional programs.

Statement of the Purpose

Little appears to have been done in investigating the relationship between the Michigan Educational Assessment Program's Experimental Reading test for grades four and seven and the concepts presented through local reading instructional programs. One result of the contested relationship between the test content and the concepts presented in the instructional materials has been a continuance of the questioning of the content validity of the assessment test.

The Michigan Department of Education appears to be moving toward resolving the question. In September, 1979, the Michigan Department of Education conducted its annual assessment program. Concurrently, the Department pilot tested an experimental assessment program. However, the experimental assessment program has been prepared and pilot tested along the same procedural lines as the current assessment program. Therefore, the potential for the debate over the content validity of the experimental assessment test remains.

The purpose, then, of this research project will be to establish the degree of concurrence between the concepts measured in the Michigan Educational Assessment Program Experimental Reading Test and the concepts presented in the selected instructional programs used in Michigan.

Specifically, the researcher will undertake to determine:

1. What knowledge, skills, abilities, or behaviors (tasks) are presented in the selected instructional programs used in Michigan?
2. What knowledge, skills, abilities, or behaviors (tasks) are presented in the experimental reading objectives and items in the Michigan Educational Assessment Program Experimental Reading Program, Grades Four and Seven?
3. What is the degree of overlap between the selected instructional reading programs and the Michigan Educational Assessment Program Experimental Reading Test?

Significance of the Study

Measurement and evaluation play a vital role in education. The predominant mode of evaluation is through written tests. More recently those tests tend to be objective-referenced tests, that is, a test based upon a set of objectives assumed to be representative of the content domain from which they have been taken. From the perspective

of program evaluation, use of such test results for either diagnostic and prescriptive or summative purposes depends upon the degree to which the test is a representative sample of the content domain.

The significance of this research project lies in its attempt to identify and appraise the relationship between the selected reading instructional materials and the Michigan Educational Assessment Program Experimental Reading Test.

The identification and appraisal of the relationship between the instructional materials and the Experimental Test is significant to several groups: 1) the Michigan Department of Education, 2) the local school districts that are using the test, and 3) the publishers of the instructional materials in use throughout the State of Michigan.

The Michigan Department of Education is attempting to create a new assessment test. The intended outcome is that the new test will more nearly reflect the instructional materials used in Michigan. The results of this study will provide the Michigan Department of Education with data showing the degree of concurrence between the Michigan Educational Assessment Program Experimental Reading Test and the instructional materials used in this study. Therefore, the actual identification and appraisal of the degree of concurrence between the Experimental Reading Test and the instructional reading materials will have a direct impact

on the policy and practice of the Michigan Department of Education in its attempt to revise and implement the Experimental Reading Test for grades four and seven.

The debate centering around the content validity issue has caused some problems for teachers and administrators at the local school district level. The public already believes the schools are granting diplomas to incompetent students.[9] In some instances, publication of test results seems to indicate the public is correct. In their own defense, school officials attempt to explain their test results on the basis of the test's lack of content validity even though neither side of the debate has been substantiated. Local school district officials and teachers need empirical data that illustrate the relationship of the instructional programs used in their district and the Experimental Reading Test being developed by the State of Michigan. The significance of this research project, then, for administrators and teachers of the local school district is that it will provide them the data concerning the relationship of the selected instructional reading materials to the Experimental Reading Test Grades Four and Seven and the relationship which exists between the instructional programs themselves.

As the Experimental Reading Test is an objective-referenced test, many local districts are developing

---

[9]Ebel, "The Case for Minimum Copetency Testing" p. 546.

objectives upon which to base their instructional programs. Textbook selection is becoming more sophisticated and the final selection is more frequently based on the degree to which the district objectives and the textbook objectives match. Knowledge of the relationship of a given instructional program to the other instructional programs, or its relationship to the Experimental Reading Test is significant to the publishers of the instructional programs used in this project.

Theory and Supportive Research

The insufficiency of available data concerning the relationship of the Michigan Educational Assessment Program's content to instructional programs used throughout the State of Michigan has been identified as the general problem this study will address. One aspect of the lack of available data is a lack of evidence to support the relationship between the concepts measured by the Michigan Educational Assessment Program Experimental Reading Test and local district instructional programs. The degree to which the relationship exists is determined by how well the test items measure the objectives and sample the content domain.[10] Whether the test is norm-referenced or criterion-referenced,

---

[10]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.

the test items should be keyed to a set of objectives and should be representative of a specified content domain. If that is the case, the test is likely to have content validity.[11]

Magnusson discusses content validity as the extent to which a test covers a field of study.[12] In this instance, the test items serve as a sample taken from a domain representing the content or aims of the course. Content validity is established by the extent to which the sample is representative of the total domain. Before one can estimate content validity, one must explicitly define the aims of instruction given in the field and the material which the students should have grasped.[13]

In his chapter "The Validity of Classroom Tests," Ebel discusses two categories of validity: 1) primary or direct validity, and 2) derived or secondary validity.[14]

---

[11] Ibid., p. 3.

[12] David Magnusson, Test Theory, Trans. by Hunter Mabon, Reading, Mass.: Addison-Wesley Publishing Company, 1967, p. 129.

[13] Ibid.

[14] Robert L. Ebel, Essentials of Educational Measurement. 2nd ed., Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972, p. 438.

Direct validity is defined as the extent to which the tasks included in the test faithfully represent and in the proper proportion, are the kinds of tasks that provide an operational definition of the trait or achievement in question, whereas derived validity is the extent to which the scores it yields correlate with criterion scores that possess direct validity.[15] Lists of various types and definitions of validity have been suggested by numerous authors in the field of educational measurement and psychology. Of particular interest are content validity, defined as being concerned with the adequacy of sampling a specified universe of content,[16] and curricular validity, defined as being determined by an examination of the content of the test itself and judging to what degree it is a true measure of the important objectives of the course, or is a truly representative sampling of the essential materials of instruction.[17] The importance of the correlation between

---

[15] Ibid.

[16] American Psychological Association, Inc., Technical Recommendations for Psychological Tests and Diagnostic Techniques, Washington, D. C: APA 1954 in Ebel, Robert L. Essentials of Educational Measurement. 2nd ed., Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972, p. 437.

[17] C. C. Ross and Julian C. Stanley, Measurement in Today's Schools, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1954 in Ebel, Robert L., Essentials of Educational Measurement, 2nd ed., Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972, p. 437.

these two definitions is made apparent from the point of view that a test author may succeed to some degree in attaining his goal if he defines his domain and writes items to represent the domain. However, from the point of view of the one who uses the test, content validity is situation-specific. Teachers teaching the same course titles are not necessarily teaching the same content domain. The result is that the test would have high content validity for one teacher and low content validity for another.[18]

Content validity and curricular validity are determined by the test content, the extent to which the test content is a representative sample of the essential materials of instruction, and is a representative sample in proportion to the total population.

---

[18]William A. Mehrens and Irvin J. Lehmann, Measurement and Evaluation in Education and Psychology, 2nd ed., New York: Holt, Rinehart and Winston, 1978, pp. 111-112.

Spool,[19] Magnusson,[20] Lennon,[21] and Tanenbaum, et al.,[22] have all suggested various components or measures of the appraisal of content validity. The basic elements of those components are: 1) the behavior to be exhibited in the performance domain, 2) the behavior to be demonstrated in testing, and 3) the relationship between the two. The relationship between behaviors in the performance domain and behaviors required by the test determines the test's validity. The goals of the test must match the goals of the instructional program. This does not constitute teaching to the test, but rather it is the selection of a test capable of measuring growth in the specific objectives of the instructional program.[23] This point is illustrated

---

[19]Mark D. Spool, Performing a Content Validity Study, Paper Presented at the Annual Meeting of the Souteastern Psychological Association (21st, Atlanta, Ga.) 1975, p. 3.

[20]Magnusson, Test Theory, P. 129.

[21]Robert T. Lennon, "Assumptions Underlying the Use of Content Validity," Readings in Measurement and Evaluation in Education and Psychology, Edited by William A. Mehrens, New York: Holt, Rinehart and Winston, 1976, p. 47.

[22]Arlene B. Tanenbaum and Christine A. Miller, The Use of Congruence Between the Items in a Norm-Referenced Test and the Content in Compensatory Educational Curricula in the Evaluation of Achievement Gains, Paper Presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York), 1977, pp. 1-10.

[23]Roger Farr, Reading: What Can Be Measured?, (Newark, Delaware: International Reading Association, 1969), p. 36.

by the study of Jenkins and Pany. This study concerned itself with the extent and direction of curriculum bias in five widely used standardized achievement tests by comparing the relative overlap in the content of these reading achievement tests with the first and second grade contents of seven commercial reading series.[24] They found that examination of scores for the curricula they studied revealed that expected annual growth would vary according to which test was used. They concluded, therefore, that it is doubtful the use of conventional achievement tests can provide an unbiased estimate of a curriculum's effect, at least with regard to the early grades.[25]

The work of Jenkins and Pany underscores the need for a high correlation of relationship between the behaviors in the performance domain and those to be demonstrated in testing. While they do raise some questions regarding the manipulation of the tests against the curriculum used to cause the scores to reflect the users bias, these questions deal with the issue of misuse of test results. Constructors of achievement tests have always emphasized the importance of defining the content domain and sampling from it in an appropriate fashion. Therefore, whether they are norm-referenced or criterion-referenced, good achievement test

---

[24]Joseph R. Jenkins and Darlene Pany, "Curriculum Biases in Reading Achievement Test," Journal of Reading Behavior, Vol. X, No. 4 (Winter, 1978), p. 348.

[25]Ibid., p. 353.

items should be based on a set of objectives and represent a specified content domain.[26] The extent to which that relationship exists will determine how much content validity a test has for a particular purpose.[27]

Limitations and Assumptions

Any comparative research faces a limitation in the extent to which terms used have a shared definition across individuals and subject groups. It is an assumption of this research that the terms used will have a high degree of meaning and similarity of meaning across reading specialists and test constructors.

This research is also limited by the fact that the source of information used to select the comparions instructional materials only provides information for the national and regional levels. The assumption is that the regional information provides a reasonable approximation of the most commonly used materials in Michigan.

Another limitation in this study is the fact that the publishers have more than one edition in use at the same time. The assumption is that skills presented tend to remain constant from one edition to the next and that the latest edition may be used for analysis.

---

[26]Mehrens and Ebel, Some Comments on Criterion-Referenced and Norm-Referenced Achievement Tests, p. 3.

[27]Ibid., p. 5.

The research is limited in that no attempt shall be made to address the issue of instructional validity, that is the degree of emphasis placed on concepts taught within and between classrooms. The assumption is that teachers tend to follow the instructional reading programs which they use.

Hypotheses

General Hypothesis I

There will be no difference between the five reading instructional programs in grades K-3 in the concepts they present or between the degree of concurrence between the concepts presented in each of the five reading instructional programs in grades K-3 and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 as shown in the Reading Concepts Checklist, (RCC).[28]

Operational H1a

There will be no difference between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the K-3 reading instructional program published by Harcourt, Brace and Jovanovich according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC)

Operational H1b

There will be no difference between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the K-3 reading instructional program published by Holt, Rinehart and Winston according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

---

[28] Appendix A

Operational H1c

There will be no difference between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the K-3 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1d

There will be no difference between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the K-3 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1e

There will be no difference between the concepts presented in the K-3 reading instructional program published by Harcourt, Brace and Jovanovich and the K-3 reading instructional program published by Holt, Rinehart, and Winston according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1f

There will be no difference between the concepts presented in the K-3 reading instructional program published by Harcourt, Brace and Jovanovich and the K-3 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1g

There will be no difference between the concepts presented in the K-3 reading instructional program published by Harcourt, Brace and Jovanovich and the K-3 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1h

There will be no difference between the concepts presented in the K-3 reading instructional program published by Holt, Rinehart, and Winston and the K-3 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1i

There will be no difference between the concepts presented in the K-3 reading instructional program published by Holt, Rinehart, and Winston and the K-3 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1j

There will be no difference between the concepts presented in the K-3 reading instructional program published by Houghton-Mifflin Company and the K-3 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1k

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1l

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Harcourt, Brace and Jovanovich and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1m

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Holt, Rinehart and Winston and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Opeational H1n

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Houghton-Mifflin Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1o

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Scott, Foresman Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concept Checklist, (RCC).

General Hypothesis II

There will be no difference between the five reading instructional programs in grades 4-6 in the concepts they present or between the degree of concurrence between the concepts presented in each of the five reading instructional programs in grades 4-6 and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 as shown in the Reading Concepts Checklist, (RCC).

Operational H2a

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the 4-6 reading instructional program published by Hartcourt, Brace and Jovanovich according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2b

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the 4-6 reading instructional program published by Holt, Rinehart, and Winston according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2c

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the 4-6 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2d

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the 4-6 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2e

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich and the 4-6 reading instructional program published by Holt, Rinehart, and Winston according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2f

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich and the 4-6 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2g

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich and the 4-6 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2h

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Holt, Rinehart, and Winston and the 4-6 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2i

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Holt, Rinehart, and Winston and the 4-6 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2j

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Houghton-Mifflin Company and the 4-6 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2k

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 according to the proportion of matches and mismatches across the Reading Concetps Checklist, (RCC).

Operational H2l

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich and the concepts tested by the Michigan Educational Program Experimental Reading Test Grade 7 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2m

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Holt, Rinehart, and Winston and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2n

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Houghton-Mifflin Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2o

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Scott, Foresman Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Organization of the Thesis

This chapter contains a statement of the problem, a statement of the purpose, the significance of the study, and the theory and research upon which the study is based. It also includes the limitations and assumptions of the study and the testable hypotheses.

In Chapter II, a review of related literature is presented. The review includes objective-referenced test construction, related studies of the relationship of test content to instructional materials, and the theory of content validity.

In Chapter III, the procedure and methodology of the study are presented. The detailed description includes selection of instructional materials, data collection, the instrumentation, and the statistical analysis treatment.

The results of the analysis of the data are presented in Chapter IV.

In Chapter V, the summary, discussion of the major findings, recommendations, and areas for further research are presented.

CHAPTER II

RELATED LITERATURE

Introduction

Inflation, technological advancements, the complexities of attempting to meet the needs of students have placed a strain on the imagination of educators across America. Educators' efforts seem to be achieving less and parents' complaints seem stronger as evidence appears to mount in support of the notion that the cost of education continues to rise while its achievements seemingly decline annually. It has become the opinion of the citizens of the community that it is necessary and proper to hold the school board members, the school administrators, the teachers, and the students accountable for their successes or failures in the learning process.[1]

One result of the demand for accountability has been renewed interest in the Competency Based Education (CBE) movement. The move toward CBE has renewed interest in the

---

[1]Robert L. Ebel, Essentials of Educational Measurement 3rd ed., Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1979, p. 3.

field of measurement and evaluation, specifically in the form of Criterion Referenced Tests and mandated assessment programs by Departments of Education at the state level. In 1978, thirty-six states were involved in some phase of an accountability program and the evidence indicates this number will increase rather than decline.[2]

Questions concerning the adequacy of these mandated tests arise from teachers and administrators. A major question concerns the correspondence between test content and instructional content. A review of the literature concerning the theory behind test validity would be inadequate if it did not include a discussion of the purpose of evaluation and the procedures involved in test construction.

Purpose of Evaluation

Tests are used in a variety of situations. They may be administered prior to instruction as a survey of prior knowledge. They may be administered during the course of instruction to monitor student understanding of the material being presented. Tests may be administered at the conclusion of the course of instruction to assess the level of student achievement. These processes can be

---

[2]Chris Pipho, State Activity Minimal Competency Testing, Educaton Commission of the States, Denver, Colorado, October 5, 1978, pp. 1-12.

adopted separately or in any or all combinations. The purpose of the evaluation process in each case is to provide a description or a representation of a person.[3]

The function of the evaluation is to aid in the decision-making process. If a test does not aid in the decision-making process, the test is useless.[4] The term "decision" is defined as all possible courses of action which might follow from test scores.[5] Linking the two terms, function and decision-making process, adds yet another dimension to the overall view of educational evaluation. The function of the evaluation can assume different meanings depending upon the perspective from which the evaluation is viewed. As Cronbach[6] defines the functions of evaluation, there are five: 1) learner feedback, 2) learner reinforcement, 3) teacher feedback, 4) counseling decision, and 5) administrative decisions.

---

[3]William A. Mehrens and Irvin J. Lehmann, Measurement and Evaluation in Education and Psychology 2nd ed., New York: Holt, Rinehart and Winston, 1978, p. 110.

[4]Jum C. Nunnally, Educational Measurement and Evaluation 2nd ed., New York: McGraw Hill Book Company, 1972, p. 4.

[5]Ibid., p. 5.

[6]Lee J. Cronbach, Educational Psychology 2nd ed., New York: Harcourt, Brace and World, Inc., 1963, p. 539.

The idea behind learner feedback is to assist the student to realize how he should change or develop his behavior, while learner reinforcement provides the student with confirmation of his own assessment regarding his level of achievement. Students who receive high scores on tests are encouraged to continue the work habits and methods of study which brought them success. Students who score poorly on tests are warned to work harder, change their methods of study, and seek help. As the students mature, the cummulative effect of the years of testing will enable the students to learn where their strengths and weaknesses are and allow them to plan for their future. Tests also provide feedback to student about the key concepts in instruction.[7] Patterns begin to develop. Concepts emphasized on previous examinations become the emphasis for students to study for future examinations.

The information provided to teachers through the use of tests helps them judge the adequacy of teaching methods. Student performance on the tests indicates to teachers what needs to be retaught and which methods are effective and can be used again to teach to the same objective.[8] Only to the extent that the average student meets the objectives

---

[7]Nunnally, Educational Measurement and Evaluation, p. 126.

[8]Cronbach, Educational Psychology, p. 540.

can the teacher feel satisfaction with the instruction as a whole, and the progress of individual students is judged largely by how well they perform with respect to the objectives. Lacking the knowledge of that progress, intelligent decisions about the individual or the class as a whole cannot be made.[9]

Opportunities for promotion within school or to advance studies in colleges and universities, recommendations to pursue a particular type of employment or types of employment not to consider are the types of decisions frequently made by counselors and administrators on the basis of test results. Some of those decisions are reached with the students and some are reached for the students. Administrative decisions concerning the total school program are based on the use of test results. Analysis of test scores provides indications of the program's strengths and weaknesses. Inferior areas will need to be brought up to standard through a change in instructional materials, different instructional strategies, or both.[10]

The purposes and functions of educational evaluation instruments are predicated upon the assumption that the tests have been constructed according to the requirements for constructing tests. It is widely recognized that

---

[9]Nunnally, Educational Measurement and Evaluation, p. 124.

[10]Cronbach, Educational Psychology, p. 540.

teachers and administrators are encouraged to use standardized test results to assess achievement, identify learning problems, and evaluate the effectiveness of instructional strategies. The use of test results to achieve any of these functions can be considered only in view of the teachers' knowledge of the extent to which the content of the test corresponds to the content of instruction.[11] The same caution holds true to a somewhat lesser degree of teacher-made tests. If the procedures of test construction are followed for teacher-made, or tailor-made achievement tests, the underlying assumption is that the teacher-made test will be more directly linked to the course objectives than the more global standardized test in that it is an assumption that the goals and objectives of tailor-made tests are tied more closely to the smaller units of instruction. The tailor-made tests are constructed for a specific purpose and are a sample of a more constricted domain.[12] The caution as stated here will be more

---

[11]Donald Freeman, Therese Kuhs, Lucy Knappen, and Andrew Porter, A Closer Look at Standardized Tests, Institute for Research on Teaching, East Lansing, Michigan, November 1978, p. 1.

[12]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), p. 4.

appropriately expanded and treated fully in the section later in this chapter concerning test validity.

Objective Referenced Tests

Distinctions Between Test Types

Tests, generally, can be classified into two major categories: 1) essay tests and 2) objective tests. Essay tests are answered in the narrative form by the examinee. The essay test requires less time to prepare, but a greater amount of time to grade. The grading of essay tests is subjective in nature and dependent upon the judgment of the rater as to whether the question has or has not been answered and the degree to which the question has or has not been answered.

Objective tests contain the distinctive characteristics of providing a greater number of items which allows for a more extensive sampling of the content domain; of not usually requiring the student to produce an answer all on his own, but rather only requiring that he recognize the correct answer by one method or another; of having rules for scoring that are absolutely clear.[13]

Objective tests are those tests which are usually classified as standardized tests. They are standardized in the sense that they conform to specific criteria.

---

[13]Nunnally, Educational Measurement and Evaluation, p. 155.

Within those criteria, there are several tests which can be classified as "standardized tests." The problem seems to be one of definition when reference is made to the various types of objective tests. The generally accepted classification of objective tests is 1) standardized achievement tests, 2) tailor-made achievement tests, 3) objective-referenced tests, and 4) domain-referenced tests.[14] In some instances, the distinctions between these tests are major and in other instances the distinctions are much more subtle. Within the standardized achievement test category, classifications are subdivided into criterion-referenced tests, norm-referenced tests, objective-referenced tests, and domain-referenced tests. While all *good* achievement tests are objective based, the major distinction is the manner in which the user wishes to use the data gathered. It is a misconception that objective-referenced tests, criterion-referenced tests, and domain-referenced tests are not "standardized tests." Rather, it is the interpretation of their use which differentiates them from the other standardized test, a "norm-referenced" test. All are commerically prepared and draw their sample from a broad domain of general interest. All can be used for normative referencing or criterion

---

[14]Mehrens and Ebel, Some Comments on Criterion-Referenced and Norm-Referenced Achievement Tests, p. 4.

referencing.[15] The difference between the normative reference interpretation and the criterion reference interpretation is that the meaning of an individual's score gains its meaning through comparison to some specific criterion of proficiency. If the comparison is to scores of other individuals in a particular group, it is normative referencing. If the comparison is to specific criterion of proficiency, it is criterion referencing. Further confusion lies in the fact that the terms criterion-referenced and objective-referenced are used interchangeably. An objective-referenced test, simply stated, is a test in which the tasks have been related directly to a set of objectives.[16]

Another major distinction between norm-referenced tests and criterion-referenced tests is that the norm-referenced test is descriptive and predictive in nature and the criterion-referenced test is generally diagnostic and prescriptive.[17] The criterion-referenced test reflects the examinee's standing relative to the curriculum. The discrimination is between the level of mastery or non-

---

[15] Ibid.

[16] Ibid.

[17] Glen E. Roudabush, Item Selection for Criterion-Referenced Tests, Paper Presented at the Annual Meeting of The American Educational Research Association, (57th, New Orleans, La.) 1973, p. 2.

mastery of the objectives making up the curriculum of interest from which the criterion-referenced test was constructed. From the information gathered as to which objectives have or have not been mastered (diagnostic information), decisions for further instruction can be made (prescriptive information). Following additional instruction based on the decisions made from the previous criterion-referenced examination, another criterion-referenced test can be administered which would reflect changes in the examinee's capability to perform. The implications of this major difference are that the items for a criterion-referenced test should be sensitive to instruction, while the items of a norm-referenced test should be sensitive to individuals.[18]

The purpose of the criterion-referenced test involves the classification of individuals into one of several mutually exclusive categories.[19] The mutually exclusive categories may be masters and non-masters, instructed and uninstructed students, or some other group in which there is a control group and a random group. By so placing the

---

[18] Ibid.

[19] Douglas A. Smith, The Effects of Various Item Selection Methods on the Classification Accuracy and Classification Consistency of Criterion-Referenced Instruments, Paper Presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada) 1978, p. 3.

individual into a mutually exclusive category, the intended behavior or instructional objective can be said to have been measured.[20]

Criterion-reference measurement differentiates from normative-reference measurement in that criterion-reference measurement is more likely to be undimensional or homogeneous.[21] Criterion-referenced tests are composed of clusters of items. Those clusters of items are keyed directly to specific objectives and are intended to indicate whether or not the objective has or has not been achieved.[22] Therefore, a criterion-referenced test is one that is deliberately constructed to yield measurements that are directly interpretable in terms of specified performance standards.[23]

---

[20]Ronald A. Berk, A Consumers' Guide to Criterion-Referenced Test Item Statistics, Paper Presented at the Annual Meeting of the National Council on Measurement in Education (Toronto, Ontario, Canada), 1978, p. 2.

[21]Albert C. Crambert, Estimation of Validity for Criterion-Referenced Tests, Paper Presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York), 1977, p. 9.

[22]Ebel, Essentials of Educational Measurement, 1979, p. 351.

[23]R. Glaser and A. J. Nitko, Measurement in Learning and Instruction. In R. L. Thorndike ed. Educational Measurement, Washington: American Council on Education, 1971, pp. 625-670, in Ronald K. Hambleton and William P. Gorth, Criterion-Referenced Testing: Issues and Applications, Paper Presented at the Annual Meeting of the Northeastern Educational Research Association (Liberty, New York), 1970, p. 1.

Because of the homogeneity of the test and the clustering of items around specified objectives, less emphasis is placed on item analysis in the item selection process; however, item analysis is used to a degree.[24]

The uses of the two types of tests, norm-referenced and criterion-referenced, depend largely on what information the user wishes to obtain. The distinctions between the two types of tests are, primarily, what Airasian[25] has called formative evaluation and summative evaluation. Formative evaluation indicates how students are changing with respect to their attainment of the instructional goals. Summative evaluation is end-of-instruction evaluation, primarily to grade student achievement. It provides information with respect to how students have changed relative to course objectives. The significant difference is the verb. Formative evaluation attempts to provide data relative to weaknesses and direct corrective teaching action. Formative evaluation should occur frequently during instruction.[26] When being used in

---

[24]Crambert, Estimation of Validity for Criterion-Referenced Tests, p. 9.

[25]Peter W. Airasian, "The Role of Evaluation in Mastery Learning," in Mastery Learning Theory and Practice, James H. Block, ed. New York: Holt, Rinehart, and Winston, Inc., 1971, p. 78.

[26]Ibid., p. 79.

formative evaluation, criterion reference measurement provide their most important information. In this stage of the evaluation process, data are used by those in charge of developing curriculum to make judgments about how to maximize the probability of learning an established set of objectives.[27]

Both the tailor-made achievement test and the domain-referenced test can be used and inferences can be drawn from them in the same manner as the norm-referenced and criterion-referenced tests. There are, however, some differences between the tailor-made tests, the domain-referenced tests, and the norm-referenced and criterion-referenced tests. The tailor-made test and the domain-referenced test sample opposite ends of the spectrum. The tailor-made test's primary distinction is that it is built for a specific purpose and samples from a constricted domain. Such a test could be commercially prepared or prepared at the local school district level. The domain-referenced test consists of tasks that are sampled from a thoroughly defined population of tasks in such a manner that one can estimate the proportion of tasks in the population

---

[27]John A. Emrick, The Experimental Validation of an Evaluation Model for Mastery Testing, Final Report, Office of Education, Washington, D. C., November, 1971, p. 1.

at which the student is likely to succeed.[28] Tailor-made tests tend to be program oriented while domain-referenced are more global representing the entire domain.

What can be concluded concerning the distinctions between the various types of objective tests is: 1) they are based on a set of objectives, 2) at least as far as administration procedures, they are all "standardized," 3) they may be used as instruments to gather norm referenced or criterion referenced data. Therefore, the proper distinctions are between the more global standardized tests and the more constricted tailor-made tests, and whether the intrepretation is to be criterion-referenced or norm-referenced.[29]

Characteristics of Objectives

Ebel[30] has said that a result of an educational achievement test should be to measure what the process of education has sought to achieve. Therefore, the test constructor must concern himself with educational objectives, objectives that relate to the total process of education

---

[28]Mehrens and Ebel, Some Comments on Criterion-Referenced and Norm-Referenced Achievement Tests, p. 4.

[29]Ibid.

[30]Robert L. Ebel, Essentials of Educational Measurement, 2nd ed., Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972, p. 57.

and objectives that relate specifically to the course, subject, or unit of instruction for which the test was constructed. The test designed should be consistent with the objectives of society, the school, and the test constructor.[31] The objective characteristic, then, is relevance.[32] The advancement in technology since the early 1950's should cause current educators to re-examine their curricular offerings. The relevance issue of objectives raises questions about society's needs and the needs of students. Automation in industry has lessened the demand for great amounts of workers. What, if any, impact does this have on the aims of education? Would a relevant objective deal with career planning? Should students be taught to deal with leisure time because the possibility exists that they will spend less and less time at work?

Another characteristic of an educational objective is feasibility. Feasibility is an umbrella of considerations. It takes into consideration striving for goals that are parallel with what psychologists know about how children develop, how they learn and how they differ one from another in these two respects, as well as whether or

---

[31] Ibid., p. 58.

