

379
N811
No. 843

THE EFFECT OF IOX OBJECTIVES-BASED READING TEST COLLECTIONS
UPON FIFTH-GRADE COMPREHENSION AND
WORD-ATTACK SKILLS

DISSERTATION

Presented to the Graduate Council of the
North Texas State University in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

By

Jean Estelle Hoff, B.S., M. Ed.
Denton, Texas
August, 1974

F. H. E.

ABSTRACT

Hoff, Jean E., The Effect of IOX Objectives-Based Reading Test Collections Upon Fifth-Grade Comprehension and Word-Attack Skills. Doctor of Philosophy (Elementary Education), August, 1974, 92 pp., 13 tables, bibliography, 90 titles.

This study compares the effect of the objectives-based test collections of the Instructional Objectives Exchange on reading comprehension and word-attack skills of fifth-grade students in a basal reader program. The IOX, a non-profit educational organization, was established in the late 1960's to provide educators with instructional materials such as criterion-referenced tests to allow realistic assessment of students in reference to specific instructional objectives. IOX Director James Popham states the Exchange's purpose as encouraging educators throughout the country to use criterion-referenced instructional procedures.

The study compares gains in reading comprehension and word-attack skills of a research group with the gains of a control group, using the Stanford Diagnostic Reading Test for both pre-test and post-test. The IOX criterion-referenced tests were added to the reading program for the research group but were not given the control group.

Of the 121 students, 68 were in the experimental group and 53 in the control group. The two groups were equated by

analysis of covariance, using IQ and pre-test scores as the covariates.

Four hypotheses were investigated to determine significance of difference between the two groups, with the following results: There was no significant difference in reading comprehension scores between the experimental group and the control group. Word-attack skills scores differed significantly at the .01 confidence level between the control and the experimental groups.

The findings indicate that the use of IOX criterion-referenced tests as practice material is likely to result in a more significant increase in knowledge of word-attack skills by fifth-grade boys and girls than other published material. However, findings do not indicate superiority of IOX materials similarly used for comprehension skills.

One possible reason for this difference is that comprehension skills are more complex than word-attack skills and that any differences due to different instructional treatment would take longer to appear than the time covered in this study. Another is that norm-referenced tests such as the Stanford Diagnostic Reading Test used in this study are not sufficiently sensitive to detect growth in the numerous comprehension skills contained in the IOX materials.

TABLE OF CONTENTS

| | Page |
|---|------|
| LIST OF TABLES | v |
| Chapter | |
| I. INTRODUCTION | 1 |
| Statement of the Problem | |
| Purposes of the Study | |
| Hypotheses | |
| Significance of the Study | |
| Definition of Terms | |
| Limitations | |
| Basic Assumptions | |
| Procedures for Collecting the Data | |
| Instruments | |
| Treatment of the Data | |
| II. SURVEY OF RELATED LITERATURE | 12 |
| III. PROCEDURES FOR OBTAINING AND TREATING DATA | 44 |
| Description of the Subjects | |
| Description of Experimental Methods and Procedures | |
| Procedures for Collecting the Data | |
| Statistical Treatment of the Data | |
| Summary | |
| IV. PRESENTATION AND ANALYSIS OF THE DATA | 54 |
| Data Related to Hypothesis One | |
| Data Related to Hypothesis Two | |
| Data Related to Hypothesis Three | |
| Data Related to Hypothesis Four | |
| Summary of Findings | |
| V. SUMMARY, FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS | 66 |
| Summary of Methods and Procedures | |
| Review of the Findings | |
| Discussion of the Findings | |
| Conclusions | |
| Recommendations | |

| | Page |
|------------------------|-------------|
| APPENDIX A | 75 |
| APPENDIX B | 76 |
| APPENDIX C | 77 |
| APPENDIX D | 78 |
| APPENDIX E | 79 |
| APPENDIX F | 80 |
| APPENDIX G | 81 |
| APPENDIX H | 82 |
| APPENDIX I | 83 |
| APPENDIX J | 84 |
| BIBLIOGRAPHY | 85 |

LIST OF TABLES

| Table | Page |
|---|------|
| I. Supplementary Materials Utilized by the Control Group. | 51 |
| II. Summary of the Data Related to the Mean Scores of the Reading Comprehension Post-Test. | 55 |
| III. Summary of the Analysis-of-Covariance of the Reading Comprehension Post-Test Scores | 56 |
| IV. Summary of the Data Related to the Mean Scores of the Word-Attack Skills Post-Test. | 57 |
| V. Summary of the Analysis-of-Covariance of the Word-Attack Skills Post-Test Scores | 57 |
| VI. Summary of the Data Related to the Mean Scores of the Reading Comprehension Scores for Girls in the Experimental and Control Groups | 58 |
| VII. Summary of the Analysis-of-Covariance of the Reading Comprehension Scores for Girls. | 59 |
| VIII. Summary of the Post-Test Reading Comprehension Scores of Boys in the Experimental and Control Groups. | 60 |
| IX. Summary of the Analysis-of-Covariance of the Reading Comprehension Scores for Boys | 61 |
| X. Summary of the Data Related to the Mean Scores of the Word-Attack Skills for Girls. | 61 |
| XI. Summary of the Analysis-of-Covariance of the Word-Attack Skills Scores for Girls. | 62 |

| | Page |
|---|------|
| Table | |
| XII. Summary of the Data Related to the Mean Scores of the Word-Attack Skills Scores for Boys | 63 |
| XIII. Summary of the Analysis-of-Covariance of the Word-Attack Skills Scores for Boys | 64 |