[32] Mehrens and Lehmann, Measurement and Evaluation in Education and Psychology, p. 20.

not the resources are available to achieve these goals successfully.[33]

Objectives provide guidance. They answer such questions as "Where do I want to go?", "How do I get there?", and "How do I know I have arrived?"[34] In this situation, objectives serve a multiple purpose. They direct the educational process toward the intended educational outcome and at the same time are the desired outcome in stated form. An outcome has been defined as what occurs as a result of an educational experience.[35] In its stated form, an objective directs both the teacher and the learner through the learning process.

To be complete, to provide the means to evaluate successful achievement of the objective, the objective needs to be specific. By adding the element of stated observable performance in which the learner will be engaged during the evaluation process, it becomes possible to determine whether or not the learner has achieved the

---

[33] Ibid., p. 21.

[34] Albert R. Wight, "Beyond Behavioral Objectives," Readings in Measurement and Evaluation in Education and Psychology, Edited by William A. Mehrens, New York: Holt, Rinehart and Winston, 1976, p. 90.

[35] Mehrens and Lehmann, Measurement and Evaluation in Education and Psychology, p. 18.

objective. The objective has become a "behavioral objective."[36]

A behavioral objective is specific and contains an action verb. The behavioral objective describes what the learner will be doing during evaluation. A behavioral objective should not contain the statement that the student will gain an appreciation for the American form of government because the learner can not be observed "appreciating" during the evaluation process. [37] In discussing the construction of criterion-referenced tests, Roudabush[38] states objectives are coherent, clearly stated and specifically describe the behavior the examinee will be able to perform if he has mastered the objective, that is, each objective specifies a limited domain of behaviors. Behavioral objectives provide a basic plan of action for the teacher and the learner from either a pre-instruction or a post-instruction vantage point. Objectives provide both, the teacher and the learner, with the information as to what is expected during the course of instruction and with the information as to the level of achievement after

---

[36] Ibid., p. 19.

[37] Ibid., p. 19.

[38] Roudabush, Item Selection for Criterion-Referenced Tests, p. 3.

instruction. By providing directive guidance to them, objectives take the surprises out of the teaching-learning process.

In their writing of an army training manual, Swezey and Pearlstein maintain that an objective only covers a single task, not a combination of tasks, that the main intent of the objective is clear, and the performance indicators are simple, direct, and part of what the trainee can already do.[39] An objective is composed of three parts: 1) a performance, 2) a condition, and 3) a standard.[40] The performance is what is to be accomplished. It is the task, action, knowledge, skill, or ability required for the job. The condition is the circumstance or situation under which the performance is to be accomplished. The condition might be the tools and equipment required, the materials required, or where it is to be accomplished. For a military trainee, the condition could conceivably be under simulated conditions in the classroom or on a training field "battle-ground." In educational terms, the condition refers to a classroom setting on the one hand or to a "job" situation under other circumstances. The standard is the level or quality of performance. It can be stated in terms of how

---

[39]Robert W. Swezey and Richard B. Pearlstein, Guidebook for Developing Criterion-Referenced Tests, Army Research Institute for the Behavioral and Social Sciences, Arlington, VA., August, 1975, p. 2:9.

[40]Ibid., p. 2:3.

well the performance is to be accomplished or in terms of time, how quickly it is to be accomplished.[41]

An author of behavioral objectives must keep in mind several features and attributes if the objective is to be adequate. It is not enough to include some aspects and exclude the rest. All must be considered. To be relevant, an objective must meet the needs of the society, the student, the school, the instructor. Some modification may be necessary to insure that the objective is feasible. Constraints of money, time, space, materials, and most importantly, the growth and development and the abilitites of the students concerned affect the feasibility of an objective. The objective must be written with enough specificity so as to define and describe its intent and limit it to a single task. The specificity should provide instructional guidance to both the learner and the teacher. It should contain an action verb describing what observable performance will take place during the evaluation process.

---

[41] Ibid., p. 2:7.

Criterion-Referenced Test Construction

Model for Test Construction

The construction techniques of objective-referenced tests, based on the principles of "standardization", is debatable. It is generally accepted that instrument adequacy depends on the extent to which the instrument is capable of assigning individuals to their true level of performance, for example, pass-fail or master-non-master, and the degree to which decisions made are consistent across repeated administrations of the instrument.[42]

These considerations conform to what Cronbach calls the diagnostic purpose of testing, that is, a test appraises the pupil's performance by observing his work on a sample of tasks or items. The sample must be representative of the area being tested and must contain enough items to give evidence which is dependable. To yield dependable evidence, the test must be given in the same way to all students.[43] For an instrument to assign individuals to their true level of performance, it must have objectivity.

---

[42]Smith, The Effects of Various Item Selection Methods of the Classification Accuracy and Classification Consistency of Criterion-Referenced Instruments, p. 1.

[43]Cronbach, Educational Psychology, pp. 549-550.

A measurement is said to be objective if it can be verified by another independent evaluator. Objectivity is not the process by which the measures are obtained, but rather a characteristic of the measure obtained.[44]

Most experts are able to agree with the above definitions and requirements. The methodology for constructing criterion-referenced tests on the basis of conventional statistical processes is the questionable issue. Mehrens and Lehmann represent the opposing point of view to the use of conventional item-analysis procedures in criterion-referenced tests construction quite well. A summary of their point of view is 1) a test item should not be discarded because it does not discriminate providing it does reflect an important attribute of the criterion, 2) a negative discriminator may be caused by one of the several reasons: a) a faulty item, b) ineffective instruction, c) inefficient learning on the pupil's part, and 3) more research is needed before any conclusive answer can be obtained regarding the usefulness of conventional item analysis procedures for criterion-referenced tests.[45]

---

[44]Ebel, Essentials of Educational Measurement, 1979, p. 62.

[45]Mehrens and Lehmann, Measurement and Evaluation in Education and Psychology, p. 334.

There appears to be, however, a consensus of opinion that conventional item analysis procedures are of value in the construction of criterion-referenced tests, and the practice is, in fact, common practice. Douglas U. Smith defends the practice by contending that it is presumptous to think each item comprising a domain of items measures it equally well. The items will vary in difficulty as well as their relationship to the domain. The use of empirical methods of item selection may enhance test characteristics by alleviating some of the subjective judgments in the item writing process.[46]

Since it is generally common practice to use conventional procedures in criterion-referenced test construction, the following steps for criterion-referenced test construction can be developed, based on the work of Rubinstein and Nassif-Royer[47] and Gavin.[48]

---

[46]Smith, The Effects of Various Item Selection Methods on the Classification Accuracy and Classification Consistency of Criterion-Referenced Instruments, . 2-3.

[47]Sherry Ann Rubinstein and Paula Nassif-Royer, The Outcomes of Statewide Assessment: Implications for Curriculum Evaluation, Paper Presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York), 1977, p. 4.

[48]Anne T. Gavin, Guide to the Development of Written Tests for Selection and Promotion: The Content Validity Model. Technical Memorandum 77-6, Civil Service Commission, Washington, D. C.: Personnel Measurement Research and Development Center, June, 1977, p. 6.

Step I: Task Analysis

Step II: Test Plan

Step III: Test Construction

Step IV: Estimate Test Reliability and Content Validity

Task Analysis

Task analysis is defined as the process of determining the purpose and parameters of the test in terms of the subject area and domain to be assessed.[49] An underlying assumption involved in this definition is that the development of the objectives and the definition of the domain to be assessed can be clearly and specifically stated. When the criterion-referenced test is designed to evaluate learning outcomes relative to objectives for a specific curriculum, the likelihood for success of the task analysis process is increased. The reason is that the criterion-referenced test was pioneered for use in the classroom, that is, criterion-referenced tests are generally administered before or after small units of instruction.[50] The greater

---

[49]Rubinstein and Nassif-Royer, The Outcomes of Statewide Assessment: Implications for Curriculum Evaluation, p. 5.

[50]Ronald K. Hambleton and M. R. Norick, "Toward an Integration of Theory and Method for Criterion-Referenced Tests," Journal of Educational Measurement, 1973, 10, 159-170, in Rubinstein and Nassif-Royer, The Outcomes of Statewide Assessment: Implications for Curriculum Evaluation, p. 6.

the diversity of curricula, the broader the task analysis must be defined. Diversity of curricula modifies the purpose of task analysis to imply that the domain being defined is one to which all students have been exposed, a "common ground" area.[51] This appears counter-productive. The more thoroughly defined the domain, the greater the possibility of building a domain referenced test.[52] The closer one comes to building a domain referenced test, the closer one comes to constructing a test sensitive to instruction. A criterion-referenced test begins with a set of objectives representing some curriculum and ends with reporting performance on each of those objectives. It should discriminate well between mastery and non-mastery of each of the objectives making up the curriculum of interest as opposed to a good norm referenced test discriminating well between examinees who have differing amounts of achievement in a general area of interest.[53]

---

[51]Rubinstein and Nassif-Royer, The Outcomes of Statewide Assessment: Implications for Curriculum Evaluation, p. 8.

[52]Mehrens and Lehmann, Measurement and Evaluation in Education and Psychology, p. 110.

[53]Roudabush, Item Selection for Criterion-Referenced Tests, p. 2.

Test Plan

If one wishes to travel from New York to California by car, one has several options available. One can randomly strike out and hope his sense of direction is sufficient to plot a course which will lead him to California. One can install a compass in his car and use it as a guide until he finally reaches California. In each case the probability of reaching the destination rests on several considerations. One would have to ask oneself if he were willing to invest the time and money, not to mention patience, to embark on such a journey. The logical course of action to follow if one wished to complete such a journey in an efficient and effective manner would be to use a map showing the major highways and the most direct route from New York to California.

The construction of a test is no different than planning a trip. One must have a plan of action, a guide determining the direction the test will take. The test plan becomes the directing force for the test. It defines, outlines, maps out the test. The test plan is, indeed, the table of specifications for the test constructor. Using a table of specifications provides that a) only the objectives involved in the instructional process will be assessed, b) each objective will receive a proportional amount of emphasis on the test in the same relation as the emphasis placed on that objective instruction, and c) no important objective or content area will be accidentally

omitted.[54] To be assured the table of specifications will yield these provisions, a set of explicit specifications should be observed. The following list is a summary of Ebel's suggestion as to what a table of specification should contain:

1. The forms of the test items to be used
2. The number of items of each form
3. The kinds of tasks the items will present
4. The number of tasks of each kind
5. The areas of content to be sampled
6. The number of items in each area
7. The level and distribution of item difficulty[55].

As the level of difficulty of intellectual objectives varies, so does the level of difficulty of test items vary. The form of the test item becomes one of the determiners of the level of difficulty. The form may be of the true-false variety, the completion (fill-in-the-blank) type, matching one column of items to their correct response in another column, or the multiple-choice method.[56] The decision must be made as to which type (form) of item is to be used.

---

[54]Mehrens and Lehmann, Measurement and Evaluation in Education and Psychology, p. 179.

[55]Ebel, Essentials of Educational Measurement, 1979, p. 69.

[56]Swezey and Pearlstein, Guidebook for Developing Criterion-Referenced Tests, p. 3:14-15.

Matching, completion, classification types of items, and short answer can be effectively used, but they have more limited applicability. The true-false form and the multiple-choice form will measure any aspect of cognitive educational achievement. What is measured by the true-false item or the multiple-choice item is determined more by its content than its form.[57]

The kinds of tasks the items will present will be determined by the objectives as defined through the task analysis process. Practical constraints such as time and cost will have a bearing on the number of items selected to measure the individual objectives.[58] The purpose of the test and the information desired, as well as the scope of the area to be measured, will determine the number of objectives to be measured. Measuring too many objectives, each with several items, causes the test length to increase. Decreasing the number of items per objective effects the reliability of the test. The reliability of a test is its ability to measure the same thing through repeated administrations of the test.[59] For the estimate of reliability to be held stable, an objective must be measured

---

[57]Ebel, Essentials of Educational Measurement, 1972, p. 103

[58]Swezey and Pearlstein, Guidebook for Developing Criterion-Referenced Tests, p. 1:10.

[59]Ibid., p. 1:11.

by at least four items. This would allow up to twenty-five objectives to be measured. However, varying item lengths would realistically bring the number of objectives closer to fifteen.[60]

Defining the content domain becomes a definition based on practical concerns. The content validity of a test has been defined as based on a hypothetical universe of situations.[61] A "universe of situations" is the whole collection of measurements that might have been made.[62] An attempt to define all possible situations would be subject to severe criticism. It would be subjective rather than objective. It would be prohibitively costly in terms of human effort. It would be unmanagably long and detailed to the extent its usefullness would be questionable. The result is that most criterion-referenced test are of the "content-specified" approach on the basis of a listing of the intended educational outcomes of the institution, a

---

[60]Rubinstein and Nassif-Royer, The Outcomes of Statewide Assessment: Implications for Curriculum Evaluation, p. 11.

[61]Roger T. Lennon, "Assumptions Underlying the Use of Content Validity," Readings in Measurement and Evaluation in Education and Psychology, Edited by William A. Mehrens, New York: Holt, Rinehart and Winston, 1976, p. 46.

[62]Marsha M. Linehart, Content Validity in Behavioral Assessment, Paper Presented at the Annual Meeting of the American Psychological Association (84th, Washington, D. C.), 1976, p. 3.

table of specifications, or some other means of detailing the intended content of the test. In a criterion-referenced test, the universe of items can be described, but not fully defined. The criterion-referenced test is considered to be only illustrative of the universe and not a sample of it.[63]

For a test to be content valid, the table of specifications requirement for the determination of the number of items to be used in each of the content areas to be sampled takes on added importance. A factor in determining the content validity of a test is documenting that the behaviors demonstrated in the test constitute a representative sample of the behaviors to be exhibited in the desired content domain.[64] If a reading instructional program devotes twenty percent of its presentation to structural analysis, ten percent of its presentation to phonic analysis, sixty percent of its presentation to the various aspects of comprehension skills, and ten percent of its presentation to study skills, the number of items should be appropriately proportioned.

---

[63]Crambert, Estimation of Validity for Criterion-Referenced Tests, p. 6.

[64]Michigan Educational Assessment Program, Technical Report, (Lansing, Michigan : MDE), 1977, p. 13.

Test Construction

At the very heart of a criterion-referenced test, specifically, or any test in a more general sense, is the "item," the "thing" that is scored as correct or incorrect. It is the item which ultimately determines the content validity of the test.[65] It is the item which, joined with other items, measures the educational objective, the desired outcome toward which the learning process is being directed. The selection of the item(s) for a test, and a criterion-referenced test in particular, is of prime importance in the test construction process. The match between the item and its objective is determined by the objective. The specificity of the objective is the factor which determines the restrictions placed on an item writer's freedom to alter the original intent of the objective. Generally, objectives written in vague generalities give item writers latitude to define the tasks required by the objective. The greater the specificity of an objective, the more likely will be the precision of the item which measures it.[66]

The item which is selected for inclusion in the final form of the test comes from an item pool. Swezey

---

[65]Michigan Educational Assessment Program, Technical Report, (Lansing, Michigan: MDE), 1977, -. 13.

[66]William Mehrens, Technical Report: The Fifth Report of the 1973-74 Michigan Educational Assessment Program. Michigan State Department of Education, Lansing, Michigan, 1975, p. 16.

and Pearlstein,[67] Rubinstein and Nassif-Royer,[68] and Roudabush[69] suggest the item pool comes from one of two sources. Either totally new items are generated by item writers or items could be obtained from existing item pools. Authoring original items offers the probability of a higher degree of precision in the match between item and objective. A constraint placed on this approach is cost: cost in terms of paying for the writers' time to develop the items themselves. Drawing items from an existing item pool saves time and money; however, a decrease in the precision of correspondence between the objective and the item may cause a mismatch between the objective and the item and require a modification of the original objective.

Once a pool of items has been established, one of two processes may be observed in selecting which items will be included in the test. Items may be included through empirical item sampling or random sampling from the item pool. One empirical item sampling method represents

---

[67]Swezey and Pearlstein, Guidebook for Developing Criterion-Referenced Tests, p. 1:10.

[68]Rubinstein and Nassif-Royer, The Outcomes of Statewide Assessment: Implications for Curriculum Evaluation, p. 14.

[69]Roudabush, Item Selection for Criterion-Referenced Tests, p. 3.

selecting items that show the greatest difference in item difficulty computed from uninstructed-instructed samples. The uninstructed-instructed sample consisted of two-hundred-fifty-eight dental students who were administered two forms of a 100-item test. The data were analyzed on two types of samples: 1) a post-instruction sample representing instructed students, and 2) a pre and post-instruction sample representing the full range of attainment in the achievement domain. The test contained both knowledge of basic dental anatomy and a collection of items defined by objectives of the text. The conclusion was that tests which are created by random sampling seem to provide the smallest errors of measurement.[70]

Smith, on the other hand, suggests the use of item selection procedures does not necessarily affect the content validity of the instrument because the developer could select only the most highly discriminating items and remain with the original test plan, retaining the same category proportions as the original item pool. The empirical approach to item selection may enhance the test characteristics by alleviating some of the subjective

---

[70]Tom Haladyna and Gale Roid, A Theoretical and Empirical Comparison of Three Approaches to Achievement Testing, (New York: ERIC Document Reproduction Service, Education 148903, May, 1978), pp. 10-18.

judgments in the item writing process.[71] However, because of the particular significance in content-referenced measurement of the relationship between the instructional objectives and the test content, it is necessary that the test development procedure be designed and executed with greater care and higher standards for consensus judgment than are usually thought to be necessary for norm-referenced measurement.[72]

Item Analysis

A characteristic of a behavioral objective has already been identified as an observable performance in which the learner will be engaged during the evaluation process. The prupose, then, of the items in a criterion-referenced test is to measure behavior in relation to the instrumental objective. Item analysis is a procedure designed to express the degree or relationship between the intent of each item and the responses of the students to each item. Nineteen different statistics have been identified as having the ability to provide quanitiative evidence of item validity.[73]

---

[71]Smith, The Effects of Various Item Selection Methods on the Classification Accuracy and Classification Consistency of Criterion-Referenced Instruments, p. 3.

[72]Crambert, Estimation of Validity for Criterion-Referenced Test, p. 9.

[73]Berk, A Consumers' Guide to Criterion-Referenced Item Statistics, p. 2.

Before item analysis can be performed, the responses of the students to each item must be tabulated.[74] Tabulation of student response to the various items yields a variety of information. An Index of Item Difficulty can be computed through calculating the proportion of students who responded to the item correctly:

$$Diff = X/N$$

where

X = the number of students responding correctly

N = the total number of students tested[75]

The result, the level of difficulty, is what has been referred to as a proportion score ("P" score) or an expression representing the frequency of correct responses to an item, giving the proportion of the total number of examinees tested who answered the item correctly. An increase in the score indicates an easier item with a lower degree of discriminating power. The maximum level of item discrimination occurs with a "P" score of 0.50. As the "P" score approaches a perfect 1.00 or 0.00, the item becomes useless because the

---

[74] James E. Wert, Charles O. Neidt and J. Stanley Ahman, Statistical Methods in Educational and Psychological Research, New York: Appleton-Century-Crofts, Inc., 1954, p. 338.

[75] Ibid.

level of difficulty is either extreme. The frequency is all or none.[76]

Estes, Colvin, and Goodwin[77] validated the items in their criterion-referenced test by using Truman Kelly's discrimination method of extreme groups. Kelly has demonstrated that using extreme groups, each formed by approximately 27 percent of the total group, the ratio of the difference in average abilitites of the groups to the standard error of their difference is maximum.[78] In so doing, Estes, et. al., used the following

$$D = H-L/N$$

where

H = the number of students in the top 27 percent who responded correctly

L = the number of students in the lower 27 percent who responded correctly

N = the number of students 27 percent represents

---

[76]David Magnusson, Test Theory, Trans. by Hunter Mabon, Reading, Mass.: Addison-Wesley Publishing Company, 1967, p. 219.

[77]Gary Estes, Lloyd W. Colvin and Coleen Goodwin, A Criterion-Referenced Basic Skills Assessment Program in a Large City School System, Paper Presented at the Annual Meeting of the American Educational Research Association (60th, San Francicso, California), 1976, p. 7.

[78]Truman Kelley, "The Selection of Upper and Lower Groups for the Validation of Test Items," Journal of Educational Psychology, Vol. 30, (1939), pp. 17-24, in Robert L. Ebel, Essentials of Educational Measurement, 2nd ed., Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972, p. 386.

and selected items whose discrimination value was at least 0.20 and whose difficulty value fell between 0.40 and 0.80.[79] There is a degree of variation in the field of measurement concerning the range of values. Nunnally establishes the range of difficulty values as 0.20 to 0.80,[80] while Ebel establishes the discriminating level for test items as 0.30 and up as reasonably good and 0.40 and up as very good items. Items below 0.29 are marginal to poor items.[81]

A very useful and frequently used statistic in item anlaysis is the one which Magnusson[82] referred to as a "short-cut" method which investigates differences between extreme groups on the test and the criterion distributions respectively. It is the Phi Coefficient, symbolized by "Ø".

---

[79]Estes, Colvin and Goodwin, A Criterion-Referenced Basic Skills Assessment Program in a Large City School System, p. 8.

[80]Nunnally, Educational Measurement and Evaluation, p. 188.

[81]Ebel, Essentials of Educational Measurement, 1979, p. 267.

[82]Magnusson, Test Theory, p. 198.

The Phi Coefficient is written

$$\phi = \frac{AD - BC}{(A+B)\ (C+D)\ A+C)\ (B+D)}$$

To validate the item, one needs to know 1) the masters and non-masters who pass the item, and 2) the masters and non-masters who did not pass the item.[83] The following illustration demonstrates how the item can be validated and the value of the information received through the process. The illustration is a summarization of work by Swezey and Pearlstein[84] and Edmonston and Randall.[85]

---

[83]Swezey and Pearlstein, Guidebook for Developing Criterion-Referenced Test, p. 5:11.

[84]Ibid., pp. 5:8-9.

[85]Leon P. Edmonston and Robert S. Randall, A Model for Estimating the Reliability and Validity of Criterion-Referenced Measures, Paper Presented at the Annual Meeting of the American Educational Research Association (56th, Chicago, Ill.), 1972, pp. 16-20.

Student	Master and Non-Master	Item Number 1	2	3	4	5	6	7	8	Number Items Passed
1	M	P	P	P	P	P	P	P	P	8
2	M	P	F	P	P	F	P	P	P	6
3	M	F	P	P	P	P	F	F	F	4
4	NM	P	P	F	P	P	F	P	P	6
5	NM	P	F	P	F	F	F	F	F	2
6	NM	F	F	P	F	F	F	F	F	1
Number Masters Passed		2	2	3	3	2	2	2	2	18
Number Non-Masters Passed		2	1	2	1	1	0	1	1	9
Total Number Passed		4	3	5	4	3	2	3	3	27

To compute the Phi coefficient for Item 4, the following grid will be used

	Item Number 4	
	Pass	Fail
Masters	a 3	b 0
Non-Masters	c 1	d 2

Substituting the values in the grid into the preceeding formula

$$\phi = \frac{(3)(2) - (0)(1)}{\sqrt{(3+0)(1+2)(3+1)(0+2)}}$$

$$= \frac{(6) - (1)}{\sqrt{(3)(3)(4)(2)}}$$

$$= \frac{5}{\sqrt{72}}$$

$$= 5/8.485$$

$$= .589 \cong .59$$

The total range of the Phi coefficient is from -1.00 through zero to +1.00. An item has acceptable discriminating power if its score falls between 0.30 and 1.00.[86] It could be concluded that the sample item above would be acceptable for inclusion in the test. The same computations should be completed for each item.

While there is a lack of consensus of opinion as to whether or not conventional methods of test construction should be used in the construction of objective-referenced tests, there appears to be a sufficient body of information where the conventional methods have been used successfully. The use of conventional methods of test construction tends

---

[86] Swezey and Pearlstein, Guidebook for Developing Criterion-Referenced Test, p. 5:12.

to identify items capable of discriminating in such a manner so as to satisfy a purpose of criterion-referenced measurement, that is, classifying individuals into mutually exclusive categories.

Test Validity

Types of Validity

A survey of the rather extensive amount of literature pertaining to test validity yields discussions of many varieties of validity. As the varieties increase, some minor changes in interpretation begin to appear. Lists have been compiled which provide definitions for these varied forms of validity. One of these lists contains ten different varieties of validity.[87] However, the American Psychological Association delimits only three kinds of validity: 1) construct validity, 2) criterion-related validity, and 3) content valididty.[88]

Construct Validity

A construct has been described as an attribute of people assumed to be reflected in test performance.[89]

---

[87]Ebel, Essentials of Educational Measurement, 1972, pp. 436-437.

[88]Mehrens and Lehmann, Measurement Evalaution in Education and Psychology, p. 110.

[89]Haladyna and Roid, A Theoretical and Empirical Comparison of Three Approaches to Achievement Testing, p. 2.

A construct, then, is a psychological trait. Construct validity is the measurement of a psychological trait, not of the trait itself but of the presence of the trait.[90] The items in a test designed to test logic measure a person's tendency to think logically in a given situation. Personnel specialists have the option of either administering a written test or require an applicant to perform the acutual job for which the application has been made. For reasons of health and safety, it might not be practical to "perform" the actual job. In this situation the written test would be preferable. The test is assumed to contain the constructs to measure the necessary attributes required to perform the job.[91]

Criterion-Related Validity

Criterion-related validity applies to the relationship between the scores on a test and an independent external measure.[92] If the personnel specialist, from the above illustration, decided on the basis of the test scores to

---

[90]Nunnally, Educational Measurement and Evaluation, p. 31.

[91]Gavin, Guide to the Development of Written Tests for Selection and Promotion: The Content Validity Model. Technical Memorandum 77-6. p. 2.

[92]Ibid.

employ the applicant, the personnel specialist could determine the criterion-related validity of the test according to the degree of success or failure of the applicant's job performance (external criterion). What criterion-related validity permits the test user to do is predict. In criterion-related validity, the aim is to determine how well one can generalize from one score to another.[93] If the comparison of test results is with data gathered at the same time as the time of test administration, it is said to have concurrent validity. However, if the comparison of test results is with data collected at some future date, it becomes predictive validity.[94] In either case (predictive validity or concurrent validity) they are both concerned with prediction.[95]

In education, measurement is primarily concerned with achievement. The measurement may concern itself with assessment of student knowledge across a broad, general

---

[93]Mehrens and Lehmann, Measurement and Evaluation in Education and Psychology, p. 112.

[94]Swezey and Pearlstein, Guidebook to Developing Criterion-Referenced Tests, p. 7:6.

[95]Mehrens and Lehmann, Measurement and Evaluation in Education and Psychology, p. 112.

area of study or it may concern itself with assessment of student mastery of the goals and objectives of the course of instruction. In either situation, the relationship of test content to the course content is of prime importance. In terms of validity, this relationship is referred to as content validity.

The American Psychological Association has stated in its Standards for Educational and Psychological Tests that to demonstrate the content validity of a set of test scores, it must be shown that the behaviors demonstrated in testing constitute a representative sample of the behaviors to be exhibited in a desired performance domain.[96] Therefore, there are three components to the content validity of a test: 1) the behavior to be exhibited in the performance domain, 2) the behavior to be demonstrated in testing, and 3) the strength of the relationship between the two.[97]

The establishment of content validity is essentially an inference of the adequacy of the sampling process. The inference of content validity requires a judgment that the specified content domain has been adequately sampled by

---

[96]American Psychological Association, Standards for Educational and Psychological Tests, p. 28.

[97]Mark D. Spool, Performing a Content Validity Study, Paper Presented at the Annual Meeting of the Southeastern Psychological Association (21st, Atlanta, Ga.) 1975, p. 3.

the test. The issue is one of reasonable (not statistical) representativeness. The term "representativeness" refers to both the types of behaviors assessed and the proportional coverage of the different knowledge, skills, and abilities.[98] The establishment of content validity is an inference, but not an ideal inference. It is a careful judgment, based on the test's apparent relevance to the behaviors which are legitimately inferable from those delimited by the criterion.[99] The establishement of content validity through careful judgment requires that specific procedures be followed to assure the accuracy of the validation process. One model for those procedures is 1) a thorough and accurate analysis of the content domain, 2) a review and evaluation of the test by experts, 3) a comparison between the test content and the instructional content to assess the extent of the relationship between the two, and 4) document each

---

[98]Gavin, Guide to Development of Written Tests for Selection and Promotion: The Content Validity Model. Technical Memorandum 77-6, p. 4.