CHAPTER I

INTRODUCTION

For several years teachers have expressed the need for additional information for use in diagnosing and prescribing learning activities for students in reading. The norm-referenced standardized achievement tests which are usually given to students in the public schools do not specify the individual strength or weaknesses of students. In recent years several companies have developed criterion-referenced tests which provide an additional dimension of student data. Such a company is the Instructional Objectives Exchange, a non-profit organization, which was established in the late 1960's.

The flexible objectives-based test collections of the IOX are designed to provide the educator with testing instruments that will allow realistic assessment of individuals in reference to explicit instructional objectives. By combining a set of clearly-stated objectives with criterion-referenced tests, the teacher should, according to proponents of this approach, be able to devise a pedagogical strategy that is more sensitive to the needs and capabilities of the pupils and which is therefore more efficacious.

Statement of the Problem

This study investigated the effect of two different sets of educational materials designed as diagnostic and instructional materials for reading comprehension and word-attack skills in a basal reader program for fifth-grade students in a suburban school district.

Purpose of the Study

The purpose of the study was to compare the gains in reading comprehension and word-attack skills of a research group with the gains of a control group, using the Stanford Diagnostic Reading Test for both the pre-test and the post-test; the difference in the program for the two groups being that the Instructional Objectives Exchange criterion-referenced tests were added to the program for the research group but were not given to the control group.

Hypotheses

To accomplish the purposes of this study, the following hypotheses were formulated:

I. The experimental group will achieve a significantly higher mean gain on the Stanford Diagnostic Reading Test in reading comprehension than will the control group.

II. The experimental group will achieve a significantly higher mean gain on the Stanford Diagnostic Reading Test in word-attack skills than will the control group.

III. A. At the culmination of the treatment the girls in the experimental group will attain a significantly higher

mean gain in reading comprehension than the girls in the control group on the Stanford Diagnostic Reading Test.

B. At the conclusions of the treatment the boys in the experimental group will attain a significantly higher mean gain in reading comprehension than the boys in the control group on the Stanford Diagnostic Reading Test.

IV. A. After the termination of the treatment the girls in the experimental group will attain a significantly higher mean gain in word-attack skills than the girls in the control group on the Stanford Diagnostic Reading Test.

B. At the end of the treatment the boys in the experimental group will attain a significantly higher mean gain in word-attack skills than the boys in the control group on the Stanford Diagnostic Reading Test.

Significance of the Study

The present study was designed to compare the effects of the use of Instructional Objectives Exchange criterion-referenced tests on the reading comprehension and word-attack skills of fifth-grade students in a basal reader program.

Today basal readers are utilized in most of the reading programs in the United States (9). From years of experience and study the classroom teachers have created most of the materials which are utilized in the basal reading programs (6). Though basal reader programs have been found to be effective in teaching reading skills, an examination of five basal reader programs widely used in the United States revealed that

teachers do not always use the suggestions concerning the procedures for individualizing instruction which are written in the preface of teacher's manuals. In most cases, these additional activities would be utilized, at the discretion of the teachers, with the students that were reading on a particular grade level; however, many teachers omit the individualized aspects of the reading program.

The present study was designed to compare the effects of Instructional Objectives Exchange criterion-referenced tests on the reading achievement of fifth-grade students. Although reports of research are available pertaining to the use of behavioral objectives, very little research can be found concerning the IOX Objectives-Based Tests. This study was an attempt to determine to what extent the IOX reading tests, used as supplementary material, effected the reading comprehension and word-attack skills of fifth-grade students in a basal reader program.

Definition of Terms

1. Reading Process is defined as the "act of reading, involving primarily the recognition of printed symbols and the meaningful reaction of the reader to these symbols; such reaction may include the reader's interpretation, appraisal, and attitudinal responses as determined by his purposes and needs" (2, p. 475).

2. Basal Reader is a "textbook, usually part of a graded series, used for instruction in reading" (2, p. 472).

3. Basal Reader Program is defined as "reading aimed at the systematic development of reading ability by means of a series of books or other materials especially suitable for each successive stage of reading development" (1, p. 473).

4. Criterion-referenced tests are "tests of essential terminal behavior as called for by the statements of learning objectives; test items should measure the student's ability to behave in performance as he was taught during the course of instruction" (2, p. 596).

5. Norm-referenced tests are "tests designed to evaluate the performance of one individual with the performance of others. Thus, the individual's score is meaningful only when it is compared to the group scores of others. This type of measurement requires that test items be modified and changed to ensure that a normal distribution occurs" (2, p. 358).

6. The Instructional Objectives Exchange (IOX) is a clearing house where instructional objectives are kept as well as created and developed. The IOX was established by the UCLA Center for the Study of Evaluation and W. James Popham is the Director of the Exchange. Presently, thirty-five different collections of objectives covering a varied range of subjects in grades K-12 can be purchased from the Exchange. The address is box 24095, Los Angeles, California, 90024 (4, p. 1).

7. The Reading Test Collections of IOX consist of a Description and Use Statement, Test Manual, Scoring Guide, Order Form, and the tests on pre-printed spirit masters. The

test manual that accompanies each test box comprises an introduction; table of contents; and a published statement of educational objectives, sample test item, and amplified objective for each test in the collection (4, p. 3).

Limitations of the Study

The following limitations were placed on this study:

1. The population involved in the study were fifth-grade students in one school which was located in a suburban, predominantly white, middle-class community. Less than one per cent of the children in this school were black.
2. There are two areas of reading considered in the study; reading comprehension and word-attack skills.
3. Only ten-weeks were allowed for the treatment.

Basic Assumptions

For this study, it was necessary to assume that:

1. the teachers participating in the experimental and control programs are equally proficient and dedicated, since no special qualifications will be stipulated for the selection of these teachers.
2. academic achievement in reading can be measured adequately by means of the Stanford Diagnostic Reading Test used in this study.
3. extraneous conditions affecting the results of the instruction being studied are evenly distributed between the groups and therefore would not significantly affect the study, and

4. the students involved in the study would cooperate and perform at their best level.

Procedures for Collecting the Data

During the month of December, 1973, permission to conduct the study was requested from the suburban school system. The superintendent wrote a letter of authorization and requested the principal and faculty to give consideration to the study. The principal of the selected school gave his permission, and six volunteer teachers were chosen to participate in conducting the research. Each of the six female teachers was assigned, randomly and without preference, to either the experimental or control group. Due to administrative regulations the intact classrooms of the various teachers were used for the study.

The subjects were selected from the total number of children in the six classrooms. A total of 121 subjects was chosen from the 188 students who participated in the research. All subjects who did not have scores on the California Short Form Test of Mental Maturity, and the Stanford Diagnostic Reading Test were eliminated from the study. Pre-existent intervening variables, such as may be observed in intact groups, were statistically controlled through the utilization of analysis of covariance, a statistical treatment which allowed the use of I. Q. and pre-test scores as the covariates. This method permits statistical rather than experimental control of variables (7, p. 254). The I. Q. score on the California Short Form Test of Mental Maturity and the pre-test scores on the

Stanford Diagnostic Test (Reading Comprehension, Syllabication, Sound Discrimination, and Blending) served as the covariates.

Upon completion of the pre-testing, the groups began receiving special reading lessons, fifteen to twenty minutes, daily, for a ten-week period. Each teacher was given a file box called a "teaching kit" on Monday morning which contained the reading lessons for each student. The material was in a manila folder with the student's name on the label. Also included in every folder were three assorted puzzles which were to be utilized only if the student finished the lessons before the allotted time. The subjects in both groups received identical puzzles each week. The folder for the pupils in the experimental group contained lessons selected from the IOX objectives-based tests collection for reading comprehension and word-attack skills. The teacher of the experimental group received instructions on how to conduct the children's lessons and a list of the instructional objectives for each lesson. A list of these materials may be seen in Appendix C. Sample students' lessons and teacher's information sheet may be viewed in Appendix A, B, and G. The control group received mimeograph lessons selected from commercial materials which can be purchased at any teachers' supply store. A list of these materials is contained in Table I. In all cases the regular classroom teacher conducted the lessons. On Friday of each week the teaching kit was collected and the materials for the week were removed and new lessons were inserted for the

following week.

Instruments

The following measures were administered to the subjects in the study:

1. The California Short-Form Test of Mental Maturity.

This test was administered to the subjects prior to the study and the I. Q. scores were obtained. This test is widely used throughout the United States (8).

2. The Stanford Diagnostic Reading Test, Level II,

Form X, Reading Comprehension, Syllabication, Sound Discrimination, and Blending. Form X was again administered at the conclusion of the study as a post-test. The test manual states that a split-half reliability coefficient for this test was .87 when the test was administered to a norm group of one hundred randomly selected fifth-grade pupils from several different communities (5). The authors state that the three-word-recognition skills at Level II should be sufficient to diagnose most pupils' word-attack skills (1).

Treatment of the Data

The subjects participating in this study were already assigned at the beginning of this experiment. Therefore, it was necessary to treat the data statistically by analysis-of-covariance. This method is used to statistically adjust the dependent variable and thus permit the use of intact classrooms (3, p. 288).

The procedures included coding the data for punch cards

and automatic data processing at the North Texas State University Computer Center, and presentation of the data in tables to facilitate interpretation. The computer output sheets will be utilized in preparing the data for presentation. All statistical findings will be reported. The null hypothesis will be rejected at the .05 level of significance.