[99]W. James Popham and T. R. Husek, "Implications of Criterion-Referenced Measurement," Journal of Educational Measurement, 1969, 6, 1-9, in Ronald K. Hambelton and William P. Gorth, Criterion-Referenced Testing: Issues and Implications, Paper Presented at the Annual Meeting of the Northeastern Educational Research Association (Liberty, New York), 1970, p. 14.

procedure of the study.[100] Although not specifically stated, there have been several studies conducted regarding content validity which have approximated this model.

Related Studies

Tallmadge and Horst[101] conducted a study related to the validity of achievement tests and the instructional programs used by local school districts involved in Title I federal programs. Their hypothesis was that not all achievement tests are sensitive to achievement gains. The purpose of their study was to argue against Title I policy allowing only one standardized test to be used as a measure of achievement gains due to the effect of Title I assistance to children with reading difficulties.

The study analyzed the instructional programs of Houghton-Mifflin, Ginn and Company, and Economy. The standardized tests were the California Achievement Test and the Metropolitan Achievement Test.

The report indicated that a poor correlation was found to exist between the instructional programs and the tests.

---

[100]Spool, Performing a Content Validity Study, p. 3.

[101]G. Kasten Tallmadge and Donald P. Horst, The Use of Different Achievement Tests in the ESEA Title I Evaluation System, Paper Presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontarion, Canada), 1978, pp. 4-8.

The conclusion is, it seemed highly probable that when the content of a test shows a low correlation with the content of a curriculum, the test will be insensitive to whatever achievement gains the curriculum might produce. The conclusion further emphasized that the only valid way to assess the effects of an instructional treatment is to use a test that measures what is taught, a test in which the items are samples from the same domains as the teaching-learning exercises.

While the results of the study are founded on the procedures to be followed in a content validation procedure, the basic issue, and therefore, the major weakness of the study, is the usage of conventional instructional programs in an unconventional fashion which results in an inappropriate application of the standardized tests. The conclusion reached, probably would have been the same had they addressed the basic issue rather than their hypothesis. Only the means of achieving the conclusion "might" have been different.

The Tallmadge and Horst study reflects an attempt to evaluate the behaviors required in the performance domain, the behaviors to be demonstrated in testing, and the interrelationship between the two. It is not an easy task. There are some features which may add to the strength of such a study.

Tanenbaum and Miller[102] formulated rules to incorporate into their procedure to compensate for what they felt to be deficiencies in instructional material outlines, tests, and teaching strategies. They devised two files: 1) showed curricula taught, and 2) showed curricula keyed to the test. These files were devised as a result of finding the outlines provided by the publishers were, in their opinion, not sufficiently precise. These files formed their own description of the content and the criterion for each item. A strategy of "near transfer" was adopted. All features had to be represented in the curricula exactly as they were found in the test format. They established the level of readability on the Dale-Chall formula. To compensate for the fact that not all teachers teach to the same degree, a word was considered taught if a pupil was exposed twice to curricula that contained the word in a well marked exercise. Using these guidelines, they conducted an evaluation of Project Information Packages (PIPS). A content analysis was performed to detect the congruence between the Metropolitan Achievement Test and six exemplary compensatory education program curricula.

---

[102]Arlene B. Tanenbaum and Christine A. Miller, The Use of Congruence Between the Items in a Norm-Referenced Test and the Content in Compensatory Education Curricula in the Evaluation of Achievement Gains, Paper Presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York), 1977, pp. 1-10.

Fall-spring testing patterns (fail-pass; pass-pass; pass-fail; fail-fail) were tallied to compare performance on congruent and non-congruent items. Eventually, a model factorial design was devised to incorporate the variables which appear to influence the patterns of achievement.

The results of the study appear very small. The degree of congruence appears to fall between 5 percent and 20 percent and decreases with an increase in grade level from grade four to grade eight. The results show that the amount of congruence was too small to make strong inferences about the quality of the PIP education programs.

The merit of this study lie in its attempt to define the domain and to compensate for the differences in teaching strategies. However, the addition of factor analysis appears to have altered the results markedly.

The work of Jenkins and Pany[103] underscores the need for a high correlation of relationship between the behaviors in the performance domain and those to be demonstrated in testing. Their research was directed toward detecting bias in achievement tests. To detect the extent of bias, Jenkins and Pany studied five standardized tests and seven first and second grade commercial reading series.

---

[103]Joseph R. Jenkins and Darlene Pany, "Curriculum Biases in Reading Achievement Test," Journal of Reading Behavior, Vol. x, No. 4. (Winter, 1978), pp. 345-357.

The procedure which they used was to use publishers' guides to determine which books were used in first and second grade levels and teachers' manuals to compile alphabetical word lists for each book in the series. Next, alphabetized lists of all words in the standardized tests of word recognition were prepared. By comparing the two lists, the extent of overlap could be established by determining the total number of word matches per grade level.

The results of their study indicate that expected annual growth would vary according to which test was administered in conjunction with which curriculum was in use. They concluded that it is doubtful that the use of conventional achievement tests can provide an unbiased estimate of a curriculum's effect, at least with regard to the early grades.

The significance of their work is that the combination of curriculum being used and the tests which are administered can be manipulated to affect the achievement gain scores. While this is an issue concerning the misues of tests and test results, it holds a high degree of relationship to content validity. The level of bias was directly proportional to the degree of congruence between the tests and the curricula.

One aspect of the work of Jenkins and Pany is the item-by-treatment interaction. Their word lists were created

from the instructional materials. Freeman, et. al.,[104] have completed a study of four commercial achievement tests in elementary school mathematics. For their analysis they devised a taxonomy of mathematics. The taxonomy consisted of a classification matrix which had three dimensions: 1) mode of presentation, 2) nature of material, and 3) operation, which specified the process which was required.

They concluded that there are striking differences between the content covered by the four most commonly used standardized tests of elementary school mathematics. They also concluded that significant discrepancies between the content a teacher presents to students and the content which is being tested on the standardized tests administered are likely to exist. These mismatches have a negative effect on the use of standardized tests for instructional purposes. In order to diagnose student strength or weaknesses or to diagnose program strength and weaknesses, either the program must be modified or the test must be selected with extreme care to insure a proper match.

---

[104]Freeman, Kuhs, Knappen, and Porter, A Closer Look at Achievement Tests, pp. 1-10.

Summary

The pressure of accountability is being applied with more intensity today than it has been in several decades. One type of response to the pressure has been for state Departments of Education to implement accountability programs. The programs have, as a major component, a mandated assessment test.

To mandate an assessment test means that an evaluation of someone or something will occur. Therefore, there needs to be a definition of the purpose of evaluation. The definition of the purpose of the evaluation process, as presented in this chapter, is to describe or represent a person. The function of the evaluation process is to aid in the decision-making process. If the evaluation process does not accomplish that function, the process is considered useless, a waste of time for both the evaluatee and the evaluator.

The evaluation process, as it relates to education, consists mainly of paper-and-pencil tests. There are two categories of tests: 1) an essay form, that is, a written narrative, and 2) an objective form, that is a short answer variety which does not require the student to provide the answer completely on his own. Within the category of objective tests, a variety of types are identified. The basic distinctions, however, were between

the more global standardized tests and the more constricted tailor-made tests, and whether the interpretation was to be norm-referenced or criterion-referenced.

All good achievement tests are objective referenced. Of particular interest is the behavioral objective and the characteristics which make up the objective. To write a behavioral objective, certain attributes must be included if the behavioral objective is to be useful and capable of being assessed. An objective must be relevant and feasible. An objective must be written with enough specificity to limit the objective to a single task and describe and define its intent. An objective describes what observable performance will be taking place during the evaluation process. A behavioral objective contains three parts: 1) the performance, 2) the condition, and 3) the standard.

Although opinions differ concerning the use of conventional methods of test construction, there appears to be a sufficient body of information where the conventional methods have been used successfully. The use of conventional methods of test construction tends to identify items capable of discriminating in such a manner as to satisfy a purpose of criterion-referenced measurement, that is, classifying individuals into mutually exclusive categories.

While many varieties of validity appear in the literature, the American Psychological Association delimits only three: 1) construct validity, 2) criterion-related validity, and 3) content validity. Of these three, educational assessment is primarily concerned with content validity. From the definition, it can be said content validity is composed of three components: 1) the behavior to be exhibited in the performance domain, 2) the behavior to be demonstrated in testing, and 3) the strength of the relationship between the two.

The establishment of content validity is based on careful judgment of the test's apparent relevance by using a thorough and accurate analysis of the content domain and the content of the test.

Several studies have been identified which indicate that the relationship between the content of several widely used instructional programs and the content of several of the more popular standardized achievement tests is suspect. The studies have revealed that the degree of match between a program and a test will vary depending on which program is matched with which test.

The significance of the related studies to the study currently under investigation is that the present study is attempting to establish the degree of concurrance between the Michigan Educational Assessment Program Experimental Reading Test (a criterion-referenced test) and five reading instructional programs.

CHAPTER III

METHODOLOGY OF THE STUDY

The present study is based on a design that makes possible the determination and analysis of the concepts presented in the five reading instructional programs and the concepts tested in the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven as measured by the Reading Concepts Checklist, (RCC).[1]

Development of the Instrument and Its Use

The Instrument

The Reading Concepts Checklist, (RCC), was developed as a means of describing, within a common framework, the concepts presented in the instructional materials and the concepts tested in the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven. It was recognized at the beginning of this study that terminology and definitions would vary to some degree across

---

[1]Appendix A

specialists. The goal, therefore, was to develop the Reading Concepts Checklist, (RCC), on the basis of conceptual consensus of agreement to insure that its terms and definitions would have a high degree of meaning and similarity of meaning across reading specialists and test constructors.

The construction of the Reading Concepts Checklist, (RCC), was based on the work of recognized authorities in the field of reading. Cohen and Hyman,[2] Barbe,[3] and Ekwall[4] agree, generally, upon the major divisions of the Reading Concepts Checklist, (RCC). Duffy and Sherman[5] use terminology which is different, but have basically the same divisions as the Reading Concepts Checklist, (RCC). Reid's[6]

---

[2]Alan S. Cohen and Joan S. Hyman, Instructional Objectives in Reading, New York: Random House, Inc., 1977, pp. 1-8, 15-19.

[3]Walter B. Barbe, Personalized Reading Instruction, 9th Printing, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1967, pp. 142-143, 152-153, 160-161, 168-169, 182-183, 192-193, 204-205.

[4]Eldon E. Ekwall, Diagnosis and Remediation of the Disabled Reader, 2nd Printing, Boston, Mass.: Allyn-Bacon, Inc., 1976, pp. 59-61.

[5]Gerald G. Duffy and George B. Sherman, Systematic Reading Instruction, 2nd ed., New York: Harper and Row, 1977, p. 82.

[6]Ethna R. Reid, Teaching Literal and Inferential Comprehension, Salt Lake City, Utah: Cove Publishers, 1978, pp. 10-11.

overall structure agrees with the Reading Concepts Checklist, (RCC); however, Reid subdivides the categories into greater detail than that contained in the Reading Concepts Checklist, (RCC). The six major divisions of the Reading Concepts Checklist, (RCC), are

1. Auditory Discrimination
2. Visual Discrimination
3. Phonic Analysis
4. Structural Analysis
5. Comprehension, and
6. Study Skills.

Each major category was subdivided into its predominant categories and numerically coded for later use with a computer program. The Reading Concepts Checklist, (RCC), formed the basis for two matrices: 1) the classification of concepts presented in the instructional materials in grades K-6, and 2) the classification of the concepts tested in the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven.

The Use of the Instrument

The matrix developed for use with the instructional materials consisted of the Reading Concepts Checklist, (RCC), being placed down the left side and the K-6 grade levels being placed across the top. The five instructional programs were alphabetically ordered and chronologically numbered. If a given program presented a concept at any or all grade levels, the code number representing the

program was placed in the cell formed by the intersection of the concept and the appropriate grade level. To determine which concepts were presented at the various grade levels, each teacher's manual for each grade level was examined in its entirety. The process was repeated for each of the five instructional programs.

The form of the matrix for the classification of the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test differed from that used for the instructional programs in that only two categories were placed across the top of the matrix. They were "Grade 4" and "Grade 7."

The classification of the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test was conducted independently by the researcher and three reading experts.

The materials used to implement the classification of the test's concepts consisted of the matrix, a copy of the draft copy of the Michigan Department of Education's Communication Skills Objectives: Reading,[7] response keys for the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven, and a copy of the Experimental Test for grades four and seven.

---

[7]Appendix B

Each of the two levels of the test consists of one-hundred-forty-one items which measure six major categories of reading skills. The sixth category, "Positive Responses to Reading," and the test items 126-141, are attitudinal in nature and have no "correct" response. Therefore, the sixth category, and its accompanying items, were not classified into the matrix. The other five categories are

1. Vocabulary Meaning
2. Literal Comprehension
3. Inferential Comprehension
4. Critical Reading Skills, and
5. Related Study Skills

The instructions given each judge were to match each item with its stated objective, read each test item and determine the nature of the task being required of the examinee. Finally, based on the above determination, the judges were instructed to list the category, objective and the test item number in the appropriate cell of the matrix according to the grade level and the concept. Each item of the test was treated in the same manner until all 125 items had been included in the matrix by the judges. The process was followed for each level of the test.

Selection of Instructional Materials

The process of selecting comparison materials involves such questions as "What are the predominant instructional programs in use in Michigan's public schools?" and "What combinations of those programs are used by a majority of

Michigan's Kindergarten through sixth grade students?" In defining the term "majority", several aspects were taken into consideration. First, a majority should be a clearly definitive number, not simply "more than half." Next, a majority should be large enough so as to insure a population of students large enough to be exposed to the defined content domain, a domain from which a representative test sample could conceivably be taken. Finally, a majority should be large enough that it represents a reasonable cross-section of Michigan's rural, suburban, urban, and large-city school children. Therefore, based on these considerations, the lower acceptable limit which defined a majority of students using the reading instructional materials to be included in the present study was established at seventy-five percent.

The basis for answering these questions and selecting the reading instructional materials for this study is the result of a 1977 national survey of reading instructors and reading supervisors. Market Data Retrieval, Inc. mailed 11,889 questionnaires to reading instructors and reading supervisors. Of that number, 2052 valid responses were used which made the response rate 17.3 percent.[8] Although

---

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

the agency was under contract to a particular publishing company, the questionnaire appears to be free from bias toward any publisher. The survey results provided statistics for both national and regional levels of the market share captured by the several publishing companies. The survey revealed that the predominant reading instructional materials used in the region which included Michigan are: 1) Ginn and Company, 2) Harcourt, Brace and Jovanovich, 3) Holt, Rinehart and Winston, 4) Houghton-Mifflin Company, and 5) Scott, Foresman Company.[9] The survey also provided data which satisfied the lower acceptable limit definition of seventy-five percent of Michigan's K-6 students using the reading instructional materials. The survey indicates the percentage to be 75.86.[10] The following table indicates the distribution of students using the reading instructional materials by area. The table does not indicate the market share of the publishers. It illustrates the concentration of the publications according to the three types of areas nationally.

---

[9] Ibid., p. 5.

[10] Ibid.

DISTRIBUTION OF STUDENTS USING TEXTS OF THE FIVE MAJOR PUBLISHERS BY AREA[11]

Publisher	Urban	Suburban	Rural
Ginn and Company	24.9	37.9	36.2
Harcourt, Brace and Jovanovich	24.0	34.5	41.4
Holt, Rinehart Winston	38.4	29.5	31.8
Houghton-Mifflin Company	23.4	26.2	49.9
Scott, Foresman Company	11.4	39.6	47.8

The national and regional levels of information which this survey provided permits a high degree of confidence to be placed in the assumption that the five reading instructional programs selected for this study do, indeed, constitute those programs which are the predominant programs in use in Michigan's K-6 grades and are used by least seventy-five percent of Michigan's K-6 students.

Treatment of the Data

Due to the nature of the Michigan Educational Assessment Program Experimental Reading Test, the data which had been compiled in the instructional materials classification matrix were grouped into a K-3 category to be

[11] Ibid., p. 7.

compared with the Grade 4 Test and a 4-6 category to be compared with the Grade 7 Test. A concept was considered presented if it appeared in the K-3 or 4-6 category. The matrix was then reduced to dichotomous data in either of the K-3 or 4-6 instructional levels. Concepts which were presented were assigned a numerical value of "1" while concepts which were not presented were assigned the value of "0".

The data were then punched and verified for IBM and computer tabulation. A separate set of data cards was prepared for the K-3 and 4-6 levels. The IBM card layout used nine columns, providing for the identification of each individual concept (3 columns); individual instructional program concept data (5 columns). The final column was reserved for data pertinent to the test. Printed IBM listing from card data was completed to facilitate computations for further statistical tests and to recheck the completeness and accuracy.

The compiled data from the test classification matrix was also converted to dichotomous data. A concept was considered tested if one or more test items were identified by the judges as measuring that concept. Concepts which were tested were assigned a numerical value of "1" while concepts which were not tested were assigned the value of "0". Value assignment was based on majority agreement among three of the four judges.

The data were then punched and verified for IBM and computer tabulation. A separate set of data cards was prepared for the Grade 4 and Grade 7 tests. The IBM card layout used seven columns, providing for the identification of each individual concept (3 columns) and invididual judges' responses (4 columns).

Statistical Methodology and Research Design

Research Design

A statistical test may be termed nonparametric if it does not test a hypothesis which characterizes one of the parameters of the parent variable of interest. Or, a statistical test may be termed distribution-free if the sampling distribution of the statistic on which the test is based is completely independent of the parent distribution of the variable. The two terms are imperfect synonyms and tend to be blurred frequently. Therefore, many statisticians tend to use them interchangeably.[12]

The research design chosen for this study falls into the category of the nonparametric, distribution-free statistical test model. It is Cochran's Q Test.

Cochran's Q test is an extention of the McNemar two-sample test and is considered appropriate in an

---

[12]Leonard A. Marascuilo and Maryellen McSweeney, Nonparametric and Distribution-Free Methods for the Social Sciences, Monterey, California: Brooks/Cole Publishing Company, p. 5.

experiment involving repeated observations or matched groups where the dependent variable can take only two values: 1) $X_{ik} = 1$ if the observation for the subject "i" under condition "k" can be termed a "success;" or 2) $X_{ik} = 0$ if the observation for the subject "i" under condition "k" is a "failure". The term success is arbitrarily applied to the outcome of interest. The role of the numerical score is to assign individuals to one of two categories.[13, 14]

Cochran's Q test has a distribution that is approximately χ^2 with v = K-1 degrees of freedom. The statistic for the test is

$$Q = \frac{K(K-1)\ \Sigma\ C_j^2 - (K-1)N}{KN - \Sigma R_i^2} \sim \chi^2_{K-1} \qquad 15$$

where

C_j = the sum of the column values

R_i = the sum of the row values

K = the number of rows or subjects

N = either the sum of the columns or the sum of the rows as they are equal values.

---

[13]William L. Hays, Statistics for the Social Sciences, New York: Holt, Rinehart and Winston, Inc., 1973, pp. 773, 775.

[14]Marascuilo and McSweeney, Nonparametric and Distribution-Free Methods for the Social Sciences, p. 177.

[15]Ibid., p. 178.

A test of the hypothesis that the proportions of success are the same for all treatments, or that treatment effects are absent, can be made by rejecting H_o if:

$$Q > \chi^2_{K-1:1-\alpha} \quad [16]$$

If H_o is rejected on the basis of the hypothesis test, it is not possible to determine the magnitude or the direction of the difference in treatments. Post hoc multiple comparions of the treatment means can be used to examine the differences among treatments more carefully. Multiple comparisons of the treatment means can be conducted through the use of the Dunn-Bonferroni inequality test. The use of the Dunn-Bonferroni test provides a narrower confidence interval than the Scheffe' technique.[17]

The research design was applied to the study under investigation in that the reading instructional programs were considered the treatments and the reading concepts were considered the subjects. If an instructional program presented a given concept, the value of "1" was assigned. Presentation of a concept was equated with "success". The lack of a program's presentation of a given concept was considered a "failure" and the value of "0" was assigned. To be considered a "success", the concept

---

[16] Ibid.

[17] Ibid., p. 180.

had to have been presented in any of the grades K-3 to be compared with the Experimental Reading Test Grade 4 or in any of the grades 4-6 to be compared with the Experimental Reading Test Grade 7. A "failure" was the total absence of the presentation of a concept by an instructional program in either of the appropriate levels K-3 or 4-6. A "success" was the presentation of a concept by an instructional program at any grade level within the appropriate levels of K-3 or 4-6 to be compared with the appropriate level of the Experimental Reading Test.

The Cochran Q test was used to obtain inter-rater reliability scores between the independent rating of the judges. The reliability was computed from the proportions of the individuals' ratings of which concepts the test items measured.

Statistical Methodology

Statistical treatments of the data in this study were conducted through the use of the facilities of the Computer Laboratory, Michigan State University. The statistical package for the Social Sciences (SPSS) routines were used to compute the proportions data. The calculations of the computer were randomly checked by performing the statistical treatments on a mechanical calculator.

The Dunn-Bonefrroni pairwise comparisons were conducted using a mechanical calculator to perform the statistical treatments to examine the differences between proportions

scores of the instructional materials and the Experimental Reading Test.

Summary

The Reading Concepts Checklist, (RCC), was developed as a means of describing, within a common framework, the concepts presented in the instructional materials and the concepts tested in the Michigan Educational Assessment Program Experimental Reading Test. The Reading Concepts Checklist, (RCC), was developed on the basis of conceptual consensus of agreement obtained from the work of several recognized authorities in the field of reading. It was formed into two matrices for the purpose of classifying the instructional materials' presented concepts and the Experimental Reading Tests' tested concepts.

The data were coded for IBM tabulation. Statistical treatments required for tests of inter-rater reliability and the significance of the difference between the proportions were processed through the use of the facilities of the Computer Laboratory, Michigan State University. The Cochran Q test was used to compute the significance of the difference between proportions. The Cochran Q test was used to obtain inter-rater reliability scores to determine the significance of difference between judges. The Dunn-Bonferroni pairwise comparisons were performed to examine the differences in significance of the proportions.

CHAPTER IV

ANALYSIS OF RELATIONSHIPS BETWEEN VARIOUS READING PROGRAMS AND THE MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM EXPERIMENTAL READING TEST

This chapter contains a restatement of the major hypotheses tested, a summary of the findings, a description and interpretation of the statistical treatment of the data, and an evaluation of each hypothesis.

The hypotheses which are being tested are stated in the null form and are designated by the symbol H_o. The level of significance used is .05. If the probability of the occurrence of the data is smaller than the level of significance, the data are considered contradictory to the hypothesis and a decision is made to reject the null hypothesis. Rejection of the null hypothesis is regarded as a decision to accept the research hypothesis. A non-rejection of the null hypothesis indicates there is no statistical difference and signifies a rejection of the corresponding research hypothesis.

This chapter contains an analysis of the degree of concurrence between the five reading instructional programs surveyed and the relationship between each of the five reading programs in the Michigan Educational Assessment

Program Experimental Reading Test, Grades Four and Seven, as measured by the Reading Concepts Checklist, (RCC).

Analysis

General Hypothesis I

The first general hypothesis and fifteen operational null hypotheses are as follows:

There will be no difference between the five reading instructional programs in grades K-3 in the concepts they present or between the degree of concurrence between the concepts presented in each of the five reading instructional programs in grades K-3 and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 as shown in the Reading Concepts Checklist, (RCC).

Operational H1a:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the K-3 reading instructional program published by Harcourt, Brace, and Jovanovich according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1b:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the K-3 reading instructional program published by Holt, Rinehart, and Winston according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1c:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the K-3 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1d:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the K-3 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1e:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Harcourt, Brace and Jovanovich and the K-3 reading instructional program published by Holt, Rinehart, and Winston according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1f:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Harcourt, Brace and Jovanovich and the K-3 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1g:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Harcourt, Brace and Jovanovich and the K-3 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1h:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Holt, Rinehart, and Winston and the K-3 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational Hli:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Holt, Rinehart and Winston and the K-3 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational Hlj:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Houghton-Mifflin Company and the K-3 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational Hlk:

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational Hll:

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Harcourt, Brace and Jovanovich and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational Hlm:

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Holt, Rinehart and Winston and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational Hln:

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Houghton-Mifflin Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational Hlo:

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Scott, Foresman Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Summary of Hypothesis I Results

1. The total proportion scores of the matches and mismatches across the Reading Concepts Checklist, (RCC), show a significant degree of mismatch between each K-3 reading instructional program and the Michigan Educational Assessment Program Experimental Reading Test Grade 4 (Table 1).

2. The total proportion scores of the matches and mismatches across the Reading Concepts Checklist, (RCC), show a significant degree of mismatch between the K-3 reading instructional programs (Table 1).

3. Pairwise comparisons using mean scores of the proportions of matches and mismatches across the Reading Concepts Checklist, (RCC), show a significant degree of

mismatch between each reading instructional program and the Michigan Educational Assessment Program Experimental Reading Test Grade 4 (Table 2).

4. Pairwise comparisons using mean scores of the proportions of matches and mismatches across the Reading Concepts Checklist, (RCC), show no statistical difference between Ginn and Company and 1) Harcourt, Brace and Jovanovich, 2) Holt, Rinehart, and Winston, and 3) Houghton-Mifflin Company; show no statistical difference between Harcourt, Brace and Jovanovich and Holt, Rinehart, and Winston; show no statistical difference between Holt, Rinehart, and Winston and Houghton-Mifflin Company; show no statistical difference between Houghton-Mifflin Company and Scott, Foresman Company at the K-3 reading instructional program level (Table 2).

5. Pairwise comparisons using mean scores of the proportions of matches and mismatches across the Reading Concepts Checklist, (RCC), show a significant degree of mismatch between Ginn and Company and Scott, Foresman Company; show a significant degree of mismatch between Harcourt, Brace and Jovanovich and 1) Houghton-Mifflin Company and 2) Scott, Foresman Company; show a significant degree of mismatch between Holt, Rinehart, and Winston and Scott, Foresman Company at the K-3 reading instructional program level (Table 2).

6. An analysis of the findings of this study indicates a strong lack of concurrence between each reading instructional program and the Michigan Educational Assessment Program Experimental Reading Test Grade 4. Differences are apparent between the reading instructional programs in the total category score but are less apparent when pairwise comparisons are performed.

7. The overall findings related to the degree of concurrence between the K-3 reading programs surveyed and each K-3 reading program and the Michigan Educational Assessment Program Experimental Reading Test Grade 4, as measured by the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC), indicate a lack of concurrence between the Michigan Educational Assessment Program Experimental Reading Test Grade 4, and each of the K-3 reading programs. The relationship of the Reading Concepts Checklist, (RCC), to the Michigan Educational Assessment Program Experimental Reading Test Grade 4, will be analyzed in detail following the results of Hypothesis II.

Statistical Tests and Treatments

The Cochran Q test, utilizing a Chi-square (χ^2) distribution, was used to test the significance of the observed differences between the proportion of matches

and mismatches across the Reading Concepts Checklist, (RCC). The limits within which the hypotheses will be accepted and outside of which they will be rejected are predicated on the .05 level of significance. The χ^2 values which cut off 2.5 percent of the area in each tail of the χ^2 distribution provide the measure of the difference between the proportion scores. The Ω statistic will be numerically larger than the χ^2 distribution when the null hypotheses are not true.

The null hypothesis will not be rejected if the χ^2 value is greater than the .05 level of significance ($p > .05$). The region of rejection for the null hypothesis is defined by the confidence limits, (.025, .975). When very strong rejections of the null hypotheses occur, higher probability levels for rejecting the null hypotheses are given, for example: $p < .01$ or $p < .001$.

Table 1. Summary of the total proportion scores of the matches and mismatches of the K-3 reading instructional programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 4 as measured by the 103 concepts contained in the Reading Concepts Checklist, (RCC).[1]

Program	Matches	Mismatches	Proportion
Ginn and Company	89	14	.8641
Harcourt, Brace Jovanovich	90	13	.8932
Holt, Rinehart and Winston	89	14	.8641
Houghton-Mifflin Company	72	31	.6990
Scott, Foresman Company	70	33	.6796
Test-Grade 4	25	78	.2427

Summaries of the results of the statistical treatments are presented in the following sections. Additional data are included in the appendices and referred to as necessary in the analysis of the results.