CHAPTER BIBLIOGRAPHY

1. Buros, Oscar Krisen, The Seventh Mental Measurements Yearbook, Highland Park, New Jersey, The Gryphen Press, 1972, 726.
2. Good, Carter, Dictionary of Education, Third Edition, New York, McGraw-Hill Book Company, Inc., 1973.
3. Ferguson, George, Statistical Analysis in Psychology and Education, New York, McGraw-Hill Book Company, 1971 288.
4. IOX Major Resources for Promoting Educational Excellence, Los Angeles, California, Instructional Objectives Exchange, 1973.
5. Karlsen, Bjorn, Madden, Richard, and Gardner, Eric, Manual for Administering and Interpreting Stanford Diagnostic Reading Tests, New York, Harcourt, Brace, and World, Inc., 1966, 27-30.
6. Olson, Arthur, "Reading: The Approaches We Use," Reading and the Elementary School Child; Selected Readings on Programs and Practices, edited by Virgil Howes and Helen Fisher Darrow, New York, Macmillan Company, 1968, 254.
7. Roscoe, John T., Fundamental Research Statistics for the Behavioral Sciences, New York, Holt, Rinehart, and Winston, Inc., 1969, 254.
8. Sullivan, Elizabeth, Clark, Willis, and Tiegs, Earnest, California Short Form Test of Mental Maturity, rev. ed., Monterey, California, California Test Bureau, 1963.
9. Zintz, Miles V., The Reading Process: The Teacher and the Learner, Dubuque, Iowa, William C. Brown Company, 1970, 404.

CHAPTER II

SURVEY OF RELATED LITERATURE

The related literature in this research is presented in four sections: (1) the research related to behavioral objectives, (2) the differentiation between norm-referenced measurement and criterion-referenced measurement, (3) the application of criterion-referenced measurement in education, and (4) a description of the nature and purposes of the Instructional Objectives Exchange.

Research Related to Behavioral Objectives

It is evident that society holds the educational system responsible for the successes and failures in educating the youth of America. It is of little wonder that during these troubled times that attention has been focused on the nation's schools, and stronger demands are being made on the educational enterprise to "reaffirm their credentials as society's formal educational agencies" (25, p. 13). Throughout the nation the message is clear: changes in education must occur. Louis Lessinger states this edict concisely: "Our schools must assume a revised commitment -- that every child shall learn" (33, p. 11). Schools are being pressured to demonstrate that their students can be educated both efficiently and effectively -- in fact, to be accountable (25, p. 13).

If education can be viewed as a process whereby skills, knowledge, and values that are necessary for the fulfillment of individuals and society can be transmitted to students, then, assuming that these educational outcomes are subject to "cost accounting," the notion of behavioral objectives can be supported (25, p. 14).

Proponents of educational accountability claim that the effectiveness of the educational program can be improved by using behavioral objectives. When a student's observed behavior may be described precisely in the terms of a particular stated objective, this is tangible evidence that objectives are being achieved; the behavior demonstrated by the student conforms with a previously formulated objective (25, p. 12-13).

Duchastel and Merrill (14, p. 2) state three main instructional functions which are served by behavioral objectives: (1) "direction for teaching and curriculum development; (2) guidance in evaluation; and (3) facilitation of learning." Behavioral objectives determine the purpose of the instruction and establish the measure through which this purpose can be equated to the results. The effectiveness of the instruction can be evaluated by employing objectives to gather evidence of the changed behavior of the learner (61, p. 7).

"If something is worth teaching, isn't it worth knowing if we have succeeded?" asked Robert Mager in 1961 in his book, Preparing Educational Objectives (37). This classical little book revolutionized the educational community in relation to the possibility of utilizing objectives, and the need for such

behavioral objectives to teaching and learning. A great deal has been written about behavioral objectives since 1961, and while many educators are enthusiastic proponents of the behavioral objectives movement, some authors have questioned the validity of this movement; e.g., Kurtz (32); Liveritte (36); Stake (57).

In an attempt to discover the truth regarding behavioral objectives, many investigators have turned to research. For example, Eisner (16) claims that it is really an empirical question as to whether or not behavioral objectives are of value in curriculum development, teaching, and learning.

Since the present study concerns itself with criterion-referenced tests which are objectives-based, a review of research should include studies which are related to behavioral objectives. No attempt will be made to categorize the studies because of their interrelatedness. However, the reader may note the level at which the study was conducted for his own convenience.

Among the arguments for using behaviorally stated objectives are the claims that their use improved the quality of teaching. Researchers who have tested this theory are: McNeil (39); Jenkins and Deno (27); Piatt (43); Bryant (6); and Schneidervent (54). The results of these investigations suggest that the utilization of behavioral objectives improves instruction, is a difficult argument to support empirically. However, this approach may allow a derivation of greater practical benefits.

Authorities often cite the importance of the value of

behavioral objectives in providing guidance for evaluation. A study by Briggs, Stoker, and Scalon (5), and one by Griffin (24) have confirmed the usefulness of behaviorally stated objectives in the area of evaluation. Another more recent study which supports this belief was conducted by Duchastel (14). He used 241 high-school students to study the effectiveness of instructional objectives as organizational support in learning. It was hypothesized that behavioral objectives would direct learning on relevant materials and that this process would also interact with previous experience. In this study the author listed three independent variables: (1) objectives (given or not given); (2) arrangement of text (structured or random); (3) previous experience (valid, non-valid, none). From a "matrix" of concepts and attributes a passage of thirty sentences and five related behavioral objectives was constructed. In addition, three other passages were developed that contained either "objectives and test-items referenced to these objectives, and test-items not referenced to these, or no objectives and an assortment of test-items." After concluding his study, the author stated that the investigation afforded only limited support for the use of objectives as organizational support; however, the results did support the hypothesized orienting function of objectives as well as the interaction between this function and previous experience. It was also noted that the passages which contained either objectives or test-items referenced to these objectives improved recall.