The determination of whether observed differences in the total proportion scores indicates the degree of concurrence is of major interest. Additional examination and analysis is concerned with the degree of concurrence

[1]See Appendices C and D for additional statistical data.

between the K-3 reading programs surveyed and each of the K-3 reading instructional programs and the Michigan Educational Assessment Program Experimental Reading Test, Grade 4.

Table 2. Interval estimate of the multiple comparison of proportion scores for the K-3 reading programs and the Exterimental Reading Test, Grade 4.

	1	2	3	4	5	T_4	C.I.[a]
1		-.0291	0	.1651	.1845	.6214	±.1764
2			.0291	.1942	.2136	.6505	
3				.1651	.1845	.6214	
4					.0194	.4563	
5						.4368	

Key: 1 = Ginn and Company

2 = Harcourt, Brace and Jovanovich

3 = Holt, Rinehart, and Winston

4 = Houghton-Mifflin Company

5 = Scott, Foresman Company

T_4 = Michigan Educational Assessment Program Experimental Reading Test Grade 4

[a]Confidence Interval

Results and Evaluation of Statistical Treatment

Total Proportion Scores

In order to determine the degree of concurrence between the K-3 reading programs surveyed and between each of the K-3 reading instructional programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 4, the total proportion scores which appear in Table 3 between each of the K-3 reading programs and the Experimental Test Grade 4, were compared by means of the Cochran Q test. Based on the significant difference in total proportion scores, the null hypothesis:

> There will be no difference between the five reading instructional programs in grades K-3 in the concepts they present or between the degree of concurrence between the concepts presented in each of the five reading instructional programs and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 as shown in the Reading Concepts Checklist, (RCC).

is rejected; therefore, the research hypothesis that there is a significant statistical difference between the K-3 reading instructional programs and between each of the K-3 reading programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 4, as shown in the Reading Concepts Checklist, (RCC), is accepted.

This difference indicates a significant lack of concurrence between each of the K-3 reading instructional programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 4, and a lack of concurrence

between the K-3 instructional programs. The difference does not indicate the magnitude nor the direction of the difference.

Table 3. Differences in total proportion scores of the K-3 instructional programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 4.[2]

Score Total Matches	1	2	3	4	5	T_4	Q	D.F.	P
Programs and Test	89	90	89	72	70	25	162.4435	5	p < .001 - S
Programs Only	89	90	89	72	70		31.9865	4	p < .001 - S

Key: S indicates a level of significance between proportion scores at a minimum of P< .05. P <.001 represents higher levels of significance than minimum required.

1 = Ginn and Company

2 = Harcourt, Brace and Jovanovich

3 = Holt, Rinehart, and Winston

4 = Houghton-Mifflin Company

5 = Scott, Foresman Company

[2]See Appendix D for additional statistical data.

Pairwise Comparison Scores

Table 4 contains the values of the pairwise comparison of the means of the proportion scores between Ginn and Company and 1) Harcourt, Brace and Jovanovich, 2) Holt, Rinehart, and Winston, 3) Houghton-Mifflin Company, 4) Scott, Foresman Company K-3 reading instructional programs, and 5) the Michigan Educational Assessment Program Experimental Reading Test Grade 4. On the basis of the lack of a significant statistical difference between the means of the proportions scores, the following null hypotheses are accepted.

Operational H1a:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the K-3 reading Instructional program published by Harcourt, Brace and Jovanovich according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1b:

There will be no difference between the concepts presented in the K-3 reading instructional progarm published by Ginn and Company and the K-3 reading instructional program published by Holt, Rinehart, and Winston according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1c:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the K-3 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

The corresponding research hypotheses that a significant statistical difference exists are rejected.

A significant statistical difference between the means of the proportion scores is evident and the following null hypotheses are rejected:

Operational Hld:

> There will be no difference between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the K-3 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational Hlk:

> There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

The corresponding research hypotheses, then, are accepted.

The values in Table 5 of the pairwise comparison of the means of the proportion scores between Harcourt, Brace and Jovanovich and 1) Holt, Rinehart, and Winston, 2) Houghton-Mifflin Company, 3) Scott, Foresman Company, and 4) the Michigan Educational Assessment Program Experimental Reading Test Grade 4, yield a non-significant statistical difference between the means of the proportion scores. Thus, the following null hypothesis is accepted.

Operational Hle:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Harcourt, Brace and Jovanovich and the K-3 reading instructional program published by Holt, Rinehart, and Winston according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

and the corresponding research hypothesis that significant statistical difference exists is rejected.

Table 4. Interval estimate of the multiple comparison of proportion scores for the K-3 reading programs and the Experimental Reading Test Grade 4.

Publisher	Ginn and Company		C.I.
Harcourt, Brace and Jovanovich	-.0291	NS	±.1764
Holt, Rinehart, and Winston	0	NS	
Houghton-Mifflin Company	.1651	NS	
Scott, Foresman Company	.1845	S	
Experimental Reading Test, Grade 4	.6214	S	

NS indicates a non-significant statistical difference between the means of the proportion scores.

S indicates statistically significant difference between the means of the proportion scores at a minimum of $p < .05$.

The occurrence of a significant statistical difference in the means of the proportion scores forms the basis for rejecting the following null hypotheses:

Operational H1f:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Harcourt, Brace and Jovanovich and the K-3 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1g:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Harcourt, Brace and Jovanovich and the K-3 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1l:

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Harcourt, Brace and Jovanovich and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

and, conversely, the basis for accepting the corresponding research hypotheses that significant statistical differences do exist.

The pairwise comparison values in Table 6 of the means of the proportion scores between Holt, Rinehart, and Winston and Houghton-Mifflin Company fail to illustrate a significant statistical difference. Therefore, the following null hypothesis is accepted and its research hypothesis stating the existence of a significant statistical difference is rejected.

Operational Hlh:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Holt, Rinehart and Winston and the K-3 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Table 5. Interval estimate of the pairwise comparion of proportion scores between Harcourt, Brace and Jovanovich and three K-3 reading programs and the Experimental Reading Test Grade 4.

Publisher	Harcourt Brace and Jovanovich		C.I.
Holt, Rinehart, and Winston	.0291	NS	±.1764
Houghton-Mifflin Company	.1942	S	
Scott, Foresman Company	.2136	S	
Experimental Reading Test Grade 4	.6505	S	

NS indicates a non-significant statistical difference between the means of the proportion scores.

S indicates statistically significant difference between the means of the proportion scores at the minimum of $p < .05$.

However, the means of the proportion scores in Table 6 exhibit a significant statistical difference between Holt, Rinehart, and Winston and 1) Scott, Foresman Company and 2) the Michigan Educational Assessment Program Experimental Reading Test Grade 4. As a result of these significant

differences, the research hypotheses that significant statistical differences exist are accepted and the following null hypotheses are rejected:

Operational Hli:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Holt, Rinehart, and Winston and the K-3 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational Hlm:

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Holt, Rinehart, and Winston and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Table 6. Interval estimate of the pairwise comparison of proportion scores between Holt, Rinehart, and Winston and the two K-3 reading programs and the Experimental Reading Test Grade 4.

Publisher	Holt, Rinehart, and Winston		C.I.
Houghton-Mifflin Company	.1651	NS	±.1764
Scott, Foresman Company	.1845	S	
Experimental Reading Test Grade 4	.6214	S	

NS indicates a non-significant statistical difference between the means of the proportion scores.

S indicates statistically significant difference between the means of the proportion scores at a minimum of $p < .05$.

Table 7 presents the results of the pairwise comparison of the means of the proportion scores between Houghton-Mifflin Company and Scott, Foresman Company K-3 reading programs. The values reveal the lack of a significant statistical difference. Based on the results of the comparison score, the following hypothesis is accepted:

Operational H1j:

> There will be no difference between the concepts presented in the K-3 reading instructional program published by Houghton-Mifflin Company and the K-3 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Because the above hypothesis is accepted, the corresponding research hypothesis advocating the existance of a significant statistical difference is rejected.

The relationship between the K-3 reading program published by Houghton-Mifflin Company and the Michigan Educational Assessment Program Experimental Reading Test Grade 4, is also presented in Table 7 in the form of the means of the proportion scores. The values of the means of the proportion scores indicate a significant statistical difference exists. Therefore, the research hypothesis declaring the existence of a significant statistical difference is accepted and the following null hypothesis is rejected:

Operational Hln:

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Houghton-Mifflin Company and the concepts Tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4, according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Table 7. Interval estimate of the pairwise comparison of proportion scores between Houghton-Mifflin Company and Scott, Foresman Company K-3 reading programs and the Experimental Reading Test Grade 4.

Publisher	Houghton-Mifflin Company		C.I.
Scott, Foresman Company	.0194	NS	±.1764
Experimental Reading Test Grade 4	.4563	S	

NS indicates a non-significant statistical difference between the means of the proportion scores.

S indicates statistically significant difference between the means of the proportion scores at a minimum of $p < .05$.

On the basis of a significant statistical difference, Table 8, between Scott, Foresman Company K-3 reading program and the Michigan Educational Assessment Program Experimental Reading Test Grade 4, the null hypothesis:

Operational Hlo:

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Scott, Foresman Company and the concepts tested by the Michigan Educational Assessment Program experimental Reading Test Grade 4, according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

is rejected and the research hypothesis that a significant statistical difference exists is accepted.

Table 8. Interval estimate of the pairwise comparison of proportion scores between the K-3 reading program published by Scott, Foresman Company and the Experimental Reading Test Grade 4.

Publisher	Scott, Foresman Company		C.I.
Experimental Reading Test Grade 4	.4369	S	±.1764

S indicates statistically significant difference between the means of the proportion scores at a minimum of $p < .05$.

Table 9 contains a summary of the values of the pairwise comparisons of the means of the proportion scores between the K-3 reading instructional programs and between each of the K-3 reading programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 4. The table contains information indicating the level of significance regarding whether or not the value is statistically significant, the variance between the proportion

scores, and the "psi" value which indicates the confidence limits beyond which rejection of the null hypothesis occurs.

Table 9. Summary of the interval estimate of the pairwise comparisons of the means of the proportion scores between the K-3 reading programs and each of the K-3 reading programs and the Experimental Reading Test Grade 4.

	1	2	3	4	5	T_4	C.I.
1		-.0291[a]	0[a]	.1651[a]	.1845	.6214	±.1764
2			.0291[a]	.1942	.2136	.6505	
3				.1651[a]	.1845	.6214	
4					.0194[a]	.4563	
5						.4369	

Key: 1 = Ginn and Company

2 = Harcourt, Brace and Jovanovich

3 = Holt, Rinehart, and Winston

4 = Houghton-Mifflin Company

5 = Scott, Foresman Company

T_4 = Michigan Educational Assessment Program Experimental Reading Test Grade 4

Var. = .0036
$p < .05$

[a]Non-significant Statistical Difference

1. The data contained in Table 9 clearly support the research hypotheses that significant statistical difference exists between each of the K-3 reading programs and the Michigan Educational Assessment Program Experimental Reading

test Grade 4, according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).[3]

2. The data contained in Table 9 indicate a non-significant statistical difference exits between the K-3 reading programs published by Ginn and Company and Harcourt, Brace and Jovanovich; Ginn and Company and Holt, Rinehart, and Winston; Holt, Rinehart, and Winston and Houghton-Mifflin Company; and Houghton-Mifflin Company and Scott, Foresman Company. Therefore, the null hypotheses are accepted and the corresponding research hypotheses that such a statistical difference exists are rejected.

3. The data contained in Table 9 support the research hypotheses that significant statistical difference exists between the K-3 reading program published by Ginn and Company and Scott, Foresman Company; Harcourt, Brace and Jovanovich and Houghton-Mifflin Company; Harcourt, Brace and Jovanovich and Scott, Foresman Company; and Holt, Rinehart, and Winston and Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC); therefore, the null hypotheses that there will be no differences between the concepts presented in the K-3 reading instructional programs are rejected.

---

[3]See Appendices D and E for additional statistical data.

The data which have been analyzed have been concerned with the proportion of matches and mismatches between the K-3 reading programs surveyed and between each of the K-3 reading programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 4. The proportion scores have involved the total proportion scores based on the 103 concepts contained in the Reading Concepts Checklist, (RCC). From the data contained in Table 9, additional analysis of data which was statistically non-significant was deemed unnecessary. Additional analysis of the statistically significant data was conducted. The additional analysis was conducted to determine the areas in which the K-3 instructional programs differed from each other and the Grade 4 test. To determine the areas of difference, the data contained in the Reading Concepts Checklist, (RCC), were analyzed according to the major categories.

The data presented in Table 10 add additional support that the null hypotheses:

Operational Hld:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the K-3 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1k:

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Ginn and Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1f:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Harcourt, Brace and Jovanovich and the K-3 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1g:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Harcourt, Brace and Jovanovich and the K-3 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1l:

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Harcourt, Brace and Jovanovich and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H1i:

There will be no difference between the concepts presented in the K-3 reading instructional program published by Houghton-Mifflin Company and the K-3 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational Hlm:

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Holt, Rinehart, and Winston and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational Hln:

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Houghton-Mifflin Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational Hlo:

There will be no difference in the degree of concurrence between the concepts presented in the K-3 reading instructional program published by Scott, Foresman Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

are rejected and the research hypotheses that significant statistical difference exists between the K-3 reading instructional programs and each of K-3 reading programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 4, are accepted.

Table 10 Interval estimate of the multiple comparison of proportion scores for the K-3 reading programs and the Experimental Reading Test Grade 4 by individual categories in the Reading Concepts Checklist, (RCC).[4]

	1	2	3	4	5	T_4	C.I.
			Category: Vocabulary Development				
1		0[a]	0[a]	.1667	.6667	.3334	±.0882
2			0[a]	.1667	.6667	.3334	
3				.1667	.6667	.3334	
4					.50	.1667	
5						-.3333	
			Category: Inferential Comprehension				
1		-.0583[a]	.0583[a]	.2353	.3530	.4118	±.1017
2			0[a]	.2941	.4118	.4707	
3				.2941	.4118	.4706	
4					.1177	.1765	
5						.0588[a]	
			Category: Study Skills				
1		.1818	.0909[a]	.3637	.2727	.5455	±.1211
2			-.0909[a]	.1891	.0909[a]	.3637	
3				.2728	.1818	.4546	
4					-.0910[a]	.1818	
5						.2728	

Continued

[4]See Appendices D and E for additional statistical data.

The null hypotheses are rejected at the 0.05 level. Higher levels are indicated.

[a]Non-significant Statistical Difference.

Key: 1 = Ginn and Company

2 = Harcourt, Brace and Jovanovich

3 = Holt, Rinehart, and Winston

4 = Houghton-Mifflin Company

5 = Scott, Foresman Company

T_4 = Michigan Educational Assessment Program Experimental Reading Test Grade 4

Analysis

General Hypothesis II

The second general hypothesis and fifteen operational null hypotheses are as follows:

There will be no difference between the five reading instructional programs in grades 4-6 in the concepts they present or between the degree of concurrence between the concepts presented in each of the five reading instructional programs in grades 4-6 and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 as shown in the Reading Concepts Checklist, (RCC).

Opeational H2a:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2b:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the 4-6 reading instructional program published by Holt, Rinehart, and Winston according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2c:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the 4-6 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2d:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the 4-6 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2e:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich and the 4-6 reading instructional program published by Holt, Rinehart, and Winston according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2f:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich and the 4-6 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2g:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich and the 4-6 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2h:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Holt, Rinehart, and Winston and the 4-6 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2i:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Holt, Rinehart, and Winston and the 4-6 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2j:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Houghton-Mifflin Company and the 4-6 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2k:

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2l:

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2m:

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Holt, Rinehart, and Winston and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2n:

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Houghton-Mifflin Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2o:

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Scott, Foresman Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Summary of Hypothesis II Results

1. The total proportion scores of the matches and mismatches across the Reading Concepts Checklist, (RCC), show a significant degree of mismatch between each of the 4-6 reading programs surveyed and the Michigan Educational Assessment Program Experimental Reading Test Grade 7, (Table 11).

2. The total proportion scores of the matches and mismatches across the Reading Concepts Checklist, (RCC), show

a significant degree of mismatch between the 4-6 reading programs (Table 11).

3. Pairwise comparisons, using mean scores of the proportions of matches and mismatches across the Reading Concepts Checklist, (RCC), show a significant degree of mismatch between each of the reading programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 7, (Table 12).

4. Pairwise comparisons, using mean scores of the proportions of matches and mismatches across the Reading Concepts Checklist, (RCC), show no statistical difference between Ginn and Company and Holt, Rinehart, and Winston; Harcourt, Brace and Jovanovich and Holt, Rinehart and Winston; Houghton-Mifflin Company and Scott, Foresman Company (Table 12).

5. Pairwise comparisons, using mean scores of the proportions of matches and mismatches across the Reading Concepts Checklist, (RCC), show a significant degree of mismatch between Ginn and Company and 1) Houghton-Mifflin Company and 2) Scott, Foresman Company; show a significant degree of mismatch between Harcourt, Brace and Jovanovich and 1) Houghton-Mifflin Company and 2) Scott, Foresman Company; show a significant degree of mismatch between Holt, Rinehart, and Winston and 1) Houghton-Mifflin Company and 2) Scott, Foresman Company (Table 12).

6. An analysis of the findings of this study indicates a strong lack of concurrence between each of the reading programs surveyed and the Michigan Educational Assessment Program Experimental Reading Test Grade 7. Differences are apparent between the reading programs in the total category score but are less apparent when pairwise comparions are performed.

7. The overall findings related to the degree of concurrence between the 4-6 reading instructional programs surveyed and each of the 4-6 reading programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 7, as measured by the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC), indicate the lack of concurrence between the Michigan Educational Assessment Program Experimental Reading Test Grade 7, and each of the 4-6 reading programs. The relationship of the Reading Concepts Checklist, (RCC), to the Michigan Educational Assessment Program Experimental Reading Test Grade 7, will be analyzed in detail following the results of Hypothesis II.

Table 11. Summary of the total proportion scores of the matches and mismatches of the 4-6 reading instructional programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 7, as measured by the 103 concepts contained in the Reading Concepts Checklist, (RCC).[5]

Program	Matches	Mismatches	Proportion
Ginn and Company	85	18	.8252
Harcourt, Brace and Jovanovich	85	18	.8252
Holt, Rinehart, and Winston	87	16	.8447
Houghton-Mifflin Company	53	50	.5147
Scott, Foresman Company	67	36	.6505
Test-Grade 7	28	75	.2718

[5]See Appendices F and G for additional statistical data.

Table 12. Interval estimate of the multiple comparison of proportion scores for the 4-6 reading programs and the Experimental Reading Test Grade 7.

	1	2	3	4	5	T_7	C.I.
1		0	-.0195	.3105	.1747	.5534	±.1714
2			-.0195	.3105	.1747	.5534	
3				.3300	.1942	.5729	
4					-.1358	.2429	
5						.3787	

Key: 1 = Ginn and Company

2 = Harcourt, Brace and Jovanovich

3 = Holt, Rinehart, and Winston

4 = Houghton-Mifflin Company

5 = Scott, Foresman Company

T_7 = Michigan Educational Assessment Program Experimental Reading Test Grade 7.

Statistical Test and Treatment

The Cochram Q test, utilizing a Chi-square distribution, was used to test the significance of the observed difference between the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC). The level of significance to determine whether the null hypotheses were rejected or not

rejected was the .05 level. The null hypotheses will be accepted if the Chi-square value is greater than the .05 level of signficance ($p > .05$), indicating concurrence between the 4-6 reading programs surveyed and each of the reading programs and the Michigan Experimental Reading Test Grade 7. The Q statistic will be numerically larger than the Chi-square distribution when the null hypotheses are not true, indicating a lack of concurrence between the 4-6 reading programs and each of the reading programs and the Michigan Experimental Reading Test Grade 7. The full tests and techniques described and used in analyzing Hypothesis I are used to analyze Hypothesis II.

Results and Evaluation of Statistical Treatment

Total Proportion Scores

In order to assess the degree of concurrence between the 4-6 reading instructional programs surveyed and each of the reading programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 7, the total proportion scores of the 4-6 reading programs and the Experimental Test Grade 7, were compared by means of the Cochran Q test.

Based on the significant difference in total proportion scores, Table 13, the null hypothesis:

There will be no difference between the five reading instructional programs in grades 4-6 in the concepts they present or between the degree of concurrence between the concepts presented in each of the five reading instructional programs in grades 4-6 and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7, as shown by the Reading Concepts Checklist, (RCC).

is rejected; therefore, the research hypothesis that there is a significant statistical difference between the 4-6 reading instructional programs surveyed and each of the reading programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 7, as shown in the Reading Concepts Checklist, (RCC), is accepted. This difference indicates a significant lack of concurrence between the 4-6 reading programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 7, and a lack of concurrence between the 4-6 reading instructional programs. The difference is not indicative of the magnitude nor the direction of the difference.

Pairwise Comparison Scores

The magnitude and the direction of the difference in total proportion scores between the 4-6 reading programs and each of the 4-6 reading program and the Michigan Educational Assessment Program Experimental Reading Test Grade 7, was determined through the use of the Dunn-Bonferroni pairwise comparisons technique.

Table 13. Differences in total proportion scores of the 4-6 reading programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 7.[6]

Score Total Matches	1	2	3	4	5	T_7	Q	D.F.	P
Programs and Test	85	85	87	53	67	28	153.224	5	p < .001S
Programs Only	85	85	87	53	67		64.579	4	p < .001S

P < .001 represents higher level of significance than minimum.

S indicates a level of significance between proportion scores at a minimum of p < .05.

Key: 1 = Ginn and Company

2 = Harcourt, Brace and Jovanovich

3 = Holt, Rinehart, and Winston

4 = Houghton-Mifflin Company

5 = Scott, Foresman and Company

T_7 = Michigan Educational Assessment Program Experimental Reading Test Grade 7.

Table 14 contains the values of the pairwise comparison of the means of the proportion scores between Ginn and Company and 1) Harcourt, Brace and Jovanovich, 2) Holt, Rinehart, and Winston, 3) Houghton-Mifflin Company,

[6]See Appendices F and G for additional statistical data.

4) Scott, Foresman Company, and 5) the Michigan Educational Assessment Program Experimental Reading Test Grade 7.

The lack of a significant statistical difference between the means of the proportion scores results in the following null hypotheses being accepted:

Operational H2a:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2b:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the 4-6 reading instructional program published by Holt, Rinehart, and Winston according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

and the corresponding research hypotheses that a significant statistical difference exists are rejected.

However, the significant statistical difference between the means of the proportion scores for the null hypotheses:

Operational H2c:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by the Ginn and Company and the 4-6 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC)

Operational H2d:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the 4-6 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2k:

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7, according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC),

are rejected and the corresponding research hypotheses stating a difference exists between the concepts presented by the reading program published by Ginn and Company and 1) Houghton-Mifflin Company and 2) Scott, Foresman Company and a difference exists between the concepts presented by the Ginn and Company 4-6 reading program and the concepts tested by the Michigan Experimental Reading Test Grade 7, are accepted.

The pairwise comparison of the means of the proportion scores, Table 15, between Harcourt, Brace and Jovanovich and Holt, Rinehart, and Winston 4-6 reading programs failed to indicate a significant statistical difference. The non-sigifnicant statistical difference indicates the value is within the confidence interval. Therefore, the research hypothesis stipulating a significant statistical difference exists is rejected and the following null hypothesis is accepted:

Operational H2e:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich and the 4-6 reading instructional program published by Holt, Rinehart, and Winston according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Table 14. Interval estimate of pairwise comparison of proportion scores between Ginn and Company and four 4-6 reading programs and the Experimental Reading Test Grade 7.

Publisher	Ginn and Company		C.I.
Harcourt. Brace and Jovanovich	0	NS	±.1714
Holt, Rinehart, and Winston	-.0195	NS	
Houghton-Mifflin Company	.3105	S	
Scott, Foresman Company	.1747	S	
Experimental Reading Test Grade 7	.5534	S	

NS indicates a non-significant statistical difference

S indicates statistically significant difference between the means of the proportion scores at a minimum of $p < .05$.

However, differences in the means of the proportion scores, Table 15, between the 4-6 reading programs published by Harcourt, Brace and Jovanovich and 1) Houghton-Mifflin Company and 2) Scott, Foresman Company exceeded the level of

probability. Furthermore, the differences in the means of the proportions between the concepts presented in the 4-6 reading program published by Harcourt, Brace and Jovanovich and the concepts tested by the Michgian Educational Assessment Program Experimental Reading Test Grade 7, are statistically significant and justify rejecting the following null hypotheses:

Operational H2f:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich and the 4-6 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2g:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich and the 4-6 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H21:

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program program published by Harcourt, Brace and Jovanovich and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7, according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Therefore, the corresponding research hypotheses declaring the existence of significant statistical differences are accepted.

The pairwise comparison values, Table 16, of the means of the proportions scores between the 4-6 reading programs

of Holt, Rinehart, and Winston and 1) Houghton-Mifflin Company and 2) Scott, Foresman Company and 3) the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7, illustrate a significant statistical difference. Therefore, the following null

Table 15. Interval estimate of the pairwise comparison of proportion scores between Harcourt, Brace and Jovanovich and three 4-6 reading programs and the Experimental Reading Test Grade 7.

Publishers	Harcourt, Brace and Jovanovich		C.I.
Holt, Rinehart and Winston	-.0195	NS	±.1714
Houghton-Mifflin Company	.3105	S	
Scott, Foresman Company	.1747	S	
Experimental Reading Test Grade 7	.5534	S	

NS indicates a non-significant statistical difference.

S indicates a statistically significant difference between the means of the proportion scores at a minimum of $p < .05$.

hypotheses are rejected and their research hypotheses claiming a statistical difference exists are accepted:

Operational H2h:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Holt, Rinehart, and Winston and the 4-6 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2i:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Holt, Rinehart, and Winston and the 4-6 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2m:

There will be no difference in the degree of concurrence between the concepts presented in 4-6 reading instructional program published by Holt, Rinehart, and Winston and the concepts tested by the Michgian Educational Assessment Program Experimental Reading Test Grade 7, according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Table 16. Interval estimate of the pairwise comparison of proportion scores between Holt, Rinehart, and Winston and two 4-6 reading programs and the Experimental Reading Test Grade 7.

Publisher	Holt, Rinehart, and Winston		C.I.
Houghton-Mifflin Company	.3300	S	±.1714
Scott, Foresman Company	.1942	S	
Experimental Reading Test Grade 7	.5729	S	

S indicates statistically significant difference between the means of the proportion scores at a minimum of $p < .05$.

Table 17 presents the results of the pairwise comparison of the means of the proportion scores between Houghton-Mifflin Company and Scott, Foresman Company 4-6 reading

programs. The values reveal the lack of a significant statistical difference. Based on the results of the comparison score, the following hypothesis is accepted:

Operational H2j:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Houghton-Mifflin Company and the 4-6 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Because the above hypothesis is accepted, the corresponding research hypothesis advocating the existence of a significant statistical difference is rejected.

The relationship between the 4-6 reading program published by Houghton-Mifflin Company and the Michigan Educational Assessment Program Experimental Reading Test Grade 7, is also presented in Table 17 in the form of the means of the proportion scores. The values of the means of the proportion scores indicate a significant statistical difference exists. Therefore, the research hypothesis declaring the existence of a significant statistical difference is accepted and the following null hypothesis is rejected:

Operational H2n:

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Houghton-Mifflin Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7, according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Table 17. Interval estimate of the pairwise comparions of proportion scores between Houghton-Mifflin Company and Scott, Foresman Company 4-6 reading programs and the Experimental Reading Test Grade 7.

Publisher	Houghton-Mifflin Company		C.I.
Scott, Foresman Company	-.1358	NS	±.1714
Experimental Reading Test Grade 7	.2429	S	

NS indicates a non-significant statistical difference.

S indicates statistically significant difference between the means of the proportion scores at a minimum of $p < .05$.

A significant statistical difference between the pairwise comparison of the means of the proportion scores shown in Table 18 negates the following null hypothesis:

Operational H2o:

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Scott, Foresman Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7, according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC),

and justifies accepting the corresponding research hypothesis which states a difference exists in the degree of concurrence between the concepts presented in the Scott, Foresman Company 4-6 reading program and the concepts tested by the

Michigan Educational Assessment Program Experimental Reading Test Grade 7.

Table 18. Interval estimate of the pairwise comparison of proportion scores between the 4-6 reading program published by Scott, Foresman Company and the Experimental Reading Test Grade 7.

Publishers	Scott, Foresman Company		C.I.
Experimental Reading Test Grade 7	.3787	S	±.1747

S indicates statistically significant difference between the mean of the proportion scores as a minimum of $p < .05$.