A study by Doty (13) investigated the hypothesis that students in an industrial arts area would achieve more if they had prior knowledge of educational objectives. He used 190 seventh-grade students from seven public schools. The author concluded that there was a significant gain for those students receiving the objectives.

Blaney and McKie (4) hypothesized that a group of conference-attendees who received the objectives and another group which received the pre-test would achieve better on a post-test than a group who would be given only verbal instructions. This hypothesis resulted in a significant difference in favor of the objectives group when compared to the verbal introduction group.

Tiemann (58) investigated relationships between behavioral objectives and two types of televised instruction in a college economics class. His conclusions were that the type of television instruction was significant to the study, and that students favored instruction associated with the provision of behaviorally stated objectives.

The writing of behavioral objectives was further investigated by Dalis (10). The 133 tenth-grade subjects from health and safety classes received one of three treatments: (1) concisely-stated objectives; (2) behavioral objectives stated vaguely; or (3) health information written in paragraph form. At the end of three weeks the concisely-stated objectives group had made significant gains in comparison to the other two groups.

One study concerned with the effectiveness of stated behavioral objectives on the mathematical achievement of college students

and was conducted by Engel (17). The forty-eight elementary majors were assigned to one of the two groups. Both groups received twelve lessons in a partially programmed unit. The control group; however, did not receive behavioral objectives while the experimental group did. At the conclusion of the study the experimental group had made statistically significant gains over the control group in mathematics.

The conclusions of Conlon (7) after her research study were that prior knowledge of behavioral objectives may be useful only as guides to the educational presentations or independent study which are not "highly structured." She had investigated the effects of behavioral objectives in a highly sequenced individualized science program for seventh graders.

Perhaps one of the most interesting studies was that of Mager and McCann (reported by Mager and Clark) (38). The subjects were newly graduated engineers who were enrolled in a specialized six-months engineering course. The participants received a twenty-four page copy of the detailed objectives of the course, and complete learning instructions. The subjects were then told that the classes were cancelled and they were to proceed at their own rate. All instructors would be available for consultation if needed. As a result, the subjects completed the course in approximately seven weeks and appeared to be as well-informed as the traditional graduates. If time involved in reading the objectives is considered negligible, behavioral objectives would appear to increase the amount of attention given to the materials themselves.

All of the numerous studies concerning behavioral objectives

may appear incongruent; however, the evidence presented in this paper reveals the complexity of the issues. At this point it would seem advisable to consider what research has suggested concerning the relationship between behavioral objectives and learning:

1. One purpose for providing students with behavioral objectives may be to give direction to their learning. If the student is aware of what is expected of him, objectives would then allow him to decide what material is relevant in his studies. Rothkopf and his associates (53) have investigated hypotheses of this nature with respect to questions, and his conclusions may be directly extended to behavioral objectives.

2. A second purpose for the use of objectives may be that behavioral objectives assist the educator in the organization of his instructional lessons. However, this hypothesis requires further research.

3. It has been stated that objectives allow students to more efficiently organize their time and learning experiences in regard to the goals of the course. This management function may encourage the student not to procrastinate, and thereby avoid the last minute cramming sessions.

4. Behavioral objectives provide feedback with respect to the attainment of one's goals. A learner would know his status at all times.

5. Finally, task reinforcement is activated and maintained by the use of behavioral objectives. A student does not need to wait until the end of the instruction to receive his reinforcement (grade) because he realizes that he is

mastering the objectives as he is working through the learning tasks (14, p. 36). In conclusion, the concept of accountability has attracted interest and is currently being implemented into the educational programs of many school systems. The incorporation of accountability into the reading program requires the use of specific objectives. Renée Westcott states:

"Objectives for reading instruction, primarily cognitive (mental processes) and affective (feelings and attitudes) should be stated in performance terms to specify exactly what it is that the learner is expected to do after instructions that he could not do before" (61, p. 5). Mager (37) suggests that one reason for using behavioral objectives is that unless the educator decides his destination beforehand, he may find that he is somewhere else.

Differentiation Between Norm-Referenced and Criterion-Referenced Testing

Since 1963 when Glaser published his article in the American Psychologist (21), specialists in both the areas of measurement and instruction have been making a distinction between norm-referenced and criterion-referenced tests (27, p.1). Flanagan (18) in his 1951 article seems to have been the first writer to actually note the differences between the two types of measurement whereas the latest article on the subject is by Popham (48). Interim articles which contain similar discussion of the two terms are: Cox (9, p. 68), Davis (11, p. 1), Ebel (15, p. 15), Johnson (28, p. 355), Nitko (41, p. 3), Prescott (49, p. 347), Roudabush (52, p. 1), and Ward (60, p. 314).

A norm-referenced test, according to Fremer (20, p. 2)

is one that "yields scores which are interpreted through the use of norms. . . . These tests rely on comparisons of performance to bring meaning to individual scores."