Table 19 contains a summary of the values of the pairwise comparisons of the means of the proportion scores between the 4-6 reading programs surveyed and each of the 4-6 reading programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 7. The table contains information indicating the significance level as to whether or not the value is statistically significant, the variance between proportion mean scores, and the "psi" value which indicates the confidence limits beyond which rejection of the null hypothesis occurs.

1. The data contained in Table 19 clearly support the research hypotheses that significant statistical difference exists between each of 4-6 reading programs and the Michigan

Educational Assessment Program Experimental Reading Test Grade 7, according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).[7]

Table 19. Summary of the interval estimate of the pairwise comparisons of the mean of the proportion scores between the 4-6 reading programs and each of the 4-6 reading programs and the Experimental Reading Test Grade 7.

	1	2	3	4	5	T_7	C.I.
1		0[a]	-.0195[a]	.3105	.1747	.5534	±.1714
2			-.0195[a]	.3105	.1747	.5534	
3				.3300	.1942	.5729	
4					-.1358[a]	.2429	
5						.3787	

Key: 1 = Ginn and Company

2 = Harcourt, Brace and Jovanovich

3 = Holt, Rinehart, and Winston

4 = Houghton-Mifflin Comapny

5 = Scott, Foresman Company

T_7 = Michigan Educational Assessment Program Experimental Reading Test Grade 7.

Var. = .0036
$p < .05$

[a]Non-significant Statistical Difference.

[7]See Appendices G and H for additional statistical data.

2. The data contained in Table 19 indicate a non-significant statistical difference exists between the 4-6 reading programs published by Ginn and Company and 1) Harcourt, Brace and Jovanovich, and 2) Holt, Rinehart, and Winston; indicate a non-significant statistical difference exists between Harcourt, Brace and Jovanovich and Holt, Rinehart, and Winston; indicate a non-signficanct statistical difference exists between Houghton-Mifflin Company and Scott, Foresman Company. Therefore, the null hypotheses indicating there would be no difference are accepted and the corresponding research hypotheses indicating a difference would exist are rejected.

3. The data contained in Table 19 support the research hypotheses that significant statistical difference exists between the 4-6 reading programs published by Ginn and Company and 1) Houghton-Mifflin Company and 2) Scott, Foresman Company; significant statistical difference exists between Harcourt, Brace and Jovanovich and 1) Houghton-Mifflin Company and 2) Scott, Foresman Company; significant statistical difference exists between Holt, Rinehart, and Winston and 1) Houghton-Mifflin Company and 2) Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concept Checklist, (RCC); therefore, the null hypotheses that there will be no difference between the concepts presented in the 4-6 reading instructional programs are rejected.

The data analyzed were concerned with the proportion of matches and mismatches between the 4-6 reading instructional programs surveyed and each of the 4-6 reading programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 7. The proportion scores have involved the total proportion scores based on the 103 concepts contained in the Reading Concepts Checklist, (RCC). From the data presented in Table 19, additional analysis of data which were statistically non-significant was deemed unnecessary. Additional analysis of the statistically significant data was conducted. The additional analysis was conducted to determine the areas in which the 4-6 reading programs differed from each other and the Grade 7 Experimental Reading Test. To determine the areas of difference, the data contained in the Reading Concepts Checklist, (RCC), were anlayzed according to the major categories.

The data presented in Table 20 add additional support that the null phyotheses:

Operational H2c:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the 4-6 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2d:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the 4-6 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concept Checklist, (RCC).

Operational H2k:

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Ginn and Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2f:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich and the 4-6 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2g:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich and the 4-6 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concept Checklist, (RCC).

Operational H2l:

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Harcourt, Brace and Jovanovich and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2h:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Holt, Rinehart, and Winston and the 4-6 reading instructional program published by Houghton-Mifflin Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2i:

There will be no difference between the concepts presented in the 4-6 reading instructional program published by Holt, Rinehart, and Winston and the 4-6 reading instructional program published by Scott, Foresman Company according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2m:

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Holt, Rinehart, and Winston and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 according to the Proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2n:

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Houghton-Mifflin Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

Operational H2o:

There will be no difference in the degree of concurrence between the concepts presented in the 4-6 reading instructional program published by Scott, Foresman Company and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

are rejected and the corresponding research hypotheses that significant statistical difference exists between the 4-6 reading instructional programs and each of the 4-6 reading programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 7, are accepted.

Table 20. Interval estimate of the multiple comparison of proportion scores for the 4-6 reading programs and the Experimental Reading Test Grade 7, by individual categories in the Reading Concepts Checklist, (RCC).[8]

1	2	3	4	5	T_7	C.I.
			Category: Phonic Analysis			
1	-.1250[a]	-.1875	.750	-.1250[a]	.8125	±.1441
2		.0625[a]	.8750	0.00[a]	.9375	
3			.9375	.0625[a]	1.00	
4				-.8750	.0625[a]	
5					.9375	
			Category: Structural Analysis			
1	0.00[a]	0.00[a]	.3636	.2727	.8182	±.1247
2		0.00[a]	.3636	.2727	.8182	
3			.3636	.2727	.8182	
4				-.0909[a]	.4546	
5					.5455	

[a]Non-Significant Statistical Difference. Continued

[8]See Appendices G and H for additional statistical data.

Table 20. Continued

	1	2	3	4	5	T_7	C.I.
			Category: Literal Comprehension				
1		.0769[a]	0.00[a]	.4616	.3077	.6154	±.1176
2			-.0769[a]	.3847	.2308	.5385	
3				.4616	.3077	.6154	
4					-.1539	.1538	
5						.3077	
			Category: Inferential Comprehension				
1		0.00[a]	-.0588[a]	.1765	.3530	.5294	±.0929
2			-.0588[a]	.1765	.3530	.5294	
3				.2353	.4118	.5882	
4					.1765	.3529	
5						.1764	

The null hypotheses are rejected at the 0.05 level. Higher levels are indicated.

[a]Non-Significant Statistical Difference

Key: 1 = Ginn and Company

2 = Harcourt, Brace and Jovanovich

3 = Holt, Rinehart, and Winston

4 = Houghton-Mifflin Company

5 = Scott, Foresman Company

T_7 = Michigan Educational Assessment Program Experimental Reading Test Grade 7.

INTER-RATER RELIABILITY
CLASSIFICATION OF TESTED CONCEPTS

The validity model upon which this study is based called for a review and an evaluation of the test by a panel of experts. The purpose of the review and evaluation by the experts was to determine the relationship of the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven to the Reading Concepts Checklist, (RCC). What concepts contained in the RCC were being measured by the Michigan Educational Assessment Program Experimental Reading Test? The establishment of this relationship provided the basis for the comparison of the Michigan Experimental Reading Test to the five reading programs.

The review and evaluation was conducted independently by a panel of three reading experts and the researcher. An inter-rater reliability study was performed to establish the strength of the relationship between the judges' classifications of the test items.

Summary of Inter-Rater Reliability Tests

1. The total proportion scores of the matches and mismatches across the Reading Concepts Checklist, (RCC), show a higher degree of agreement among the judges for the Grade 7 Test than the Grade 4 Test (Table 21).

2. The total proportion scores of the matches and mismatches across the Reading Concepts Checklist, (RCC), show a strong positive relationship among the judges' classification of the items of the Michigan Experimental Reading Test Grade 4 (Table 21).

3. The total proportion scores of the matches and mismatches across the Reading Concepts Checklist, (RCC), show a strong positive relationship among the judges' classification of the items of the Michigan Experimental Reading Test Grade 7 (Table 21).

4. The findings of the judges' rating indicate the fourth grade Michigan Educational Assessment Program Experimental Reading Test failed to measure any portion of the Reading Concepts Checklist, (RCC), subcategories of "Auditory Discrimination," "Visual Discrimination," and "Phonic Analysis," and the seventh grade test completely omitted measuring the subcategory of "Phonic Analysis."

5. An analysis of the findings of the inter-rater reliability study indicates a strong positive agreement among the judges. The non-significant statistical difference between the ratings of the judges eliminated the need for further analysis.

6. The overall findings related to the inter-rater reliability study indicate the judgments related to the concepts tested by the Michigan Educational Assessment

Program Experimental Reading Test Grades Four and Seven can be validity compared with the concepts presented by the five reading instructional programs according to the Reading Concepts Checklist, (RCC).

Statistical Tests and Treatments

The Cochran Q Test, compared to a Chi-square distribution, was used to test the significance of agreement among the judges between the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC). The limits within which the significance of agreement will be accepted and beyond which it will be unacceptable are based on the .05 level of significance. The Q statistic will be numerically large with the level of agreement is low. The level of inter-rater reliability will be accepted when the Chi-square value is greater than the .05 level of significance ($p > .05$). The region of rejection is defined by the confidence limits, (.025, .975).

Results and Evaluation of Statistical Treatment

In order to determine the relationship of the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven to the Reading Concepts Checklist, (RCC), the total proportion scores of the judges were compared by means of the Cochran Q Test (Table 21). Based

on the lack of a significant statistical difference in total proportion scores, it is accepted that there is strong positive agreement among the independent ratings of the judges and that their judgments may be compared to the five reading instructional programs according to the Reading Concepts Checklist, (RCC).

Table 21. Inter-rater reliability total proportion scores for the Experimental Test Grades 4 and 7.

	1	2	3	4	Q	D.F.	P
Grade 4							
Matches	28	26	25	24	2.8378	3	p > .05
Mismatches	75	77	78	79			
Grade 7							
Matches	27	29	29	27	1.3333	3	p > .05
Mismatches	76	74	74	76			

The findings of the test indicate the ratings of the judges show a greater proportion of the concepts contained in the Reading Concepts Checklist, (RCC), are not measured by the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven than the proportion of concepts which are measured by the Michigan Experimental Reading Test Grades Four and Seven.

CHAPTER V

SUMMARY, CONCLUSIONS, IMPLICATIONS AND RECOMMENDATIONS

This chapter contains a brief summary of the study's purpose, procedures, limitations, major findings and conclusions. Implications of the study and recommendations specifically associated with the data presented are also included.

Summary

Purpose and Major Hypotheses

This study is an attempt to establish the degree of concurrence between the concepts measured by the Michigan Educational Assessment Program Experimental Reading Test for grades four and seven and the concepts presented in the most widely used reading instructional programs used in Michigan. This study is designed to analyze and compare the concepts tested in the Michigan Educational Assessment Program Experimental Reading Test according to the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC). Also included in the purpose of this study

is the degree of concurrence between each of the reading instructional programs. Achieving the purpose of this study also requires a review and evaluation of the Michigan Educational Assessment Program Experimental Reading Test by a panel of reading experts.

Two major hypotheses were formulated concerning the degree of concurrence between the reading instructional programs and between each of the reading instructional programs and the Michigan Educational Assessment Program Experimental Reading Test. The major hypotheses are:

1. There will be no difference between the five reading instructional programs in grades K-3 in the concepts they present or between the degree of concurrence between the concepts presented in each of the five reading instructional programs in grades K-3 and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 4 as shown in the Reading Concepts Checklist, (RCC).

2. There will be no difference between the five reading instructional programs in grades 4-6 in the concepts they present or between the degree of concurrence between the concepts presented in each of the five reading instructional programs in grades 4-6 and the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grade 7 as shown in the Reading Concepts Checklist, (RCC).

Selection of Instructional Materials

A statistical analysis comparing the concepts presented by the reading instructional programs to the concepts tested by the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven, requires data from all levels of the reading instructional programs. The reading instructional programs used in this study provided 1) data from grades K-3 to be compared with the Michigan Educational Assessment Program Experimental Reading Test Grade 4; 2) data from grades 4-6 to be compared with the Michigan Educational Assessment Program Experimental Reading Test Grade 7; 3) reading concepts to which a majority of Michigan's K-6 students are exposed; and 4) assurance that the K-6 students using these programs represent a reasonable cross-section of Michigan's rural, suburban, urban, and large-city school children. The reading instructional programs selected for this study were chosen on the basis of a national survey of K-8 reading teachers and supervisors by an independent research organization.

Instrumentation and Data Collection

The Reading Concepts Checklist, (RCC), was developed as a means of describing, within a common framework, the concepts presented in the reading instructional materials and the concepts tested in the Michigan Educational Assessment

Program Experimental Reading Test. Its six major divisions, subdivided into nine major categories, contain 103 concepts. The Reading Concepts Checklist, (RCC), was developed on the basis of conceptual consensus of agreement to insure a high degree of meaning and similarity of meaning across reading specialists and test constructors. The Reading Concepts Checklist, (RCC), formed the basis of two matrices: 1) the classification of concepts presented in the reading instructional materials in kindergarten through grade six, and 2) the classification of the concepts tested in the Michigan Educational Assessment Program experimental Reading Test Grades Four and Seven.

The data from the reading instructional materials were collected through surveying the sixty-five teachers' manuals of the five reading instructional programs. Each concept presented in the manual by a specific program was recorded in the matrix for the classification of instructional materials in the cell connecting the appropriate grade level and Reading Concepts Checklist, (RCC), concept.

The data from the Michigan Educational Assessment Program Experimental Test Grades Four and Seven were collected through a review and evaluation of the test by a panel of reading experts. The panel matched the test items with their stated objectives, published by the Michigan Department of Education, and recorded the items in the Reading Concepts Checklist, (RCC), matrix for classification

of tested concepts in the cell connecting the appropriate grade level of the test and the Reading Concepts Checklist, (RCC), concept.

Concepts which were identified as being in either the reading instructional programs or the Michigan Experimental Reading Test were assigned the value of "1", while the missing concepts were assigned the value of "0".

Treatment of the Data and Analysis

Achievement of the objectives set forth in this study required the determination of the significance between the observed differences between the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC). The nonparametric Cochran Q Test, compared to a Chi-square distribution, was used to test the significance between the observed differences between the proportion of matches and mismatches across the Reading Concepts Checklist, (RCC).

The second statistical step was the determination of the magnitude and direction of the significance of the difference between the proportion scores. Multiple comparisons of the means of proportion scores were conducted through the use of the Dunn-Bonferroni pairwise comparison technique.

The Cochran Q Test was employed to determine the level of reliability and degree of inter-rater agreement of the panel of reading experts.

The data were scored and coded for IBM tabulation and processed on a high-speed computer. Statistical treatments of the data in this study were conducted through the use of the facilities of the Computer Laboratory, Michigan State University.

Scope and Delimitations of the Study

1. The study is delimited to the degree of concurrence between the concepts presented in the reading instructional programs and between each of the programs and the concepts measured by the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven as measured by the Reading Concepts Checklist, (RCC).

2. The study treats the concepts contained in the Reading Concepts Checklist, (RCC), as the defined content domain of the domain of reading concepts. The concepts are not intended to be inclusive.

3. The study treats the concepts presented in the selected reading instructional programs as those concepts to which a majority of Michigan K-6 students are exposed and are not interpreted as having been taught.

4. The conclusions and implications of this study regarding instructional programs are not interpreted to indicate the quality of the programs, merely their differences.

Major Findings

1. The Reading Concepts Checklist, (RCC), findings indicate that, according to pairwise comparison scores for four K-3 reading instructional programs (Ginn and Company; Harcourt, Brace, and Jovanovich; Holt, Rinehart, and Winston; Houghton-Mifflin Company), concurrence between each of the K-3 reading instructional programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 4, is lacking in a significant degree in all nine subcategories of the Reading Concepts Checklist, (RCC), (see Appendix E).

2. The Reading Concepts Checklist, (RCC), findings indicate concurrence between Scott, Foresman Company K-3 reading program and the Michigan Educational Assessment Program Experimental Reading Test Grade 4, is lacking in a significant degree in eight subcategories of the Reading Concepts Checklist, (RCC), while concurrence is present in a significant great degree in subcategory VII: Inferential Comprehension.

3. According to pairwise comparison scores, the Reading Concepts Checklist, (RCC), findings indicate that concurrence between the K-3 reading instructional programs is present in a significant degree between Ginn and Company and Harcourt, Brace and Jovanovich; between Ginn and Company and Holt, Rinehart, and Winston; between Holt, Rinehart, and

Winston and Houghton-Mifflin Company; and between Houghton-Mifflin Company and Scott, Foresman Company (see Table 9).

4. The findings, however, indicate a lack of concurrence in a significant degree between Ginn and Company and Scott, Foresman Company; between Harcourt, Brace and Jovanovich and Houghton-Mifflin Company; between Harcourt, Brace and Jovanovich and Scott, Foresman Company; and between Holt, Rinehart, and Winston and Scott, Foresman Company K-3 reading programs (see Table 9).

5. The Reading Concepts Checklist, (RCC), findings indicate that according to pairwise comparison scores for six subcategories, (I: "Auditory Discrimination," II: "Visual Discrimination," III: "Phonic Analysis," IV: "Structural Analysis," VI: "Literal Comprehension," VII: "Critical Comprehension"), concurrence between the K-3 reading instructional program is present in a significant degree in all five reading instructional programs (see Appendix D).

6. The Reading Concepts Checklist, (RCC), findings indicate that according to pairwise comparison scores for three subcategories, (V: "Vocabulary Development," VII: "Inferential Comprehension," and IX: "Study Skills"), concurrence between the K-3 reading instructional programs is lacking between Ginn and Company and Houghton-Mifflin Company; between Ginn and Company and Scott, Foresman Company;

between Harcourt, Brace and Jovanovich and Houghton-Mifflin Company; between Harcourt, Brace and Jovanovich and Scott, Foresman Company; between Holt, Rinehart, and Winston and Houghton-Mifflin Company; between Holt, Rinehart, and Winston and Scott, Foresman Company; between Houghton-Mifflin Company and Scott, Foresman Company. The findings further indicate, according to pairwise comparison scores, concurrence is lacking in a significant degree for the subcategory IX: "Study Skills" between Ginn and Company and Harcourt, Brace and Jovanovich K-3 reading programs (see Table 10).

7. The Reading Concepts Checklist, (RCC), findings indicate that two major divisions, I: "Auditory Discrimination" and II: "Visual Discrimination," were neither taught in the 4-6 reading instructional programs nor tested in the Michigan Educational Assessment Program Experimental Reading Test Grade 7, leaving four major divisions with seven subcategories in the Reading Concepts Checklist, (RCC).

8. The Reading Concepts Checklist, (RCC), findings indicate that according to pairwise comparison scores for four 4-6 reading instructional programs, Ginn and Company; Harcourt, Brace and Jovanovich; Holt, Rinehart, and Winston; Scott, Foresman Company, concurrence between the Michigan Educational Assessment Program Experimental Reading Test Grade 7 and the reading programs is lacking to a significant

degree in all seven subcategories of the Reading Concepts Checklist, (RCC), (see Appendix H).

9. The findings of the Reading Concepts Checklist, (RCC), indicate that, according to pairwise comparions scores for Houghton-Mifflin Company's 4-6 reading program, concurrence between the Michigan Educational Assessment Program Experimental Reading Test Grade 7 and the reading program is lacking in five subcategories of the Reading Concepts Checklist, (RCC), while concurrence is present in a significantly greater degree in the subcategories III: "Phonic Analysis" and V: "Vocabulary Development" (see Appendix H).

10. According to pairwise comparison scores, the Reading Concepts Checklist, (RCC), findings indicate that concurrence between each of the 4-6 reading instructional programs is present in a significantly greater degree between Ginn and Company and Harcourt, Brace and Jovanovich; between Ginn and Company and Holt, Rinehart, and Winston; between Harcourt, Brace and Jovanovich and Holt, Rinehart, and Winston; and between Houghton-Mifflin Company and Scott, Foresman Company (see Table 19).

11. The findings also indicate a lack of concurrence in a significant degree between Ginn and Company and Houghton-Mifflin Company; between Ginn and Company and Scott, Foresman Company; between Harcourt, Brace and Jovanovich and Houghton-Mifflin Company; between Harcourt,

Brace and Jovanovich and Scott, Foresman Company; between Holt, Rinehart, and Winston and Houghton-Mifflin Company; and between Holt, Rinehart, and Winston and Scott, Foresman Company (see Table 19).

12. The findings of the pairwise comparison scores in three subcategories, V: "Vocabulary Development," VII: "Critical Comprehension," and IX: "Study Skills," indicate concurrence between the 4-6 reading instructional programs is present in a significant degree in all five reading programs (see Appendix G).

13. The Reading Concepts Checklist, (RCC), findings indicate that scores in three subcategories, IV: "Structural Analysis," VI: "Literal Comprehension," and VII: "Inferential Comprehension," concurrence between the 4-6 reading instructional programs is lacking in a significant degree between Ginn and Company and Houghton-Mifflin Company; between Ginn and Company and Scott, Foresman Company; between Harcourt, Brace and Jovanovich and Houghton-Mifflin Company; between Harcourt, Brace and Jovanovich and Scott, Foresman Company; between Holt, Rinehart, and Winston and Houghton-Mifflin Company; and between Holt, Rinehart, and Winston and Scott, Foresman Company. The findings further indicate that in two subcategory scores, VI: "Literal Comprehension," and VII: "Inferential Comprehension," there is a lack of concurrence between Houghton-Mifflin Company and Scott, Foresman Company 4-6 reading programs (see Table 19 and Appendix G).

14. The findings of the inter-rater reliability study indicate a strong positive relationship among the judges regarding the relationship of the concepts being tested by the Michigan Educational Assessment Program Experimental Reading Test Grades 4 and 7, as measured by the Reading Concepts Checklist, (RCC).

Conclusions

The findings of the empirical study of the degree of concurrence between the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven and the selected K-6 reading instructional programs as measured by the Reading Concepts Checklist, (RCC), can be evaluated from several perspectives. A major concern of the analysis was to test the total proportional measurement of the content domain. A second major concern of this study was the investigation of the relationships between the K-6 reading instructional programs. A final component of this study involved the use of a panel of reading experts to review and evaluate the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven and perform an inter-rater reliability test to measure the strength of the relationship of their judgments. All three components of this study are interrelated and will be evaluated in terms of their significant interrelationships.

Relationships Between Michigan Experimental Reading Test and K-6 Reading Instructional Programs

1. The predominant aspect of the results is the lack of concurrence between the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven and the selected K-6 reading instructional programs as shown by the total pairwise comparison scores, and the pairwise comparison scores of the individual categories contained in the Reading Concepts Checklist, (RCC). This lack of congruence between the concepts presented in the K-6 reading instructional programs and the concepts tested in the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven show the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven not to be content valid.

Relationship Between Inter-rater Reliability Study to the Michigan Educational Assessment Program Experimental Reading Test Grades 4 and 7

2. There is agreement among the independent judges pertaining to which concepts contained in the Reading Concepts Checklist, (RCC), are being measured by the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven. This demonstrates the reliability of the data which were recorded in the Reading Concepts Checklist, (RCC), and compared with the reading instructional programs. The reliability study shows more concepts

contained in the Reading Concepts Checklist, (RCC), were not measured proportionally by the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven than were measured. Therefore, the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven does not fulfill the requirement of constituting a representative sample of the behaviors to be exhibited in the desired performance domain.

3. The agreement among the independent judges that the Michigan Educational Assessment Program Experimental Reading Test Grade Four and Seven leaves large portions of the Reading Concepts Checklist, (RCC), unmeasured shows that the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven, is insensitive to instruction, based upon those reading programs reviewed in this study.

Relationships Between the K-6 Reading Instructional Programs

4. The major feature of the results of all of the statistical tests concerning the data regarding the concepts presented in the K-6 reading instructional programs is that the five instructional programs may be classified as belonging to one of two groups: 1) Ginn and Company; Harcourt, Brace and Jovanovich; and Holt, Rinehart, and Winston and 2) Houghton-Mifflin Company and Scott, Foresman

Company. The differences between reading instructional programs apparently reflect differences in program emphasis or in philosophical approaches to the teaching of reading.

5. The major differences between the two groups of K-3 reading instructional programs are in categories V, (Vocabulary Development), VII, (Inferential Comprehension), and IX, (Study Skills). These differences may indicate a high degree of variation in student performance on the Michigan Educational Assessment Program Experimental Reading Test Grade Four.

6. The major differences between the two groups of 4-6 reading instructional programs are in categories III, (Phonic Analysis), IV, (Structural Analysis), VI, (Literal Comprehension), and VII, (Inferential Comprehension). A result of these differences may be a high degree of variation in student performance on the Michigan Educational Assessment Program Experimental Reading Test Grade Seven.

7. The results of the analyses provide confirmation of the expected relationship between the K-6 reading instructional programs and the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven. The total proportion scores confirm the absence of congruence between the K-6 reading instructional programs and the Michigan Experimental Reading Test. The results of the inter-rater reliability study establish the relationship between the Reading Concepts Checklist, (RCC), and the reading instructional programs.

8. The results indicate that according to scores of each of six divisions and nine individual categories, concurrence between the Michigan Educational Assessment Program Experimental Reading Test Grade Four and Seven, and the selected K-6 reading instructional programs is absent to a significant degree.

Implications

The findings of the study are based on data collected through surveying the five reading instructional programs' sixty-five teachers' manuals. The five reading programs were selected on the basis of a 1977 national survey of reading instructors and reading specialists. The survey revealed that the predominant reading materials used in the region which includes Michigan are 1) Ginn and Company, 2) Harcourt, Brace and Jovanovich, 3) Holt, Rinehart, and Winston, 4) Houghton-Mifflin Company, and 5) Scott, Foresman Company. Indicating the percentage to be 75.86, the survey also provided data which satisfied the lower acceptable limit definition of seventy-five percent of Michigan's K-6 students using the reading materials.

The findings indicate significant differences between what the Michigan Educational Assessment Program Experimental Reading Test Grade Four and Seven presumes to test and the concepts presented in the selected K-6 reading instructional

programs. Some explanations for these findings are given in the implications which follow.

1. Some may assume that the Michigan Educational Assessment Program Experimental Reading Test Grade Four and Seven is a fourth or seventh grade test and tests curriculum from those grades.[1] Since the tests are administered during the initial weeks of the school year for fourth and seventh grade students, the tests are a measure of the preceding grades.

The findings of the study indicate large blocks of the Reading Concepts Checklist, (RCC), are not measured by the Michigan Educational Assessment Program Experimental Reading Test, Category III, (Phonic Analysis), while the reading instructional programs emphasize this decoding skill. A major consideration: if this area is not measured by the Michigan Experimental Test, does the failure of a student to achieve the goal established for successfully completing the test's tasks for inferential comprehension signal faulty comprehension skills? Or, does the fault rest with the test for not measuring a representative sample of the concepts presented in the reading instructional programs? The first major implication is that the Michigan Educational

---

[1]The Michigan Department of Education has stressed, however, that the fourth and seventh grade tests are measures of learning in the preceding grades.

Assessment Program Experimental Reading Test Grades Four and Seven is not sensitive to the curriculum.

2. It was recognized early in this study that the accountability movement has brought public pressure to bear upon boards of education and educators at all levels and in various capacities of education. The Michigan Educational Assessment Program Experimental Reading Test is symbolic of one of the responses to the movement. It might be assumed by educators or boards of education that the results of the Michigan Educational Assessment Program Experimental Reading Test are a reflection of the quality of education within the local district. The findings of this study indicate the Michigan Educational Assessment Program Experimental Reading Test is not an accurate measure of the effectiveness of the local curriculum. The implication is that before boards of education make decisions about curricular effectiveness on the basis of scores achieved on the Michigan Educational Assessment Program Experimental Reading Test, additional data concerning the effectiveness of the curriculum needs to be assembled.

3. Building administrators and classroom teachers should exercise caution in attempting to assess the needs of the building or the individual classroom on the basis of the Michigan Educational Assessment Program Experimental Reading Test's results. The findings of this study that the Michigan Experimental Test is not sensitive to the curriculum indicate the success or failure of a student on

the Experimental Test is not an indication of the student's achievement. Reprogramming to meet the presumed needs of the student may well be inappropriate and uncessary, if not potentially an impediment to student progress.

4. Some may suggest the lack of congruence between the Michigan Educational Assessment Program Experimental Reating Test, Grade Four and Seven, and each of the reading instructional programs results from the Experimental Reading Test's measurement of minimal performance objectives. The implication is that the reading instructional programs are so comprehensive in nature that the test can not fit the reading programs.

The measurement of minimal performance levels neither eliminates the requirement that the test be a "representative sample" of the content domain nor its obligation to measure the essential elements of the content domain. If decoding is an essential reading skill, presented by the reading programs and not measured by the Experimental Reading Test, it can not be stated with certainty that the Experimental Reading Test's measurement of minimal performance levels is a measurement of essential minimal performance levels.