Standardized tests of intellectual ability or achievement are examples of norm-referenced tests. Popham and Husek (47, p. 2) note that the term norm-referenced is the result of the fact that the student is compared to a normative group. The purpose for which norm-referenced measures was devised is to make comparisons between individuals on the same test. Norm-referenced measurement has some limitations which must be considered by educators. The following defects are suggested by Otto (42, p. 17).

1. Norm-referenced tests may not be suitable for use with all individuals or groups. Some test items may be too easy or too hard, or they may be inappropriate or meaningless for certain students.

2. Unrealistic time limits may be too brief for students who work slowly but accurately. The scores of these students would therefore be meaningless.

3. Test results could become invalid in a group situation for some individuals if this person did not respond to items because he failed to comprehend the directions which were given in a standardized setting.

4. The format of norm-referenced tests may limit the type of test item which can be utilized. The use of multiple-choice items which can be machine scored might restrict the

sampling of certain behaviors.

5. The appropriate instructional level is not estimated accurately by norm-referenced measurement because this type of measurement tends to "yield over-estimates of appropriate instructional level."

Popham (48, p. 614) adds two additional limitations which should be considered. He states that:

1. standardized achievement tests which are the most common example of norm-referenced measurement are "relatively insensitive to instruction."

2. test production firms which produce these norm-referenced tests plan for the same tests to be utilized by an entire country which is composed of schools with "diverse curricular preferences." Test items therefore may not be suitable for the local curriculum.

The limitations of norm-referenced measurement resulted in a dissatisfaction with these tests. Educators began a search for more efficient measurement devices. The result of this search was the new and greater emphasis on criterion-referenced measurement. That criterion-referenced measurement is not a product of the Twentieth Century may surprise many educators. Roberts states:

Criterion-referenced testing is new terminology for old concepts. In Ancient China, leaders were chosen by their ability to perform at specific tasks that were deemed necessary for leadership. Today with the advent of criterion-referenced testing, children are assessed in relationship to those skills that are deemed as necessary to perform a given task (51, p. 1).

Prescott (49, p. 345) maintains that the new interest being given to criterion-referenced measurement is due to the creation of the "controversial National Assessment of Education Progress," by the increased utilization of programmed and computer-assisted instruction, and by the advent of accountability and individualized instruction.

The definition most frequently quoted for criterion-referenced measurement is provided by Glaser and Nitko (22, p. 653). "A criterion-referenced test is one that is deliberately constructed to yield measurements that are directly interpretable in terms of specified performance standards."

There are two primary purposes served by criterion-referenced measures. These are suggested by Smith (56, p. 4) who states that the first of these is to give the individual's performance level in reference to the instructional objectives, and the other purpose is to assist in the evaluation of instruction by providing feedback information.

It has been suggested that in appearance it is very difficult to tell the difference between criterion-referenced tests and norm-referenced tests. However, Smith (56, p. 5) has listed nine points of contrast between these two types of measurement.

1. The quantitative scale which is utilized to reveal performance is different for each of the measurements. The student's performance is calculated to determine how far he has deviated from the group's performance in norm-referenced

measurement. In criterion-referenced measurement the student's performance is measured against a specified standard.

2. The second difference is variability. This point was mentioned as a limitation of norm-referenced measurement. More variability in test scores is desired in norm-referenced measurement since the meaningfulness of the score, for the most part, is dependent on the relative position of this score in relation to the other scores. In criterion-referenced measurement the score's meaning is not dependent upon the scores of others; therefore, variability is not as important. It has been stated by Popham and Husek (47, p. 3) that the "core of difference between norm-referenced and criterion-referenced tests is the issue of variability."

3. The third difference is that of "scope." Norm-referenced tests would probably test only a sample of the objectives of the course; however, criterion-referenced tests would probably test each of the essential behaviors listed in the objective.

4. "Style," is the fourth difference. In a norm-referenced test the questions are answered indirectly by the student in contrast to criterion-referenced tests where the student may be asked to directly demonstrate the desired behavior.

5. A passing score on the norm-referenced test may require that a student answer correctly only a third to a half of the items. Perhaps as much as eighty per cent of the items may be required to be answered correctly on a criterion-

referenced test. Another difference in scoring is that passing scores may be determined after completion of norm-referenced tests; however, the passing score is set prior to the administration of a criterion-referenced measure.

6. After the criterion-referenced test is completed, it is possible that the student would be given remedial work for the missed items on the test. The objective is stated and is known by the learner. Most norm-referenced tests, on the other hand, are competitive in nature and are therefore more secretive.

7. A continuously missed item on a criterion-referenced test could be an indication of poor instruction. On a norm-referenced test a continuously missed item is more likely to be revised.

8. The administration and construction of criterion-referenced tests are more difficult than norm-referenced tests. The development of criterion-referenced tests require educators to delimit the tasks to be tested, to set performance standards, and to select the precise test items which represent a specific objective. This precision of construction is lacking in norm-referenced measures. One problem noted in the creation of criterion-referenced measurement is concerned with the sampling behavior. For a specific objective the possibility of developing all of the representative items is a difficult task.

Criterion-referenced tests are designed to be diagnostic and prescriptive. Norm-referenced tests are designed to be

diagnostic and predictive. The validity and reliability for the two tests are viewed differently because of this distinction. The validity and reliability for criterion-referenced measures only closely resembles that of norm-referenced measures. Content validity is necessary for criterion-referenced tests in order that the tests will be "sensitive to appropriate instruction." The remediation of students requires that the appropriate decisions be made regarding this treatment. These decisions are more important than the reliability (52, p. 2).

Both criterion-referenced and norm-referenced measures can be helpful to the educator if he knows the specific use of each. By using a combination of the two measurements educators can better determine the needs of students (51, p. 6).

Application of Criterion-Referenced Measurement in Education

The writings of Ebel (15) and Glaser (21) in the sixties were responsible for the increased attention to criterion-referenced measurement. In 1965 Coulson and Cogswell (8, p. 59) suggested that criterion-referenced materials could be employed in the individualized instructional programs which utilized programmed instructional materials.

The latest universally recommended trend in education is the individualizing of the curriculum. As a result of this emphasis on the individual, new measurement problems have developed, and criterion-referenced measurement has been introduced into the curriculum. The purpose for using these tests is to adequately assess and to evaluate a child's growth

within a particular learning process. "Criterion-referenced tests should be thought of in terms of the educational system; if any skill or task is worth teaching it is worth measuring for growth" (51, p. 4).

Tse-Chi Hsu (26, p. 9) at the Learning Research and Development Center of the University of Pittsburgh states: "In view of recent developments in individualized instruction and instructional technology, the traditional approach of comparing a student's performance with his peers is not enough. Prescott (49, p. 349) further suggests that the interpretation involved with criterion-referenced measurement is appealing because the teacher does not need to make "invidious" comparisons between individuals.

The main reason why criterion-referenced measurement has been incorporated into the individualized instructional program is due to the fact that an educator can determine whether or not a student has mastered the concepts prerequisite to subsequent learning (19, 34, 35, 50).

One example of a specific type of program which advocates individualized instruction is the IPI curriculum (Individually Prescribed Instruction) which utilizes behavioral objectives which have been sequenced into areas, levels, and units. Criterion-referenced measurement is required in this program to "help make decisions about pupils in reference to specific instructional objectives," and to allow educators to place students properly and to permit continuous diagnosis (34, p. 15).

The utilization of criterion-referenced measurement and individualized instruction will necessitate changes in education. Millman (40, pp. 228-229) suggests that a new report card format will be required which will show both the parents and the students what has been accomplished and what now needs to be studied. Present grading systems which use numbers or letter scales are norm-referenced and are utilized by schools to compare the performance of students and ranks them. The new proposed report card should contain a listing of objectives, space to record whether the proficiency has been achieved, and a "checking system which identifies objectives" accomplished since the last reporting period. One difficulty in the development of a new report card format will be to state the objectives narrowly enough in order that the students will achieve test proficiency on at least one objective during any one reporting period.

Another change brought about by the introduction of criterion-referenced measurement is advocated by Trow (59, p. 171). He believes that the attitudes of students will be improved when there is no longer any competition connected with grades. The student will "acquire proficiencies in much the same way that a scout earns his badges."

Other educational problems in addition to individualized instructional programs where other criterion-referenced measures have been advocated are as follows: areas of diagnostic assessment (60, p. 317), accountability (40, p. 228; 49, p. 349; 60, p. 322), assessment of developmental needs

(60, p. 316), instructional program evaluation (11, p. 3; 47, p. 2; 60, pp. 315-316), construction of and/or guidance in remedial programs (25, p. 8; 60, pp. 317-318), economy of testing time (47, p. 9), reporting student progress (40, p. 227; 35, p. 16; 47, p. 7), and curriculum development (41, p. 2; 60, p. 318).

More recently educators have discovered two unique applications for criterion-referenced measurement. These are to prevent racial and ethnic bias and as well as to promote educational assistance for the disadvantaged. Green (23, p. 8) suggests that the traditional achievement tests contain racial and ethnic bias. He states that there are two possible remedies for this problem. One solution would be the construction of unbiased tests, and the other would be to use criterion-referenced tests in the diagnosis of educational problems.

Such a federal program as Title I which financially aids education lacks an adequate formula for the distribution of funds. Presently the state of Michigan is developing criterion-referenced tests which will be utilized by the state to distribute state funds to the educationally disadvantaged. The author of this article recommends that criterion-referenced tests could also be utilized by the federal government to establish relative needs among the states (31, pp. 1-22).

The preceding discussion confirms the need for criterion-referenced measurement. Of special application to this study is the report by Knipe and Kraemer (31, p. 22) regarding a

study in the Grand Forks School District in Grand Forks, Dakota. Their research concerns the utilization of criterion-referenced measurement in mathematics for grades three to nine. Further consideration was given to three other concerns of the study:

(1) adoption experiences when using criterion referenced testing, (2) research conclusions as a byproduct of this testing, and (3) attitudes of classroom teachers to this method as compared to nationally formed tests.