Recommendations

Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven

1. It is recommended that the Michigan Department of Education undertake a complete revision of the Michigan

Educational Assessment Program Experimental Reading Test Grades Four and Seven. The Michigan Educational Assessment Program Experimental Reading Test should be redeveloped on the Basis of item-by-treatment interaction where the items of the test are in direct proportion to the concepts presented in the reading instructional programs.

2. It is recommended that the Michigan Department of Education engage the services of a nationally known panel of reading experts to review and evaluate the revised version of the Michigan Educational Assessment Program Experimental Reading Test to establish the relationship between the K-8 reading instructional programs used throughout Michigan and the revised version of the Michigan Educational Assessment Program Experimental Reading Test Grades Four and Seven.

Development of a Communications Process and Favorable Attitudes

3. The present controversy surrounding the current Michigan Educational Assessment Program Experimental Reading Test has created a schism between those in support of the

test and those who are its critics. The local Board of Education's or the individual educator's opportunity to influence mandated statewide educational policy is preceived to be greatly reduced. If the communications process is lacking or totally inadequate, the schism will increase. Therefore, it is recommended that the Michigan Department of Education and all educators recognize the challenge before them and use their ingenuity to develop new avenues of communicating with each other.

Revision, Continued Development and Use of the Reading Concepts Checklist, (RCC)

4. It is recommended that a revision of the categories having a relatively low correlation with total proportion scores and/or pairwise comparison scores between reading instructional programs should be made. The individual reading concepts within categories III, (Phonic Analysis), IV, (Structural Analysis), V, (Vocabulary Development), VI, (Literal Comprehension), VII, (Inferential Comprehension), and IX, (Study Skills) should be revised with higher levels of specificity and studied further to identify the bases for lack of concurrence between the reading instructional programs.

5. It is recommended that periodic use of the Reading Concepts Checklist, (RCC), should include an investigation of the stability of the measures derived from the instrument

to determine the extent of fluctuations in the concepts presented in the K-6 reading instructional programs. Knowledge of these variations in the concepts presented in the reading instructional programs could effectively supplement improvements to the quality of Michigan Educational Assessment Program testing in Michigan.

APPENDICES

APPENDIX A

READING CONCEPTS CHECKLIST: CLASSIFICATION OF INSTRUCTIONAL CONCEPTS

AND

READING CONCEPTS CHECKLIST: CLASSIFICATION OF TESTED CONCEPTS

Reading Concepts Checklist: Classification of Instructional Concepts

KEY:
1. Ginn and Company
2. Harcourt, Brace and Jovanovich
3. Holt, Rinehart, and Winston
4. Houghton-Mifflin Company
5. Scott, Foresman Company

Concept	Grade K	1	2	3	4	5	6
1.0 Auditory Discrimination							
1.001 Word Sounds							
1.002 Words in Sentences							
1.003 Beginning Consonants							
1.004 Ending Consonants							
1.005 Consonant Blends							
1.006 Rhyming Words							
2.0 Visual Discrimination							
2.007 Upper Case Letter Names							
2.008 Lower Case Letter Names							
2.009 Words in Sentences							
2.010 Words in Paragraph							
3.0 Phonic Analysis							
3.011 Beginning Consonants							
3.012 Ending Consonants							
3.013 Medial Consonants							
3.014 Beginning Blends							
3.015 Ending Blends							
3.016 Beginning Consonant Digraphs							
3.017 Beginning Blends and Digraphs							
3.018 Ending Blends and Digraphs							
3.019 Medial Consonants and Digraphs							

Appendix A Continued.

Concept	Grade K	1	2	3	4	5	6
3.01 Vowel Sounds							
3.020 Short Vowel Sounds							
3.021 Long Vowel Sounds							
3.022 Vowel Digraphs							
3.023 Vowel Diphthongs							
3.024 The Schwa Sound							
3.025 Context Clues							
3.026 "R" Controlled Vowel							
4.0 Structural Analysis							
4.027 Root Words							
4.028 Word Endings							
4.029 Word Families							
4.030 Contractions							
4.031 Compound Words							
4.032 Possessives							
4.033 Prefixes							
4.034 Suffixes							
4.035 Syllabication							
4.036 Accent Clues							
5.0 Comprehension							
5.01 Vocabulary Development							
5.037 Synonyms							
5.038 Antonyms							
5.039 Homonyms							
5.040 Context Clues							
5.02 Literal Comprehension							
5.041 Multiple Meaning of words							
5.042 Word Recognition							
5.043 Likenesses and Differences							

Continued

Appendix A Continued.

Concepts	Grade K	1	2	3	4	5	6
5.044 Syntax							
5.045 Word Meaning							
5.046 Sentence Meaning							
5.047 Paragraph Meanings							
5.048 Punctuation							
5.049 Character Development							
5.050 Main Idea							
5.051 Details							
5.052 Place Events in Proper Sequence							
5.053 Plot and Setting							
5.054 Cause and Effect							
5.055 Gathering Information from Pictures							
5.056 Classifying							
5.02 Inferential Comprehension							
5.057 Idiom							
5.058 Similie							
5.059 Metaphor							
5.060 Alliteration							
5.061 Onomatopoeia							
5.062 Personification							
5.063 Author's Style							
5.064 Mood or Tone							
5.065 Draw Logical Conclusions							
5.066 Predict Outcomes							
5.067 Character Development							
5.068 Main Idea							

Continued

Appendix A. Continued.

Concept	Grade K	1	2	3	4	5	6
5.069 Details							
5.070 Place Events in Proper Sequence							
5.071 Plot and Setting							
5.072 Cause and Effect							
5.073 Analogies							
5.04 Critical Comprehension							
5.074 Judge Accuracy							
5.075 Judge Validity							
5.076 Distinguish Fact from Opinion							
5.077 Author's Purpose							
5.078 Author's Point of View							
5.079 Distinguish Realims From Fantasy							
5.080 Detect Propaganda, Persuasion, Bias							
5.081 Verify Conclusions							
6.0 Study Skills							
6.082 Use Table of Contents							
6.083 Use Index							
6.084 Use Glossary							
6.085 Use Encyclopedia							
6.086 Use Index Volume							
6.087 Find a Topic							
6.088 Cross Reference							
6.089 Read Maps							
6.090 Read Charts, Graphs, Diagrams							
6.091 Dictionary Skills							
6.092 Alphabetize 1st, 2nd, 3rd Letter etc.							

Continued

Appendix A Continued.

Concept	Grade K	1	2	3	4	5	6
6.093 Use Pronunciation Key							
6.094 Locate Entry Word							
6.095 Guide Words							
6.096 Parts of Speech							
6.097 Skimming and Scanning							
6.098 Follow Written Directions							
6.01 Organizational Study Skills							
6.099 Topic Selection							
6.100 Subtopic Selection							
6.101 Outlining							
6.102 Summarizing Selection							
6.103 Reading Newspapers and Magazines							

Reading Concepts Checklist:
Classification of Tested Concepts

Concept	Grade Level Tested Grade 4	Grade 7
1.0 Auditory Discrimination		
1.001 Word Sounds		
1.002 Words in Sentences		
1.003 Beginning Consonants		
1.004 Ending Consonants		
1.005 Consonant Blends		
1.006 Rhyming Words		
2.0 Visual Discrimination		
2.007 Upper Case Letter Names		
2.008 Lower Case Letter Names		
2.009 Words in Sentences		
2.010 Words in Paragraph		
3.0 Phonic Analysis		
3.011 Beginning Consonants		
3.012 Ending Consonants		
3.013 Medial Consonants		
3.014 Beginning Blends		
3.015 Ending Blends		
3.016 Beginning Consonant Digraphs		
3.017 Beginning Blends and Digraphs		
3.018 Ending Blends and Digraphs		
3.019 Medial Consonants and Digraphs		
3.01 Vowel Sounds		
3.020 Short Vowel Sounds		

Continued

Appendix A Continued.

Concept	Grade Level Tested	
	Grade 4	Grade 7
3.021 Long Vowel Sounds		
3.022 Vowel Digraphs		
3.023 Vowel Diphthongs		
3.024 The Schwa Sound		
3.025 Context Clues		
3.026 "R" Controlled Vowel		
4.0 Structural Analysis		
4.027 Root Words		
4.028 Word Endings		
4.029 Word Families		
4.039 Contractions		
4.031 Compound Words		
4.032 Possessives		
4.033 Prefixes		
4.034 Suffixes		
4.035 Syllabication		
4.036 Accent Clues		
5.0 Comprehension		
5.01 Vocabulary Development		
5.037 Synonyms		
5.038 Antonyms		
5.039 Homonyms		
5.040 Context Clues		
5.041 Multiple Meaning of Words		
5.02 Literal Comprehension		
5.042 Word Recognition		
5.043 Likenesses and Differences		
5.044 Syntax		

Continued

Appendix A Continued.

Concept	Grade Level Tested Grade 4	Grade 7
5.045 Word Meaning		
5.046 Sentence Meaning		
5.047 Paragraph Meaning		
5.048 Punctuation		
5.049 Character Development		
5.050 Main Idea		
5.051 Details		
5.052 Place Events in Proper Sequence		
5.053 Plot and Setting		
5.054 Cause and Effect		
5.055 Gathering Information From Pictures		
5.056 Classifying		
5.03 Inferential Comprehension		
5.027 Idiom		
5.058 Similie		
5.059 Metaphor		
5.060 Alliteration		
5.061 Onomatopoeia		
5.062 Personification		
5.063 Author's Style		
5.064 Mood or Tone		
5.065 Draw Logical Conclusions		
5.066 Predict Outcomes		
5.067 Character Development		
5.068 Main Idea		
5.069 Details		
5.070 Place Events in Proper Sequence		

Continued

Appendix A Continued

Concept	Grade Level Tested Grade 4	Grade 7
5.071 Plot and Setting		
5.072 Cause and Effect		
5.073 Analogies		
5.04 Critical Comprehension		
5.074 Judge Accuracy		
5.075 Judge Validity		
5.076 Distinguish Fact From Opinion		
5.077 Author's Purpose		
5.078 Author's Point of View		
5.079 Distinguish Realism From Fantasy		
5.080 Detect Proaganda, Persuasion, Bias		
5.081 Verify Conclusions		
6.0 Study Skills		
6.082 Use Table of Context		
6.083 Use Index		
6.084 Use Glossary		
6.085 Use Encyclopedia		
6.086 Use Index Volume		
6.087 Find a Topic		
6.088 Cross Reference		
6.089 Read Maps		
6.090 Read Charts, Graphs, Diagrams		
6.091 Dictionary Skills		
6.092 Alphabetize 1st, 2nd, 3rd Letter etc.		
6.093 Use Pronunciation Key		
6.094 Locate Entry Word		

Continued

Appendix A Continued.

Concept	Grade Level Tested Grade 4	Grade 7
6.095 Guide Words		
6.096 Parts of Speech		
6.097 Skimming and Scanning		
6.098 Follow Written Directions		
6.01 Organizational Study Skills		
6.099 Topic Selection		
6.100 Subtopic Selection		
6.101 Outlining		
6.102 Summarizing Selection		
6.103 Reading Newspapers and Magazines		

APPENDIX B

COMMUNICATION SKILLS OBJECTIVES

COMMUNICATION SKILLS OBJECTIVES

--Reading

-- Speaking/Listening

-- Writing

With Examples of Experiences and Activities
and Suggested Measurement Approaches

Michigan Department of Education

January, 1979

READING

PROPOSED READING OBJECTIVES

Competency	Measureable Behavior (3rd Grade)	Measurable Behavior (6th Grade)	Measurable Behavior (9th Grade)
I. Vocabulary Meaning	By the end of the third grade, the student will be able to:	By the end of the sixth grade, the student will be able to:	By the end of the ninth grade, the student will be able to:
	A. Determine the meaning of a word in a sentence whose meaning has been affected by prefixes.	A. Determine the meaning of a word in a sentence whose meaning has been affected by prefixes.	A. Determine the meaning of a word in a sentence whose meaning has been affected by common prefixes.

Example Experiences and/or Activity	Measurement
1. Give students words whose meanings can be affected by prefixes. Also, give them lists of prefixes to use with the words, or have them think of their own prefixes to use. Discuss in what way the words have changed in meaning and what the various prefixes must, therefore, mean.	1. Give students words with prefixes and have them choose from four or more choices the meaning of the prefix. For example, given the word "reorganize," the student should choose the response "to organize again."
2. Compile lists of prefixes. Have students discuss or verify their meanings in the dictionary. Have them use the prefixes in their own writing.	2. Have students write sentences or a selection using a given list of prefixes correctly.
3. Have students locate prefixes in their textbooks and keep a list of these prefixes.	

B. Determine the meaning of a word in a sentence whose meaning has been affected by suffixes.	B. Determine the meaning of a word in a sentence whose meaning has been affected by suffixes.	B. Determine the meaning of a word in a sentence whose meaning has been affected by common suffixes.

Example Experiences and/or Activity	Measurement
1. Give students words whose meanings can be affected by suffixes. Also, give them lists of suffixes to use with the words, or have them think of their own suffixes to use. Discuss in what ways the words have changed in meaning and what various suffixes must, therefore, mean.	1. Give the students words with suffixes and have them choose from four or more choices the meaning of a suffix. For example, given the word "careless," the student should choose the response, "careless means without care."

2. Compile lists of suffixes. Have students discuss or verify their meanings in the dictionary. Have them use the suffixes in their own writing.

3. At upper grade levels, students may learn that suffixes often affect the way a word is used in a sentence; i.e., the part of speech. For example, "careless" is an adjective; "carelessly" is an adverb.

2. Have students write sentences or a selection using a given list of suffixes correctly.

C. Determine the meaning of a word that has multiple meanings, depending on its use in a sentence.

D. Determine the meaning of a word that has multiple meanings, depending on its use in a sentence.

C. Determine the meaning of a word that has mutliple meanings, depending on its use in a sentence.

Example Experiences and/or Activity

1. Write a word that has multiple meanings on the board. Ask the students to think of as many meanings for the word as possible and use the word in these various ways in sentences. For example, the word "circle" may mean to walk around something, to draw a round line, a ring, or a private group of people. Thus, "The cat circled the wounded bird," "Cicle the right answer." "We sat in a circle." "Do you belong to the inner circle?"

2. Have students look up a word that has multiple meanings in the dictionary. Discuss the various meanings and use in sentences. More complex words may have many slightly different meanings.

3. Use library books, such as Amelia Bedelia, The King Who Rained, and Jake, which make humorous use of the multiple meaning of words, to illustrate this principle. Have students write similar selections, either as individuals or as grours.

4. At the more advanced levels, discuss how words differ in connotations as well as denotations. Also discuss how words may differ in various subject areas; such as "culture" in social studies, "culture" in science, and "cultured" in the arts.

Measurement

1. Give the students a sentence with an underlined word that can have multiple meanings. Ask them to choose from a list of four or more meanings the one that is appropriate to its use in the sentence.

2. Give the students a word that has multiple meanings. Ask them to write sentences using the word according to its various meanings.

3. Give the students a sentence containing an underlined word that may have multiple meanings. Also, give the students a list of dictionary definitions of that word. Ask them to check the definition most appropriate to its use in the sentence.

D. Identify a word that has a similar meaning to another word (identifying synonyms).

D. Identify a word that has a similar meaning to another word (identifying synonyms).

D. Identify a word that has a similar meaning to another word (identifying synonyms).

Example Experiences and/or Activity	Measurement
1. Present the students with a word and ask them to think of as many synonyms as possible. Students may use dictionaries, thesauruses, and so on to locate additional synonyms.	1. Give students a sentence with an underlined word. Also give them a choice of four or more words from which to select a synonym for the underlined word.
2. Have students read poetry in which synonyms are used for artistic purposes, such as "The Cataract of Lodore." Discuss how even synonyms have fine differences in meaning.	2. Give students a word that has many synonyms. Ask them to list at least three (or some other number) synonyms for the word.
3. Have students re-write their own selections, using synonyms for the words they originally used.	

E. Identify a word that has an opposite meaning to another word (identifying antonyms).

E. Identify a word that has an opposite meaning to another word (identifying antonyms).

E. Identify a word that has an opposite meaning to another word (identifying antonyms).

Example Experiences and/or Activity	Measurement
1. Present the students with a word and ask them to think of as many antonyms as possible. Students may use dictionaries, thesauruses, and so on to locate additional antonyms.	1. Give students a sentence with an underlined word. Also give them a choice of four or more words from which to select an antonym for the underlined word.
2. Arrange students in groups and have them compete to find as many antonyms for a given number of words as possible.	2. Give students a word that has many antonyms. Ask them to list at least three (or some other number) antonyms for the word.

F. Determine the meaning of a word on the basis of the context of a sentence.

F. Determine the meaning of a word on the basis of the context of a sentence.

F. Determine the meaning of a word on the basis of the context of a sentence.

Example Experiences and/or Activity

1. When listening to students read, if they have difficulty decoding a word, encourage them to consider context clues.

2. Present students with sentences containing words they may not know the meaning of. Have them discuss what they think the word might mean on the basis of its use in the sentence. Have students verify their guesses in the dictionary.

3. Prior to having the students read a section of one of their textbooks, such as a social studies, science, health textbook, list the words that they may not know on the board. Have them discuss what they think the words may mean; then have them read the words in the context of the passage and continue the discussion. If necessary, verify their guesses in the dictionary or the glossary of the textbook. Then have them proceed to read the assigned chapter or selection.

Measurement

1. Give the students a sentence containing a word they probably will not know. Also, give them a list of possible definitions for the word. Ask them to select the most appropriate definition, given its use in the sentence.

2. Give the students a sentence containing a nonsense word. Also, give them a list of possible definitions for the word. Ask them to select the most appropriate definition, given its use in the sentence.

Competency	Measurable Behavior (3rd Grade)	Measurable Behavior (6th Grade)	Measurable Behavior (9th Grade)
II. Literal Comprehension	By the end of the third grade, the student will be able to: A. Read a selection using a knowledge of structure of the language including syntactic and semantic clues (cloze procedure).	By the end of the sixth grade, the student will be able to: A. Read a selection using a knowledge of the structure of the language including syntactic and semantic clues (cloze procedure).	By the end of the ninth grade, the student will be able to: A. Read a selection using a knowledge of the structure of the language including syntactic and semantic clues (cloze procedure).

Example Experiences and/or Activity

The cloze procedure may be used to determine the student's approximate reading level and to match her or his reading level and needs to the materials being used. It is probably one of the simplest ways to determine a student's literal comprehension level. Procedures are as follows:

1. Select a paragraph. Perhaps it may be from the reading material the student is to read for the class. The length of the paragraph may vary, depending upon the level of difficulty. For pupils in the third grade and above, passages should be at least 25 words long.

2. Delete every fifth word in the selection and replace each omitted word with a blank of standard length. Do not delete a word in the first or last sentence.

3. Ask the student to read the selection and fill in the missing words.

4. Score the test by counting the number of words correctly supplied by the student. Do not penalize students for incorrect spellings. If a student supplies a word that makes as much sense to the meaning as the original word (such as supplying the word "blue" for the phrase "...the ____ball"), it may be counted as acceptable.

Measurement

1. Student Instructions:
In this exercise you are to read several paragraphs. Every fifth word in each paragraph has been left out. As you read the paragraphs, figure out which word was taken out of each space and write it in. Only one word goes in each blank. If you are not sure of the word, you may guess.

Example:

TODAY'S CATTLE RANCHERS
(Third grade level)

John's father is a rancher who owns many cattle. Once each year John _____his father take their _____to market to sell _____. Many years ago ranchers _____to take their cattle _____ market by herding them _____ horses.

John and his _____do not do this. _____ load their cattle in _____cars owned by the _____company. Then the engineer _____ his train and hooks _____ to the cattle cars _____ hauls them to market. John and his father ride on the train with the cattle.

5. There is no standard way to "score" a cloze procedure. The teacher should use her or his own judgment as to "levels" of difficulty. Below is a suggested standard for making a judgment:
If a student supplies 70 to 100% of the missing words correctly, he/she is reading the passage at an independent level; that is, he/she can read it quite easily.

If a student supplies 40 to 69% of the missing words correctly, he/she is reading the passage at an instructional level; that is he/she can read it with some effort and perhaps assistance from the teacher.

If a student supplies 39% or less of the missing words correctly, he/she is reading the passage at a frustration level; that is, the material is probably too difficult to use even for instructional purposes.

B. Identify the stated main idea within a selection.

B. Identify the stated main idea within a selection.

B. Identify the stated main idea within a selection.

Example Experiences and/or Activity

1. Present students with a selection in which the main idea is clearly stated. Go through each sentence and have students discuss which one seems to best describe what the whole selection is about. Discuss what the term "main idea" means.

2. Give students a sentence that can serve as a "main idea" for a selection. Have them write a selection using the sentence as the main idea. Have other students locate the sentence in the selection . Or have students think of their own main-idea sentences and then have them develop selections using these sentences.

3. Before having students read a chapter or a section in one of their textbooks, such as a science or social studies textbook, go through the chapter or section as a whole and attempt to locate sentences that may represent what the main idea of the whole chapter is likely to be.

Measurement

1. Have students read a paragraph in which one sentence or phrase represents the main idea. Ask them to identify that sentence or phrase.

Do the same with parts of the chapter or paragraphs within the chapter. After they have read the selection, discuss whether the sentences were indeed the main ideas.

4. Have students find examples in newspapers and magazines of sentences or phrases that state the main idea of the selection.

C. Identify details that support the main idea of a selection.

C. Identify details that support the main idea of a selection.

C. Identify details that support the main idea of a selection.

Example Experiences and/or Activity	Measurement
1. Give students a selection in which the main idea is stated. Have them find statements within the passage that support the main idea. 2. Give students a sentence that can serve as a main idea, such as "Australia has a lot of unusual animals." Have them write a paragraph that justifies this statement. The justifying statements thus support the main idea. 3. Have students locate main ideas in their textbooks and in magazines and newspapers. Have them point out the sentences that support the main idea.	1. Have students read a selection in which the main idea is stated. List four or more choices that support the main idea and have students select the appropriate choice. For example, the selection may be "Family Life on the Prairie." The main idea is that all members of the family had work to do. The question might be: "What did little girls do to help?" The correct choice, on the basis of the selection, might be "...they helped prepare the meals."

D. Identify information within a selection on the basis if recall.

D. Identify information within a selection on the basis of recall.

D. Identify information within a selection on the basis of recall.

Example Experiences and/or Activity	Measurement
1. Have students read a selection. After they have finished, ask them about specific information contained within the passage without referring back to the selection. Or allow varying lengths of time to lapse before asking them to recall the information.	1. Have students read a selection. Without having them refer back to the selection, ask them to identify information presented in the selection, perhaps through multiple choice questions.

Practice this regularly. Over a period of time, students who may have difficulty recalling information will acquire more of a facility to do so. The activity can be made into a game, the winners being those who can recall the most information. This can be done in groups as well as with individuals.

2. Have students read a selection. Then ask them to make a list of all the information they can remember from their reading. On the basis of their lists, have them re-write the selection without referring to the original. Then have them compare their versions to the original. Discuss in what ways the re-written paragraphs are better than or not as good as the original.

3. Have students discuss mental techniques they may use to recall information. Discuss various factors that seem to affect the ability to recall. Is the time lapse between the reading and the recall important? Is the content itself a factor? Do those who understand the whole passage more fully recall the details better?

2. Have students read a selction. Present them with lists of information that may have been presented in the selection. Have them check those items that are factually correct.

E. Identify the sequence within a selection.

E. Identify the sequence within a selection.

E. Identify the sequence within a selection.

Example Experiences and/or Activity	Measurement
1. Have students read a selection in which sequenitality is clearly stated, especially by such as words as "first," "second," "thirdly," "then," "later," "next," "soon," "finally," "before," and so on. Following their reading, have students discuss or list the elements in the selection as they were presented. Include selections in which actual events are not related in the specific order they occurred. (For example, when a character in a story is walking home from school, he may be thinking back to how he got into trouble in school--and that trouble all started with something that happened yesterday. He dreads	1. Have students read a selection in which sequentiality is clearly stated. Ask questions to determine if they understand what followed what. For example, "What did Fred do as soon as knew he was lost?" Correct answer: "... he climbed a tree." (The selection says he built a fire later.)

getting home, because he knows the teacher has called his father. When he gets home, his father meets him at the door, and the concluding events are related.)

2. Give students a list of events. Have them write a narrative about the events relating them in various orders. Some may tell the story in the order of the events, some may start in the middle, some at the end.

3. Have students read books and stories, such as mystery stories, in which the sequence of events as they actually occurred (and not where they were actually related in the story) is a key factor.

4. Have students write expository selections that require a step by step treatment. Encourage them to use words that guide the reader through the exposition clearly.

F. Identify stated cause and effect relationships within a selection.	F. Identify stated cause and effect relationships within a selection.	F. Identify stated cause and effect relationships within a selection.

Example Experiences and/or Activity	Measurement
1. Have students read selections in which cause and effect relationships are clearly stated. In discussions or through individual work, have them identify the stated causes and effects. 2. Have students list words and phrases that denote cause and effect relationships. Such words and phrases may include "because," "as a result," "therefore." Sentence structure may also suggest cause and effect relationship; as in "The Civil War, brought on by the slavery issue, occurred in the 1860's" and "The War contributed much to the North's industrial development." Have students locate examples of cause and effect relationships that are clearly stated, but not through the use of words typically used to denote these relationships.	1. Give students a selection in which cause and effect relationships are clearly stated. Ask questions to determine if they comprehend these relationships. For example, "Why did Mary start crying?" Correct response"...because her friend left without her."

3. Have students locate examples of cause and effect relationships in their textbooks.

G. Identify stated likenesses and differences within a selection.	G. Identify stated likenesses and differences within a selection.	G. Identify stated likenesses and differences within a selection.

Example Experiences and/or Activity	Measurement
1. Present students with selections in which likenesses and differences are clearly stated. For example, "Wolves are like dogs in many ways, but they're also different from dogs." (The selection goes on to explain these likenesses and differences.) Have students list or discuss the stated likenesses and differences.	1. Give students a passage in which likenesses and differences are clearly stated. Have them identify these likenesses and differences. Question: "Who did Jane look like?" Answer: "Jane looked like Mary."
2. Have students find examples of stated likenesses and differences in their textbooks and other reading material.	
3. Have students discuss ways in which their school building is like other school buildings and ways it differs. List as many likenesses and differences on the board. Do this with various words, ranging from words that denote concrete objects ("How are a basketball and a baseball alike and different?") to words that denote abstractions ("How are nations and states alike and different?")	
4. Have students group various objects and words together (formulate concepts) according to their likenesses. Have them justify their groupings (concepts). ("I put "doll, "ball," and "blocks" together because they're all toys.") Have students discuss how things can be alike in some ways; different in others.	

H. Identify the meaning of a sentence based on punctuation--periods, commas, question marks, exclamation marks, and quotation marks.

H. Identify the meaning of a sentence based on punctuation--periods, commas, question marks, exclamation marks, and quotation marks.

Example Experiences and/or Activities	Measurement
1. Give the students various versions of a selection --one that is punctuated according to common usage, one that is poorly, or even ludicrously, punctuated, and one that is not punctuated at all. Have students read the passages either aloud or silently and discuss what effect the punctuation or lack of punctuation had on their ability to read the selection easily.	1. Give students a selection with punctuation omitted. Have them supply correct punctuation according to meaning. For example, "Was the house painted white ( )" ____(.) ____(,) ____(?) ____(;)
2. Have students read aloud sentences that are punctuated in various ways to show that punctuation may affect the way they would read the sentence. For example: "Mary, will you come here!" "Mary, will you come here?" "Mary, come here."	2. In the following selection, who is speaking? "John," said Phil, "where is Mary going?" ____John ____Phil ____Mary ____We don't know
3. Have students think of sentences in which the actual meaning is affected by punctuation marks. For example: "Kill Godzilla." "Kill, Godzilla." "Kill Godzilla?"	
4. Have students re-write a story containing dialogue as a play.	

Competency	Measurable Behavior (3rd Grade)	Measurable Behavior (6th Grade)	Measurable Behavior (9th Grade)
III. Inferential Comprehension	By the end of the third grade, the student will be able to: A. Infer the main idea of a selection.	By the end of the sixth grade, the student will be able to: A. Infer the main idea of a selection.	By the end of the ninth grade, the student will be able to: A. Infer the main idea of a selection.

Example Experiences and/or Activity

1. Have students read a selection in which the main idea is not explicitly stated, but is to be inferred. Go through the sentences contained in the selection and have the students discover for themselves that no one sentence alone states the main idea of the whole. Ask them to state as clearly and succinctly as possible what they think the main idea is.

2. Have students read a selection. Have them list as many ideas contained in the selection as possible. Then have them decide which of the ideas are more important to the whole selection, which less important. Have them select from the more important ideas the one they think best states the main idea. Then discuss the concept of "main idea."

3. Give students a sentence that can be used as a main idea. Ask them to write a selection about that idea without actually stating it in the selection. Have other students state what they think the main idea is.

4. As a matter of course when students read, ask them what they think the main idea was. Accustom them in various kinds of reading activities, both reading for pleasure and in instructional material, to distinguish between major (main) ideas and subordinate or minor ideas in that same selection.

Measurement

1. Have students read a selection in which the main idea is not actually stated. Have them choose from a list of possible main ideas the one that most clearly states the main idea of the selection.

2. Have students read a selection. Present them with a list of ideas to be inferred from, but not stated in, the selection, ranging from the more important ideas to the lesser ideas. Have them check the most important ideas and the least important to be inferred from the selection. Have students justify their choices.

5. Ask students to consider the question, "Is the main idea more often stated or inferred?" in regard to a variety of reading materials; i.e., stories, fables, science materials, social studies materials, newspapers, plays, novels, short stories, and so on. In what kinds of reading materials does one tend to find the main idea stated and in what kind of materials is it likely to be inferred? Why?

B. Infer the cause and effect relationships within a selection.

B. Infer the cause and effect relationships within a selection.

B. Infer the cause and effect relationships within a selection.

Example Experiences and/or Activity	Measurement
1. Have students read selections in which cause and effect relationships are not actually stated. In discussions and through individual work have students state the inferred cause and effect relationships. Have them justify the causes and effects they state. For example, "Joe is very good at carpentry. His father is a carpenter." Though not stated explicitly, one might justifiably infer that Joe learned something about carpentry from his father.	1. Give students a selection in which cause and effect relationships may be inferred. Ask them to identify the appropriate inferred causes and effects.
2. Have students locate inferred cause and effect relationships in their textbooks. Discuss the concept of multiple causes and multiple effects, especially in regard to science and the social sciences. In discussing stories, novels, and plays, make a point of asking students to discuss what they think caused the characters to act as they did and what effect these actions had on other characters. Having students discuss inferred causes of human behavior and inferred effects are as appropriate to discussing "Peter Rabbit" as to "Hamlet."	

C. Predict the probable outcome of a selection.

C. Predict the probable outcome of a selection.

C. Predict the probable outcome of a selection.

Example Experiences and/or Activity	Measurement
1. Have students read a selection from which the ending has been eliminated. Have them speculate how, on the basis of everything else in the story, it will in all probability end. Have them justify why they say the story will end that way. Then have them read the actual ending. 2. Have students read stories and trade books (library books) of their own choosing. Have them speculate what events may occur or what may happen to the characters in the years following the end of the story or novel. Have them justify their ideas. 3. Discuss the idea of "probable outcomes" in relation to such literary devices as surprise endings, ironic twists, unforeseeable outcomes, and so on. Assist them to understand the difference between "probable outcomes" and more creative and literary outcomes. Have students complete a story in the most probable way and then in a less predictable way. Have them discuss which outcome is better. Why?	1. Give the students a selection from which the ending has been eliminated. List some probable outcomes. On the basis of what the reader is told in the selection, which of the listed outcomes is the most probable?

D. Infer details that support the main idea of a selection.

D. Infer details that support the main idea of a selection.

D. Infer details that support the main idea of a selection.

Example Experiences and/or Activity	Measurement
1. Have students read a selection and then make inferences about details not explicitly stated in the selection. Have them list all inferences they can think of about details not stated. For example, if the story describes a gradually darkening, brilliant red sky that makes the sea "look blood red," the reader can infer the story occurs at sunset. Much of what we	1. Give the students a selection followed by a list of details about the selection that were not explicitly stated. Have the students choose the details that may be justifiably inferred.

read in a selection is inferred by the reader, rather than is explicitly stated--and appropriately so. But some inferences are more justifiable than others.

2. Have students make a list of details they want to include in a story, such as the day the story occurs will be an extremely hot one. It will occur in July in the 1860's, and the setting will be Pennsylvania. The main character will be a thirteen-year-old deaf girl who is the youngest of four children. And so on. Now have them write the story without explicitly stating these details. Then have other children read the selection and make inferences about details. Have them verify their inferences on the basis of the original list.

E. Infer the sequence within a selection.

E. Infer the sequence within a selection.

E. Infer the sequence within a selection.

Example Experiences and/or Activity

1. Have students read a selection in which sequentiality is inferred, though not stated specifically. Following their reading, have students list the elements in the story as they were presented and as they actually occurred. Include examples that show that the order of presentation within the selection may not be the order of the actual event. For example,in the following selection, the events are not presented in actual sequence: "I took the cake out of the oven and was so pleased, I decided to frost it with extra deluxe frosting. As I was making the frosting and then putting it on the cake, I thought back to the difficulty I had getting the batter just right." "Getting the batter right" is presented last, but actually occurred before anything else mentioned in the selection.

2. Have students do a "time line" on the basis of a story or book they have read. A section of a history book might be particularly appropriate for the activity.

Measurement

1. Have students read a selection in which sequentiality is to be inferred. Ask questions to determine if they understand the inferred sequentiality. For example, if the student were asked what occurred first in the cake-baking selection (see opposite), he/she should choose "...tried to get the batter right" and <u>not</u> "made the frosting."

3. Have students write a selection involving sequentiality. Have other students read the selection and list the elements sequentially.

F. Infer likenesses and differences within a selection.

F. Infer likenesses and differences within a selection.

F. Infer likenesses and differences within a selection.

Example Experiences and/or Activity

1. Present students with selections in which likenesses and differences are to be inferred rather than actually stated. Ask them to identify these likenesses and differences. For example, a selection may be about animals that rely on speed to escape their enemies. Two animals so discussed may be antelopes and rabbits. The student would infer that antelopes and rabbits are alike in that they both can run fast.

2. (See also identifying stated likenesses and differences.)

Measurement

1. Give the students a selection in which likenesses and differences are to be inferred. Have students identify these likenesses and differences.

G. Draw conclusions from given information.

G. Draw conclusions from given information.

G. Draw conclusions from given information.

Example Experiences and/or Activity

1. Present students with selections and passages on the basis of which conclusions may be drawn. Have them reach various conclusions as individuals or as members of groups. List the various conclusions drawn on the board and discuss which ones are the most justifiable. Discuss what constitutes "a safe conclusion."

2. Make conclusions on the basis of material presented in a wide range of written material, such as textbooks, poems, novels, stories, advertisements, news-

Measurement

1. Give the students a selection upon which a conclusion may be drawn. From a list of possible conclusions, have them select the most justifiable one. For example, if the selection states that dikes have been built around a city, we can conclude "that the city is located close to the sea," but not necessarily "that it is a city that dates back to the Middle Ages."

paper and magazine articles, research studies, and so on. What kinds of material are easiest and safest to draw conclusions on the basis of? Are some conclusions more justifiable than others? Why?

3. Make it a practice to ask the students "What do you think we can conclude from what you've read?" Encourage students to present various conclusions and to justify them.

H. Identify relationships of words (analogies).	H. Identify relationships of words (analogies).	H. Identify relationships of words (analogies).

Example Experiences and/or Activity	Measurement
1. Have students practice word analogies of various degrees of difficulty. For example, "Shoe is to foot as glove is to _____." Have students make up their own analogies to give other students. Leave various parts of the analogies blank. ("_____ is to foot as glove is to hand.") Use for vocabulary builders as well: "Mauve is to purple as gray is to _____." Student may have to look up "mauve."	1. Students choose from a list the appropriate word to complete an analogy.
2. Organize "spelling-bee" type games and other group games, using word analogies as the vehicle.	

III-I was inadvertently left out. It should read as follows:

I. Make inferences about characters in a story.	I. Make inferences about characters in a story.	I. Make inferences about characters in a story.

Competency	Measurable Behavior (3rd Grade)	Measurable Behavior (6th Grade)	Measurable Behavior (9th Grade)
IV. Critical Reading Skills	By the end of the third grade, the student will be able to:	By the end of the sixth grade, the student will be able to:	By the end of the ninth grade, the student will be able to:
	A. Determine the author's purpose for a selection.	A. Determine the author's purposes for a selection.	A. Determine the author's major purposes for a selection.

Example Experiences and/or Activity	Measurement
1. Have students read selections that have an obvious purpose, such as Aesop's Fables. Have students discuss the purpose of the selection. Encourage various ideas. Then have students decide the best statement of the purpose. At higher levels, discuss the author's purpose in terms of materials in which the purpose is not as clear cut or where there may be a number of purposes. Also, at the higher levels, have students discuss the author's purpose in regard to a wide variety of materials; i.e., fiction, non-fiction, expository writing, newspaper and magazine articles, advertisements, and so on.	1. Give students a brief selection and list four or more possible purposes. Have them choose the "best purpose" or the "main purpose." For example, the selection may be on the gradual decline of elephants because of hunters. The intended author's purpose is to "prevent the extinction of elephants."
2. Have students select "a purpose" for writing something, such as to tell a moral, to convince people they should give to Community Chest, to entertain, or to inform; and then write a selection based on that purpose. Have other students read the selections and guess the intended purposes.	2. Given a selection, the students will identify major purposes and possible minor purposes.
3. Have students read brief selections. List possible purposes on the board. Discuss why one particular purpose is the best choice.	

B. Distinguish between fantasy and reality.	B. Distinguish between fact and opinion.	B. Distinguish between fact and opinion.

Example Experiences and/or Activity	Measurement
1. Discuss with students how some stories are "real" (could actually happen), while others are fantasies	1. Under such phrases as "Which of the following could really happen?" "Which of the

(could not actually happen). Have them read selections and discuss which ones "could really happen" and which ones "are make believe." Have them discuss the reasons for distinguishing between fact and fantasy.

2. At the higher levels, discuss how stories may present "truths," even though they are not actaully true or real. Thus, although fables are not real, they present truths about life. Also, in much writing, fact and fiction seem to merge.

3. Give students reading material containing both fact and opinion. Ask them to identify the facts and the opinions and tell why they have identified these parts as such. At the upper levels, consider material in which fact and opinion are less distinguishable--for example, in cases where facts are arranged and presented so as to convey the author's opinion.

4. Have students locate examples of writing that contain facts and opinions, especially in newspapers and magazines.

following is make-believe?" "Which of the following could a person really do" list various choices and ask the student to choose the appropriate choice. For example, given "Which could not really happen?" the student would choose "The airplane laughed and laughed."

2. Give students various selections and have them decide if they are fact or fantasy.

3. Give the students a questions such as: "Which of the following are statements of fact?" and list choices. Students will select the factual statement. Do the same for opinion.

C. Determine the author's viewpoint from a selection.

C. Determine the author's viewpoint from a selection.

Example Experiences and/or Activity

1. Have the students read a selection and discuss what they believe the author's point of view to be; i.e., what opinion does the author have regarding the topic. For example, if the selection is on crime, does the author believe it is hopeless to do anything about it, everyone should try to do something, it's the governor's job, or it's the natural result of social ills.

2. Have students read a wide variety of material and discuss what they believe to be the author's viewpoint.

Measurement

1. Give students a brief selection and list four or five points of view on the topic. Have students select the point of view expressed by the author of the selection.

3. Have students read selections on the same topic, but written from various viewpoints.

4. Have students write on topics from various viewpoints. For example, have them write about the American Revolution for an English history book, a Canadian, a French, and a Russion textbook.

5. Discuss the topic of bias and point of view. Can any writing be free of bias or a point of view? Especially discuss the question in relation to the various subject areas: history, the social science, science, health education, literature, the arts, and so on.

D. Identify examples of propaganda techniques.

Example Experiences and/or Activity	Measurement
1. Discuss various propaganda techniques and have students read examples of these techniques. Have students find examples of their own in a variety of written material, including advertisements.	1. Give students a brief selection using a particular propaganda technique. Ask the student to identify the technique used.
2. List various types of propaganda techniques and have students write selections using these techniques. Have other students read the selections and discuss the techniques used.	2. Ask students to identify selections that are heavily propagandized and selections that are relatively free of propaganda.
3. Have students construct a montage art of sections of advertisements that use various propaganda techniques.	
4. Discuss the various uses of propaganda, both in contemporary society and from a historical perspective.	

Competency	Measurable Behavior (3rd Grade)	Measurable Behavior (6th Grade)	Measurable Behavior (9th Grade)
V. Related Study Skills	By the end of the third grade, the student will be able to:	By the end of the sixth grade, the student will be able to:	By the end of the ninth grade, the student will be able to:
	A. Identify the major use of dictionaries, tables of contents, and glossaries.	A. Identify the major uses of dictionaries, encyclopedias, atlases, newspapers, magazines, telephone books, tables of contents, glossaries, indexes, maps, graphs, charts, and tables.	A. Identify the major uses of dictionaries, encyclopedias, atlases, newspapers, magazines, telephone books, thesauruses, almanacs, card catalogues, periodical guides, tables of contents, glossaries, indexes, maps, graphs, charts, tables, appendixes, footnotes and bibliographies.

Example Experiences and/or Activity	Measurement
1. Give students instruction in the use of the various reference materials listed in the objectives. Discuss the various situations in which materials would be used and how they are used.	1. Give the student a type of information to be located. Have the student identify the appropriate reference material to locate the information.

	B. Locate information within reference materials using dictionaries, tables of contents and glossaries.	B. Locate information within reference materials using dictionaries, encyclopedias, atlases, newspapers, magazines, telephone books, tables of contents, glossaries, indexes, maps, graphs, charts, and tables.
		B. Locate information within reference materials using dictionaries, encyclopedias, atlases, newspapers, magazines, telephone books, thesauruses, almanacs, card catalogues, periodical guides, tables of contents, glossaries, indexes, maps, charts, graphs, tables, appendixes, footnotes, and bibliographies.

Example Experiences and/or Activity	Measurement
1. Have the students use the various reference materials listed in the objective in their everyday work, especially in a variety of subject areas.	1. Given a kind of information to locate, the student will locate the information in the appropriate reference material.

C. Follow written directions. C. Follow written directions. C. Follow written directions.

Example Experiences and/or Activity	Measurement
1. Give students various sets of directions for a wide variety of tasks, such as how to construct something, how to get from one place to another, and how to fill out a form, and ask them to follow the directions exactly.	1. Give the student a form with written directions. Ask the student to complete the form accurately.

D. Summarize a selection. D. Summarize a selection. D. Summarize a selection.

Example Experiences and/or Activity	Measurement
1. Have students read various kinds of selections and present them with summaries of the selections. Discuss which summaries are the best and why. 2. Have students write summaries of a variety of selections.	1. Give students a brief selection and four or five summaries. Have the student select the best summary.

E. Organize information in an outline form. E. Organize information in an outline form. E. Organize information in an outline form.

Example Experiences and/or Activity	Measurement
1. Have students at the lower levels construct rudimentary outlines of written material. At upper levels have them construct more complete outlines. Discuss various types of outlines, the logic behind outline forms, and the various uses of outlines. 2. Given a completed outline, have the student write a selection or give a speech using the outline. 3. Have students outline material before writing a selection.	1. On the basis of a set of material, the student will identify the best outline for a given purpose.

F. Alphabetize words correctly through the second letter.

F. Use alphabetizing skills to locate information in common references.

Example Experiences and/or Activity	Measurement
1. Have students use alphabetizing skills in locating information in reference materials.	1. Given a word, the student will choose from a list of words the one that would come next.
2. Give students lists of words and ask them to alphabetize them. Activity may be done individually or in groups, and may be carried out as a game.	2. Given a list of words, the student will alphabetize them.
3. Have students locate a group of words in a dictionary as rapidly as possible. Conduct as a race--the winners being those students who find all the words first.	3. Given two guide words (as are found on a dictionary page), the student will identify words that would fall between those words. For example, "mind" falls between "mill" and "minor," but "mock," "minority," and "mug" will not.
4. Stress the use of guide words in using dictionaries, telephone books, and so on.	

Competency	Measurable Behavior (3rd Grade)	Measurable Behavior (6th Grade)	Measurable Behavior (9th Grade)
VI. Positive Responses to Reading	By the end of the third grade, the student will demonstrate her/his enjoyment of reading by:	By the end of the sixth grade, the student will demonstrate her/his enjoyment of reading by:	By the end of the ninth grade, the student will demonstrate her/his enjoyment of reading by:
	A. Reading materials of her/his choice during free time, both in school and at home.	A. Reading materials of her/his choice during free time, both in school and at home.	A. Reading materials of her/his choice during free time, both in school and at home.

Example Experiences and/or Activity	Measurement
1. Allow time in school for students to read for their own pleasure. Students should be free to read the kind of materials they themselves select.	1. Given the opportunity to do so, students will freely select and read books, magazines, or whatever appeals to them. The observer will set his or her own objective. It may be: "Given the opportunity to do so, 90% of the students will read of their own choice for at least ___ minutes."
2. Have the students and/or their parents keep a log of what they (the students) are reading at home. Any kind of reading material should be considered allowable.	

B. Going frequently to places where reading materials are available, such as libraries, reading rooms, book sales, and book exchanges.	B. Going frequently to places where reading materials are available, such as libraries, reading rooms, book sales, and book exchanges.	B. Going frequently to places where reading materials are available, such as libraries, reading rooms, book sales, and book exchanges.

Example Experiences and/or Activity	Measurement
1. Provide time for students to go to the school library or other places where they can select reading materials.	1. The teacher's objective may be: "Given the opportunity to go to the library, 90% of the students will choose to go and select a book or some other reading material."
2. Especially allow individual students to go to the library as the need arises, or as they wish to do so.	

C. Requesting reading materials in addition to those assigned by the teacher.

C. Requesting reading materials in addition to those assigned by the teacher.

C. Requesting reading materials in addition to those assigned by the teacher.

Example Experiences and/or Activity	Measurement
1. Teachers can encourage students to ask for additional reading materials by making attractive, high interest materials readily available.	1. The teacher's objective might be: "Over the course of ___ weeks, 90% of the students will at least once ask for or seek out additional reading materials that are not 'required'."

D. Responding to the opportunity to talk about and/or discuss what he/she has read.

D. Responding to the opportunity to talk about and/or discuss what he/she has read.

D. Responding to the opportunity to talk about and/or discuss what he/she has read.

Example Experiences and/or Activity	Measurement
1. Give students ample opportunity to talk about what they are reading with other students or to adults. Conversations and discussions may be conducted class-wide or in small groups. Informal and open-ended discussions are particularly appropriate.	1. The classroom objective might read: "Given the opportunity to do so, 90% of the students will, during the course of a week, choose to talk with someone else about what they have read."

E. Taking part in creative activities related to reading such as puppet shows, dramatizations, creative dramatics, art/music activities, creative writing activities, investigative activities, and so on.

E. Taking part in creative activities related to reading such as puppet shows, dramatizations, creative dramatics, art/music activities, creative writing activities, investigative activities, and so on.

E. Taking part in creative activities related to reading such as puppet shows, dramatizations, creative dramatics, art/music activities, creative writing activities, investigative activities, and so on.

Example Experiences and/or Activity	Measurement
1. Give the students opportunities to relate their reading activities to a variety of creative activities. 2. See Speaking and Listening objectives (especially Creative Dramatics).	1. A classroom objective might read: "Given the opportunity to do so, 90% of the students will, sometime during the course of a three-week period of time, choose to take part in a creative activity related to reading."

APPENDIX C

Appendix C

Proportion scores of the reading instructional programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 4 as measured by the Reading Concepts Checklist, (RCC).

Category	P_1	P_2	P_3	P_4	P_5	T_4
Total Score (103 Concepts)	.8641	.8932	.8641	.6990	.6796	.2427
Auditory Discrimination (6 Concepts)	.6666	1.00	.3333	.8333	1.00	0.00
Visual Discrimination (4 Concepts)	.75	1.00	.75	.50	.75	0.00
Phonic Analysis (16 Concepts)	.9375	1.00	1.00	1.00	1.00	0.00
Structural Analysis (11 Concepts)	.8181	1.00	1.00	.9090	.7272	0.00
Comprehension: Vocabulary Development (6 Concepts	1.00	1.00	1.00	.8333	.3333	.6666

Appendix C Continued

Category	P_1	P_2	P_3	P_4	P_5	T_4
Literal Comprehension (13 Concepts)	1.00	.9231	1.00	.8462	.8462	.3077
Inferential Comprehension (17 Concepts)	.8824	.9412	.9412	.6471	.5294	.4707
Critical Comprehension (8 Concepts)	.875	.75	.875	.375	.50	.1250
Study Skills (22 Concepts)	.7727	.5909	.6818	.4090	.50	.2272

Key: P_1 = Ginn and Company

P_2 = Harcourt, Brace and Jovanovich

P_3 = Holt, Rinehart and Winston

P_4 = Houghton-Mifflin Company

P_5 = Scott Foresman Company

T_4 = Michigan Educational Assessment Program Experimental Reading Test Grade 4

APPENDIX D

Appendix D

Differences between proportion scores between the reading instructional programs and between the Michigan Educational Assessment Program Experimental Reading Test Grade 4 as measured by the Reading Concepts Checklist, (RCC).

Value	G-C	HBJ	HRW	HMC	SFC	T_4	Q	D.F.	P	H_o
Category: Total Score Difference in Proportions Between Instructional Programs and the Experimental Reading Test (103 Concepts)										
1	89	90	89	72	70	25	162.4435	5	p <.001	Reject
0	14	13	14	31	33	78				
Total Score Differences in Proportions Between Instructional Programs (103 Concepts)										
1	89	90	89	72	70		31.9685	4	p ≤.001	Reject
0	14	13	14	31	33					
Category I: Auditory Discrimination - Differences in Proportions Between Instructional Programs and the Experimental Reading Test (6 Concepts)										
1	4	6	2	5	6	0	18.4043	5	p<.001	Reject
0	2	0	4	1	0	6				
Auditory Discrimination - Differences in Proportions Between Instructional Programs (6 Concepts)										
1	4	6	2	5	6		9.33	4	p>.05	Not
0	2	0	4	1	0					Rejected

The null hypotheses are rejected at the 0.05 level. Higher levels are indicated. Con't.

Appendix D. Continued

Value	G-C	HBJ	HRW	HMC	SFC	T_4	Q	D.F.	P	H_o
	Category II: Visual Discrimination - Differences in Proportions Between Instructional Programs and the Experimental Reading Test (4 Concepts)									
1	3	4	3	2	3	0	12.3913	5	p <.05	Reject
0	1	0	1	2	1	4				
	Visual Discrimination - Differences in Proportions Between Instructional Programs (4 Concepts)									
1	3	4	3	2	3		5.00	4	p > .05	Not
0	1	0	1	2	1					Rejected
	Category III: Phonic Analysis - Differences in Proportions Between Instructional Programs and the Experimental Reading Test (16 Concepts)									
1	15	16	16	16	16	0	75.4819	5	p < .001	Reject
0	1	0	0	0	0	16				
	Phonic Analysis - Differences in Proportions Between Instructional Programs (16 Concepts)									
1	15	16	16	16	16		4.00	4	p > .05	Not
0	1	0	0	0	0					Rejected

The null hypotheses are rejected at the 0.05 level. Higher levels are indicated. Con't.

Appendix D. Continued

Value	G-C	HBJ	HRW	HMC	SFC	T_4	Q	D.F.	P	H_o
	Category IV: Structural Analysis - Differences in Proportions Between Instructional Programs and the Experimental Reading Test (11 Concepts)									
1	8	10	10	9	8	2	27.4490	5	$p < .001$	Reject
0	3	1	1	2	3	9				
	Structural Analysis - Differences in Proportions Between Instructional Programs (11 Concepts)									
1	8	10	10	9	8		5.7143	4	$p > .05$	Not
0	3	1	1	2	3					Rejected
	Category V: Comprehension: Vocabulary Development - Differences in Proportaions Between Instructional Programs and the Experimental Reading Test (6 Concepts)									
1	6	6	6	5	2	4	16.1111	5	$p < .01$	Reject
0	0	0	0	1	4	2				
	Comprehension: Vocabulary Development - Differences in Proporations Between Instructioanl Programs (6 Concepts)									
1	6	6	6	5	2		15.6667	4	$p < .01$	Reject
0	0	0	0	1	4					

The null hypotheses are rejected at the 0.05 level. Higher levels are indicated. Con't.

Appendix D. Continued.

Value	G-C	HBJ	HRW	HMC	SFC	T_4	Q	D.F.	P	H_o
	Category VI: Literal Comprehension - Differences in Proportions Between Instructional Programs and the Experimental Reading Tests (13 Concepts)									
1	13	12	13	11	11	4	30.7143	5	$p < .001$	Reject
0	0	1	0	2	2	9				
	Literal Comprehension - Differences in Proportions Between Instructional Programs (13 Concepts)									
1	13	12	13	11	11		5.7143	4	$p > .05$	Not
0	0	1	0	2	2					Rejected
	Category VIII: Inferential Comprehension - Differences in Proportions Between Instructional Programs and the Experimental Reading Test (17 Concepts)									
1	15	16	16	11	9	8	19.4554	5	$p < .01$	Reject
0	2	1	1	6	8	9				
	Inferential Comprehension - Differences in Proportions Between Instructional Programs (17 Concepts)									
1	15	16	16	11	9		13.2903	4	$p < .01$	Reject
0	2	1	1	6	8					

The null hypotheses are rejected at the 0.05 level. Higher levels are indicated. Con't.

Appendix D. Continued

Value	G-C	HBJ	HRW	HMC	SFC	T_4	Q	D.F.	P	H_o
	Category VIII: Critical Comprehension - Differences in Proportions Between Instructional Programs and the Experimental Reading Test (8 Concepts)									
1	7	6	7	3	4	1	15.7143	5	p<.01	Reject
0	1	2	1	5	4	7				
	Critical Comprehension - Differences in Proportions Between Instructional Programs (8 Concepts)									
1	7	6	7	3	4		8.80	4	p>.05	Not
0	1	2	1	5	4					Rejected
	Category IX: Study Skills - Differences in Proportions Between Instructional Programs and the Experimenatl Reading Test (22 Concepts)									
1	17	13	15	9	11	5	22.5806	5	p<.001	Reject
0	5	9	7	13	11	17				
	Study Skills - Differences in Proportions Between Instructional Programs (22 Concepts)									
1	17	13	15	9	11		10.8108	4	p<.05	Reject
0	5	9	7	13	11					

The null hypotheses are rejected at the 0.05 level. Higher Levels are indicated.

KEY: G-C - Ginn and Company
HBJ - Harcourt Brace and Jovanovich
HRW - Holt, Rinehart, and Winston
HMC - Houghton-Mifflin Company
SFC - Scott Foresman Company
T_4 - Michigan Educational Assessment Program Experimental Reading Test Grade Four

APPENDIX E

Appendix E

Summary of the values of the pairwise comparions of the means of the proportion scores between the K-3 reading instructional programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 4 by individual category scores within the Reading Concepts Checklist, (RCC).

Program	Experimental Reading Test	ψ
Category: Auditory Discrimination		
Ginn and Company	.6666	±.2258
Harcourt, Brace and Jovanovich	1.00	
Holt, Rinehart, and Winston	.3333	
Houghton-Mifflin Company	.8333	
Scott, Foresman Company	1.00	
Category: Visual Distrimination		
Ginn and Company	.750	±.2868
Harcourt, Brace and Jovanovich	1.00	
Holt, Rinehart, and Winston	.750	
Houghton-Mifflin Company	.500	
Scott, Foresman Company	.750	

Continued

Appendix E. Continued

Program	Experimental Reading Test	ψ
Category: Phonic Analysis		
Ginn and Company	.9375	±.2772
Harcourt, Brace and Jovanovich	1.00	
Holt, Rinehart, and Winston	1.00	
Houghton-Mifflin Company	1.00	
Scott, Foresman Company	1.00	
Category: Structural Analysis		
Ginn and Company	.5454	±.1017
Harcourt, Brace and Jovanovich	.7273	
Holt, Rinehart, and Winston	.7273	
Houghton-Mifflin Company	.6363	
Scott, Foresman Company	.4545	

Continued

Appendix E Continued

Program	Experimental Reading Test	ψ
Category: Comprehension - Vocabulary Development		
Ginn and Company	.3334	±.1470
Harcourt, Brace and Jovanovich	.3334	
Holt, Rinehart, and Winston	.3334	
Houghton-Mifflin Company	.1667	
Scott, Foresman Company	-.3333	
Category: Literal Comprehension		
Ginn and Company	.6923	±.0882
Harcourt, Brace and Jovanovich	.6154	
Holt, Rinehart, and Winston	.6923	
Houghton-Mifflin Company	.5385	
Scott, Foresman Company	.5385	

Continued

Appendix E. Continued

Program	Experimental Reading Test	ψ
Category: Inferential Comprehension		
Ginn and Company	.4118	±.1017
Harcourt, Brace and Jovanovich	.4706	
Holt, Rinehart, and Winston	.4706	
Houghton-Mifflin Company	.1765	
Scott, Foresman Company	.0588[a]	
Category: Critical Comprehension		
Ginn and Company	.7500	±.2037
Harcourt, Brace and Jovanovich	.6250	
Holt, Rinehart, and Winston	.7500	
Houghton-Mifflin Company	.2500	
Scott, Foresman Company	.3750	

[a]Non-Significant Difference

Continued

Appendix E. Continued

Program	Experimental Reading Test	ψ
Category: Study Skills		
Ginn and Company	.5455	±.1211
Harcourt, Brace and Jovanovich	.3637	
Holt, Rinehart, and Winston	.4546	
Houghton-Mifflin Company	.1818	
Scott, Foresman Company	.2728	

APPENDIX F

Appendix F

Proportion scores of the reading instructional programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 7 as measured by the Reading Concepts Checklist, (RCC).

Category	P_1	P_2	P_3	P_4	P_5	T_7
Total Score (103 Concepts)	.8252	.8252	.8447	.5147	.6505	.2718
Auditory Discrimination (6 Concepts)	.00	.00	.00	.00	.00	.00
Visual Discrimination (4 Concepts)	.00	.00	.00	.00	.00	.00
Phonic Analysis (16 Concepts)	.8152	.9375	1.00	.0625	.9375	.00
Structural Analysis (11 Concepts)	1.00	1.00	1.00	.6364	.7273	.1818
Comprehension Vocabulary Development (13 Concepts)	.8333	1.00	.8333	.6777	.500	.6777

Continued

Appendix F Continued

Category	P_1	P_2	P_3	P_4	P_5	T_7
Literal Comprehension (13 Concepts)	.9231	.8462	.9231	.4615	.6154	.3077
Inferential Comprehension (17 Concepts)	.9412	.9412	1.00	.7647	.5882	.4118
Critical Comprehension (8 Concepts)	.750	1.00	.8750	.6250	.6250	.3750
Study Skills (22 Concepts)	1.00	.8182	.8636	.7727	.7273	.3636

Key: P_1 = Ginn and Company

P_2 = Harcourt, Brace and Jovanovich

P_3 = Holt, Rinehart, and Winston

P_4 = Houghton-Mifflin Company

P_5 = Scott, Foresman Company

T_7 = Michigan Educational Assessment Program Experimental Reading Test Grade 7

APPENDIX G

Appendix G

Differences between proportion scores between the reading instructional programs and between the Michigan Educational Assessment Program Experimental Reading Test Grade 7 as measured by the Reading Concepts Checklist, (RCC).

Value	G-C	HBJ	HRW	HMC	SFC	T_7	Q	D.F.	P	H_o
Category: Total Score Difference in Proportions Between Instructional Programs and the Experimental Reading Test (103 Concepts)										
1	85	85	87	53	67	28	153.2440	5	p < .001	Reject
0	18	18	16	50	36	75				
Total Score Difference in Proportions Between Instructional Programs										
1	85	85	87	53	67		64.5797	4	p < .001	Reject
0	18	18	16	50	36					
Category I: Auditory Discrimination - Differences in Proportions Between Instructional Programs and the Experimental Reading Test Grade 7 (6 Concepts)										
Not Considered: Neither Taught nor Tested.										
Category II: Visual Discrimination - Differences in Proportions Between Instructional Programs and the Experimental Reading Test Grade 7 (4 Concepts)										
Not Considered: Neither Taught nor Tested.										

The null hypotheses are rejected at the 0.05 level. Higher levels are indicated. Con't

Appendix G. Continued

Value	G-C	HBJ	HRW	HMC	SFC	T_7	Q	D.F.	P	H_o
	Category III: Phonic Analysis - Differences in Proportions Between Instructional Programs and the Experimental Reading Test Grade 7 (16 Concepts)									
1	13	15	16	1	15	0	63.6923	5	p < .001	Reject
0	3	1	0	15	1	16				
	Phonic Analysis - Differences in Proportions Between Instructional Programs (16 Concepts)									
1	13	15	16	1	15		44.5714	4	p < .001	Reject
0	3	1	0	15	1					
	Category IV: Structural Analysis - Differences in Proportions Between Instructional Programs and the Experimental Reading Test Grade 7 (11 Concepts)									
1	11	11	11	7	8	2	29.5875	5	p < .001	Reject
0	0	0	0	4	3	9				
	Structural Analysis - Differences in Proportions Between Instructional Programs (11 Concepts)									
1	11	11	11	7	8		12.6667	4	p < .05	Reject
0	0	0	0	4	3					

The null hypotheses are rejected at the 0.05 level. Higher levels are indicated. Con't

Appendix G. Continued

Value	G-C	HBJ	HRW	HMC	SFC	T_7	Q	D.F.	P	H_o
Category V: Comprehension: Vocabulary Development - Differences in Proportions Between Instructional Programs and the Experimental Reading Test Grade 7 (6 Concepts)										
1	5	6	5	4	3	4	7.1739	5	p > .05	Not
0	1	0	1	2	3	2				Rejected
Comprehension: Vocabulary Development - Differences in Proportions Between Instructional Programs (6 Concepts)										
1	5	6	5	4	3		6.5000	4	p > .05	Not
0	1	0	1	2	3					Rejected
Category VI: Literal Comprehension - Differences in Proportions Between Instructional Programs and the Experimental Reading Test (13 Concepts)										
1	12	11	12	6	10	4	23.3562	5	p < .05	Reject
0	1	2	1	7	3	9				
Literal Comprehension - Differences in Proportions Between Instructional Programs (13 Concepts)										
1	12	11	12	6	10		12.4000	4	p < .05	Reject
0	1	2	1	7	3					

The null hypotheses are rejected at the 0.05 level. Higher levels are indicated. Con't

Appendix G. Continued

Value	G-C	HBJ	HRW	HMC	SFC	T_7	Q	D.F.	P	H_o
	Category VII: Inferential Comprehension - Differences in Proportions Between Instructional Programs and the Experimental Reading Test (17 Concepts)									
1	16	16	17	13	10	7	25.4301	5	p < .001	Reject
0	1	1	0	4	7	10				
	Inferential Comprehension - Differences in Proportions Between Instructional Programs (17 Concepts)									
1	16	16	17	13	10		13.8333	4	p < .01	Reject
0	1	1	0	4	7					
	Category VIII: Critical Comprehension - Differences in Proportions Between Instructional Programs and the Experimental Reading Test (8 Concepts)									
1	6	8	7	5	5	3	11.5000	5	p < .05	Reject
0	2	0	1	3	3	5				
	Critical Comprehension - Differences in Proportions Between Instructional Programs (8 Concepts)									
1	6	8	7	5	5		6.8000	4	p > .05	Not
0	2	0	1	3	3					Rejected

The null hypotheses are rejected at the 0.05 level. Higher levels are indicated. Con't

Appendix G. Continued

Value	G-C	HBJ	HRW	HMC	SFC	T_7	Q	D.F.	P	H_o
	Category IX: Study Skills - Differences in Proportions Between Instructional Programs and the Experimental Reading Test (22 Concepts)									
1	22	18	19	17	16	8	28.3051	5	p < .01	Reject
0	0	4	3	5	6	14				
	Study Skills - Differences in Proportions Between Instructional Programs (22 Concepts)									
1	22	18	19	17	16		7.3103	4	p > .05	Not
0	0	4	3	5	6					Rejected

The null hypotheses are rejected at the 0.05 level. Higher Levels are indicated.

Key: G-C = Ginn and Company
HBJ = Harcourt, Brace and Jovanovich
HRW = Holt, Rinehart, and Winston
HMC = Houghton-Mifflin Company
SFC = Scott, Foresman Company
T_7 = Michigan Educational Assessment Program Experimental Reading Test Grade 7

APPENDIX H

Appendix H

Summary of the values of the pairwise comparisons of the means of the proportion scores between the 4-6 reading instructional programs and the Michigan Educational Assessment Program Experimental Reading Test Grade 7 by individual category scores within the Reading Concepts Checklist, (RCC).

Program	Experimental Reading Test	ψ
Category: Auditory Discrimination		
Not Considered: Neither Taught nor Tested.		
Category: Visual Discrimination		
Not Considered: Neither Taught nor Tested.		
Category: Phonic Analysis		
Ginn and Company	.8125	±.1441
Harcourt, Brace and Jovanovich	.9375	
Holt, Rinehart, and Winston	1.00	
Houghton-Mifflin Company	.0625[a]	
Scott, Foresman Company	.9375	

[a]Non-Significant Statistical Difference.

Continued

Appendix H. Continued

Program	Experimental Reading Test	ψ
Category: Structural Analysis		
Ginn and Company	.8182	±.1247
Harcourt, Brace and Jovanovich	.8182	
Holt, Rinehart, and Winston	.8182	
Houghton-Mifflin Company	.4546	
Scott, Foresman Company	.5455	
Category: Comprehension - Vocabulary Development		
Ginn and Company	.1666	±.1347
Harcourt, Brace and Jovanovich	.3333	
Holt, Rinehart, and Winston	.1666	
Houghton-Mifflin Company	0.00[a]	
Scott, Foresman Company	.1667	

[a]Non-Significant Statistical Difference.

Continued

Appendix H. Continued

Program	Experimental Reading Test	ψ
Category: Literal Comprehension		
Ginn and Company	.6154	±.1176
Harcourt, Brace and Jovanovich	.5385	
Holt, Rinehart, and Winston	.6154	
Houghton-Mifflin Company	.1538	
Scott, Foresman Company	.3077	
Category: Inferential Comprehension		
Ginn and Company	.5294	±.0929
Harcourt, Brace and Jovanovich	.5294	
Holt, Rinehart, and Winston	.5882	
Houghton-Mifflin Company	.3529	
Scott, Foresman Company	.1764	

Continuted

Appendix H. Continued

Program	Experimental Reading Test	ψ
Category: Critical Comprehension		
Ginn and Company	.3750	±.1411
Harcourt, Brace and Jovanovich	.6250	
Holt, Rinehart, and Winston	.5000	
Houghton-Mifflin Company	.2500	
Scott, Foresman Company	.2500	
Category: Study Skills		
Ginn and Company	.6364	±.0779
Harcourt, Brace and Jovanovich	.4546	
Holt, Rinehart, and Winston	.5000	
Houghton-Mifflin Company	.4019	
Scott, Foresman Company	.3637	

BIBLIOGRAPHY

BIBLIOGRAPHY

Aaron, Ira E. and Carter, Sylvia, Step Right Up, Glenview, Illinois: Scott, Foresman and Company, 1978.

Aaron, Ira E.; Davis, Charles and Schelly, Joan, Flying Hoofs, Glenview, Illinois: Scott, Foresman and Company, 1978.

Aaron, Ira E.; Jackson, Dauris; Riggs, Carole; Smith, Richard G. and Tierney, Robert, Racing Stripes, Glenview, Illinois: Scott, Foresman and Company, 1978.

Aaron, Ira E. and Koke, Rena, Ride A Rainbow, Glenview, Illinois: Scott, Foresman and Company, 1978.

Airasian, Peter W., "The Role of Evaluation in Mastery Learning," in Mastery Learning Theory and Practice, James H. Block, ed. New York: Holt, Rinehart and Winston, Inc., 1971.

American Psychological Association, Standards for Educational and Psychological Tests, Washington, D. C.: American Psychological Association, 1974.

American Psychological Association, Inc., Technical Recommendations for Psychological Test and Diagnostic Techniques. Washington, D. C.: APA 1954 in Ebel, Robert L. Essentials of Educational Measurement. 2nd. ed., Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972.

Barbe, Walter B., Personalized Reading Instruction, 9th Printing, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1967.

Berk, Ronald A., A Consumers' Guide to Criterion-Referenced Test Item Statistics, Paper Presented at the Annual Meeting of the National Council on Measurement in Education (Toronto, Ontario, Canada), 1978.

Botko, Louise Gorgos; Kerbs, JoAnn; Manning, John and Klassen, Verla Krocker, Dragon Wings, Glenview, Illinois: Scott, Foresman and Company, 1978.

Cairns, Joanna; Galloway, Elizabeth and Tierney, Robert, Daisy Days, Glenview, Illinois: Scott, Foresman and Company, 1978.

Cairns, Joanna; Galloway, Elizabeth and Tierney, Robert, Hootenanny, Glenview, Illinois: Scott, Foresman and Company, 1978.

Clymer, Theodore and Barrett, Thomas C., A Pocketfull of Sunshine, Lexington, Mass.: Ginn and Company, 1979.

Clymer, Theodore and Bissett, Donald J., A Lizard to Start With, Lexington, Mass.: Ginn and Company, 1979.

Clymer, Theodore; Bissett, Donald J. and Wulfing, Gretchen, Inside Out, Lexington, Mass.: Ginn and Company, 1979.

Clymer, Theodore and Christenson, Bernice M., Ready for Rainbows, Lexington, Mass.: Ginn and Company, 1979.

Clymer, Theodore and Fenn, Priscilla Holton, How It Is Nowadays, Lexington, Mass.: Ginn and Company, 1979.

Clymer, Theodore and Martin, Patricia Miles, The Dog Next Door, Lexington, Mass.: Ginn and Company, 1979.

Clymer, Theodore; Martin, Patricia Miles and Gates, Doris, May I Come In, Lexington, Mass.: Ginn and Company, 1979.

Clymer, Theodore; McCracken, Blair and McCullough, Constance M., Measure Me, Sky, Lexington, Mass.: Ginn and Company, 1979.

Clymer, Theodore; Parr, Billie; Gates, Doris and Robinson, Eleanor G., A Duck is a Duck, Lexington, Mass.: Ginn and Company, 1979.

Clymer, Theodore; Parr, Billie; Gates, Doris and Robinson, Eleanor G., Helicopters and Gingerbread, Lexington, Mass.: Ginn and Company, 1979.

Clymer, Theodore; Stein, Ruth Meyerson; Gates, Doris and McCullough, Constance M., Tell Me How the Sun Rose, Lexington, Mass.: Ginn and Company, 1979.

Clymer, Theodore; Wong, Olive and Benedict, Virginia Jones, One to Grow On, Lexington, Mass.: Ginn and Company, 1979.

Cohen, S. Alan and Hyman, Joan S., Instructional Objectives in Reading, New York: Random House, Inc., 1977.

Crambert, Albert C., Estimation of Validity for Criterion-Referenced Tests, Paper Presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York), 1977.

Cronbach, Lee J., Educational Psychology, 2nd ed., New York: Harcourt, Brace and World, Inc., 1963.

Duffy, Gerald G. and Sherman, George B., Systematic Reading Instruction, 2nd. ed., New York: Harper and Row, 1977.

Durr, William K.; LePere, Jean M. and Alsin, Mary Lou, Footprints, Boston, Mass.: Houghton-Mifflin Company, 1979.

Durr, William K.; LePere, Jean M. and Alsin, Mary Lou, Rockets, Boston, Mass.: Houghton-Mifflin Company, 1979.

Durr, William K.; LePere, Jean M. and Alsin, Mary Lou, Surprises, Boston, Mass.: Houghton-Mifflin Company, 1979.

Durr, William K.; LePere, Jean M.; Alsin, Mary Lou; Bunyan, Ruth Patterson and Shaw, Susan, Cloverleaf, Boston, Mass.: Houghton-Mifflin Company, 1979.

Durr, William K.; LePere, Jean M.; Alsin, Mary Lou; Bunyan, Ruth Patterson and Shaw, Susan, Honeycomb, Boston, Mass.: Houghton-Mifflin Company, 1979.

Durr, William K.; LePere, Jean M. and Brown, Ruth Hayek, Passports, Boston, Mass.: Houghton-Mifflin Company, 1979.

Durr, William K.; LePere, Jean M. and Brown, Ruth Hayek, Windchimes, Boston, Mass.: Houghton-Mifflin Company, 1979.

Durr, William K.; LePere, Jean M.; Niehaus, Bess and York, Barbara, Sunburst, Boston, Mass.: Houghton-Mifflin Company, 1979.

Durr, William K.; LePere, Jean M.; Niehaus, Bess and York, Barbara, Tapestry, Boston, Mass.: Houghton-Mifflin, Company, 1979.

Durr, William K.; Windley, Vivian O. and Earnhardt, Kay S., Impressions, Boston, Mass.: Houghton-Mifflin Company, 1979.

Durr, William K.; Windley, Vivian O. and Yates, Mildred C., Keystone, Boston, Mass.: Houghton-Mifflin Company, 1979.

Durr, William K.; Windley, Vivian O. and McCourt, Anne A., Medley, Boston, Mass.: Houghton-Mifflin Company, 1979.

Early, Margaret, Look, Listen, and Learn, New York: Harcourt Brace and Jovanovich, 1979.

Early, Margaret; Canfield, Robert; Karlin, Robert and Schottman, Thomas A., Building Bridges, New York: Harcourt Brace and Jovanovich, 1979.

Early, Margaret; Canfield, Robert; Karlin, Robert and Schottman, Thomas A., Moving Forward, New York: Harcourt Brace and Jovanovich, 1979.

Early, Margaret; Canfield, Robert; Karlin, Robert and Schottman, Thomas A., Reaching Out, New York: Harcourt Brace and Jovanovich, 1979.

Early, Margaret; Cooper, Elizabeth K. and Santeusanio, Nancy, Happy Morning Magic Afternoon and Reading Skills 2/3, New York: Harcourt Brace and Jovanovich, 1979.

Early, Margaret; Cooper, Elizabeth K. and Santeusanio, Nancy, People and Places and Reading Skills 7, New York: Harcourt Brace and Jovanovich, 1979.

Early, Margaret; Cooper, Elizabeth K. and Santeusanio, Nancy, Ring Around the World and Reading Skills 9, New York: Harcourt Brace and Jovanovich, 1979.

Early, Margaret; Cooper, Elizabeth K. and Santeusanio, Nancy, Sun and Shadow and Reading Skills 4, New York: Harcourt Brace and Jovanovich, 1979.

Early, Margaret; Cooper, Elizabeth K. and Santeusanio, Nancy, Sun Up and Reading Skills, New York: Harcourt Brace and Jovanovich, 1979.

Early, Margaret; Cooper, Elizabeth K. and Santeusanio, Nancy, Together We Go and Reading Skills 5, New York: Harcourt Brace and Jovanovich, 1979.

Early, Margaret; Cooper, Elizabeth K. and Santeusanio, Nancy, Widening Circles and Reading Skills 8, New York: Harcourt Brace and Jovanovich, 1979.

Early, Margaret; Cooper, Elizabeth K. and Santeusanio, Nancy, World of Surprises and Reading Skills 6, New York: Harcourt Brace and Jovanovich, 1979.

Ebel, Robert L., Essentials of Educational Measurement, 2nd ed., Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972.

Ebel, Robert L., Essentials of Educational Measurement, 3rd ed., Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1979.

Ebel, Robert L., "The Case for Minimum Competency Testing," Phi Delta Kappan (April, 1978).

Edmonston, Leon P. and Randall, Robert S., A Model for Estimating the Reliability and Validity of Criterion-Referenced Measures, Paper Presented at the Annual Meeting of the American Educational Research Association 56th, Chicago, Illinois) 1972.

Ekwall, Eldon E., Diagnosis and Remediation of the Disabled Reader, 2nd. Printing, Boston, Mass.: Allyn and Bacon, Inc., 1976.

Emrick, John A., The Experimental Validation of an Evaluation Model for Mastery Testing, Final Report, Office of Education, Washington, D. C., November, 1971.

Estes, Gary; Colvin, Lloyd W. and Goodwin, Coleen, A Criterion-Referenced Basic Skills Assessment Program in a Large City School System, Paper Presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California), 1976.

Evertts, Eldonna L. and Weiss, Bernard J., Never Give Up, New York: Holt, Rinehart and Winston, 1977.

Evertts, Eldonna L. and Weiss, Bernard J., People Need People, New York: Holt, Rinehart and Winston, 1977.

Evertts, Eldonna L. and Weiss, Bernard J., Special Happenings, New York: Holt, Rinehart and Winston, 1977.

Evertts, Eldonna L. and Weiss, Bernard J., The Way of the World, New York: Holt, Rinehart and Winston, 1977.

Evertts, Eldonna L.; Weiss, Bernard J. and Cruikshank, Susan B., A Place For Me, New York: Holt, Rinehart and Winston, 1977.

Evertts, Eldonna L.; Weiss, Bernard J. and Cruikshank, Susan B., A Time for Friends, New York: Holt, Rinehart and Winston, 1977.

Evertts, Eldonna L.; Weiss, Bernard J. and Cruikshank, Susan B., Books and Games, New York: Holt, Rinehart and Winston, 1977.

Evertts, Eldonna L.; Weiss, Bernard J. and Cruikshank, Susan B., Can You Imagine, New York: Holt, Rinehart and Winston, 1977.

Evertts, Eldonna L.; Weiss, Bernard J. and Cruikshank, Susan B., Hear, Say, See, Write, New York: Holt, Rinehart and Winston, 1977.

Evertts, Eldonna L.; Weiss, Bernard J. and Cruikshank, Susan B., Pets and People, New York: Holt, Rinehart and Winston, 1977.

Everetts, Eldonna L.; Weiss, Bernard J. and Cruikshank, Susan B., Rhymes and Tales, New York: Holt, Rinehart and Winston, 1977.

Farr, Roger, Reading: What Can Be Measured? Newark, Delaware: International Reading Association, 1969.

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

Gavin, Anne T., Guide to the Development of Written Tests for Selection and Promotion: The Content Validity Model. Technical Memorandum 77-6, Civil Service Commission, Washington, D.C.: Personnel Measurement Research and Development Center, 1977.

Glasser, R., and Nitko, A. J., Measurement in Learning and Instruction. In R. L. Thorndike ed. Educational Measurement, Washington: American Council on Education, 1971. In Ronald K. Hambleton and William P. Gorth, Criterion-Referenced Testing: Issues and Applications, Paper Presented at the Annual Meeting of the North-eastern Educational Research Association (Liberty, New York), 1970.

Haladyna, Tom and Roid, Gale, A Theoretical and Empirical Comparison of Three Approaches to Achievement Testing, (New York: ERIC Document Reproduction Service, Education 148903, May, 1978).

Hambleton, Ronald K. and Norick, M. R., "Toward an Integration of Theory and Method for Criterion-Referenced Tests," Journal of Educational Measurement. In Sherry Ann Rubinstein and Paula Nassif-Royer, The Outcomes of Statewide Assessment: Implications for Curriculum Evaluation, Paper Presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York), 1977.

Hays, William L., Statistics for the Social Sciences, 2nd ed., New York: Holt, Rinehart and Winston, Inc., 1973.

Jackson, Dauris, Jumping Jamboree, Glenview, Illinois: Scott, Foresman and Company, 1978.

Jackson, Dauris, No Cages Please, Glenview, Illinois: Scott, Foresman and Company, 1978.

Jackson, Dauris, Puppy Paws, Glenview, Illinois: Scott, Foresman and Company, 1978.

Jenkins, Joseph R. and Pany, Darlene, "Curriculum Biases in Reading Achievement Tests," Journal of Reading Behavior, Vol. X, No. 4, (Winter, 1978).

Jennings, Robert E. and Prince, Dorothy E., Calico Caper, Glenview, Illinois: Scott, Foresman and Company, 1978.

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

Kelley, Truman, "The Selection of Upper and Lower Groups for the Validation of Test Items," Journal of Educational Psychology, Vol. 30, (1939), in Robert L. Ebel, Essentials of Educational Measurement, 2nd ed., Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972.

Lennon, Roger T., "Assumptions Underlying the Use of Content Validity," In Readings in Measurement and Evaluation in Education and Psychology, Edited by William A. Mehrens, New York: Holt, Rinehart, and Winston, 1976.

Lewis, Juanita; Harrison, M. Lucile; Durr, William K. and McKee, Paul, Getting Ready to Read, Boston, Mass.: Houghton-Mifflin Company, 1979.

Linehart, Marsha M., Content Validity in Behavioral Assessment, Paper Presented at the Annual Meeting of the American Psychological Association (84th, Washington, D.C.), 1976.

Magnusson, David, Test Theory, Trans. By Hunter, Mabon, Reading, Mass.: Addison-Wesley Publishing Company, 1967.

Marascuilo, Leonard A. and McSweeney, Maryellen, Non-parametric and Distribution-Free Methods for the Social Sciences, Monterey, California: Brooks/ Cole Publishing Company, 1977.

Market Data Retrieval, Inc., HM Co. Market Research Report No. 17, Reading K-8 Survey, New York: Market Data Retrieval, Inc., 1977.

McCormick, Dean Richard, "The Controversial Development of the Michigan Educational Assessment Program 1969-1977" (unpublished Ph.D. dissertation, Michigan State University, 1978).

Mehrens, William A., Technical Report: The Fifth Report of the 1973-74 Michigan Educational Assessment Program. Michigan State Department of Education, Lansing, Michigan 1975.

Mehrens, William A. and Ebel, Robert L., 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.

Mehrens, William A. and Lehmann, Irvin J., Measurement and Evaluation in Education and Psychology, 2nd ed., New York: Holt, Rinehart and Winston, 1978.

Michigan Department of Education, Michigan Accountability 1967-77, Lansing, Michigan: Michigan Department of Education, undated.

Michigan Educational Assessment Program, First Report of the 1977-78 Michigan Educational Assessment Program, Interpretive Manual, Lansing, Michigan, 1978.

Michigan Educational Assessment Program, Technical Report, Lansing, Michigan: Michigan Department of Education, 1977.

Nunnally, Jum C., Educational Measurement and Evaluation, 2nd ed., New York: McGraw Hill Book Company, 1972.

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

Popham, W. James and Husek, T. R., "Implications of Criterion-Referenced Measurement," Journal of Educational Measurement, 1969. In Ronald K. Hambleton and William P. Gorth, Criterion-Referenced Testing: Issues and Implications, Paper Presented at the Annual Meeting of the Northeastern Educational Research Association (Liberty, New York), 1970.

Reid, Ethna R., Teaching Literal and Inferential Comprehension, Salt Lake City, Utah: Cove Publishers, 1978.

Riggs, Carole, First Feathers, Glenview, Illinois: Scott, Foresman and Company, 1978.

Riggs, Carole, Hello, Sunshine, Glenview, Illinois: Scott, Foresman and Company, 1978.

Ross, C. C. and Stanley, Julian C., Measurement in Today's Schools, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1954 in Ebel, Robert L., Essentials of Educational Measurement, 2nd ed., Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1972.

Roudabush, Glen E., Item Selection for Criterion-Referenced Tests. Paper Presented at the Annual Meeting of the American Educational Research Association, (57th, New Orleans, La.) 1973.

Rubinstein, Sherry Ann and Nassif-Royer, Paula, The Outcomes of Statewide Assessment: Implications for Curriculum Evaluation, Paper Presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York), 1977.

Smith, Douglas A., The Effects of Various Item Selection Methods on the Classification Accuracy and Classification Consistency of Criterion-Referenced Instruments, Paper Presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada), 1978.

Smith, Richard G. and Tierney, Fins and Tales, Glenview, Illinois: Scott, Foresman and Company, 1978.

Spool, Mark D., Performing a Content Validity Study, Paper Presented at the Annual Meeting of the Southeastern Psychological Association (21st, Atlanta, Ga.), 1975.

Swezey, Robert W. and Pearlstein, Richard B., Guidebook for Developing Criterion-Referenced Tests, Army Research Institute for the Behavioral and Social Sciences, Arlington, Va., 1975.

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

Tanenbaum, Arlene B., and Miller, Christine A. The Use of Congruence Between the Items in a Norm-Referenced Test and the Content in Compensatory Education Curricula in the Evaluation of Achievement Gains, Paper Presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York), 1977.

Weiss, Bernard J., Freedom's Ground, New York: Holt, Rinehart and Winston, 1977.

Weiss, Bernard J., Riders on the Earth, New York: Holt, Rinehart and Winston, 1977.

Weiss, Bernard J. and Stener, Loreli Olson, Time To Wonder, New York: Holt, Rinehart and Winston, 1977.

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

Wight, Albert R., "Beyond Behavioral Objectives," Readings in Measurement and Evaluation in Education and Psychology, William A. Mehrens, ed., New York: Holt, Rinehart and Winston, 1976.