The authors, Knipe and Krahmer, concluded that criterion-referenced measurement has much to offer education and this type of measurement should be implemented into practice without delay. A list of adoption experiences of the Grand Forks School District as a result of the use of criterion-referenced measurement is as follows: (1) criterion-referenced tests were favored by more teachers than the norm-referenced achievement tests; (2) the "group printout" lists together all of the students who have a similar skill deficiency, which allows the teacher to plan a more efficient remedial program; (3) the tests revealed that within the school district and at the various grade levels the skill objectives received different emphasis; (4) criterion-referenced tests enable teachers to discover what the learner has accomplished and what his needs are for the future learning situation, and provides a basis for measuring instructional accountability," which is advantageous in conferences between parents, teachers, and students; (5) the problem of a time for administering the tests have been a matter

of concern in the past to the school district because of the large numbers of transient students. However, by utilizing criterion-referenced tests this problem will be eliminated by the fact that this type of measurement may be administered when the need arises at any time throughout the year; (6) evaluation and subsequent improvement of the curriculum was accomplished successfully by utilizing criterion-referenced measurement; (7) it was noted that the students were more conscientious in answering the criterion-referenced tests because they did not want to repeat skills which they had previously learned; and, (8) criterion-referenced measurement is the only type of measurement which any school district can utilize to determine whether or not the goals of the curriculum have been accomplished. The authors conclude that criterion-referenced measurement as employed in the Grand Forks School District has been successful and helpful.

Criterion-referenced measurement with all its characteristics is not new to education. Klein and Kosecoff finally suggest that what is new is the various educational applications and implications for which this type of measurement has been utilized. It is also important to note the recognition that is presently being given to criterion-referenced measurement by experts in both of the fields of curriculum and measurement (1; 3; 12; 29).

The implementation of criterion-referenced measurement into the educational programs will lead to further improvements and will result in further ramifications for educational

practice. One such example is suggested by Quie (50).

A bill is now pending before the United States Congress that would require criterion referenced test data in order to make funding decisions affecting thousands of schools and involving several million dollars.

Purpose and Defining Characteristics of The Instructional Objectives Exchange

In spite of the many advantages offered by the application of criterion-referenced measurement there are numerous reasons why this type of measurement is not used more often. Smith (56, p. 10) suggests that "one hindrance is the time, skill, and energy required to state behavioral objectives, to choose instructional procedures that will most likely assure reaching one's objectives, and to analyze tasks to determine the types of performance that are most apt to indicate mastery or lack of it." However, these are simple hurdles which can be overcome by educators who are concerned and dedicated professionals. Smith further recommends that as a beginning place educators could order materials from the Instructional Objectives Exchange.

The Instructional Objectives Exchange was created by the UCLA Center for the Study of Evaluation in 1968. The Exchange became a nonprofit educational corporation in 1970 following a separation from the Center. W. James Popham (44) who is the Director of the IOX notes that the creation of the Exchange was to encourage educators throughout the country to "initiate criterion-referenced instructional strategies."

After its creation a decision was made that the IOX

would perform the following functions: (1) encourage educators to be aware of the final learning outcome when planning the instructional lessons; (2) serve as a clearing house for the distribution of instructional objectives throughout the nation. It would utilize the already created objectives of educators and thus reduce the total amount of energy expenditures that are spent in developing instructional objectives; (3) gather and develop new measuring instruments for determining the success of attainment of objectives which are distributed by the Exchange; and (4) continually update and create additional instructional objectives in the various areas (46).

Popham (44, p. 174) explains how the Exchange operates. First, an attempt was made to publicize the existence of the Instructional Objectives Exchange and to make educators aware of the services that are available to them. Advertisements, such as articles in magazines, brochures, business letters, and releases to the news media, contained a description of the Exchange and a request for any "behaviorally stated instructional objectives" which have been created by any educational agency or school.

At the same time this appeal was made, the staff at the Instructional Objectives Exchange was already engaged in the development of new objectives and criterion-referenced related item pools. A continual refining process is constantly being conducted to improve the procedures which involve the creation of objectives and test items.

Early efforts at the IOX were dedicated to an emphasis on cognitive objectives; however, it was planned that noncognitive

goals will also be developed. With the utilization of these IOX materials Popham suggests that the Exchange can have an impact on education and could conceivably change the educational practices in America.

In addition to the opinion of Popham, a more convincing argument might be an examination of the rationale behind the Exchange. It is a belief that if teachers can plan an instructional lesson around a measurable objective, then the teacher can judge the quality of the instruction by noticing the measured effects on students.

Identifying the instructional objectives is a difficult task; therefore the Exchange has identified six criteria to be utilized in the development of objectives:

1. "General Acceptance." Do educators believe that a particular objective is worthwhile?
2. "Transferability Within the Domain." Can the learner transfer a highly specific behavior to similar skills in that identical behavior domain?
3. "Transferability Outside the Domain." Can the learner transfer behavior into different kinds of behavior domains?
4. "Terminality." Will the objectives selected be terminal rather than in route?
5. "Amenability to Instruction." Will the objective measure a "native trait" which is relatively insensitive to instruction?
6. "Ease of Scoreability." Can the learner's responses to the objective be easily scored?

These criteria are then employed by the external consultants (usually master-teachers) who then evaluate the instructional objectives. In addition, the Instructional Objectives Exchange materials are given to small groups of learners to determine if there is any confusion regarding the objectives, directions, etc.

At the same time there are three different internal IOX reviews which determine the congruence between the objectives and the tests designed to measure them. Instead of the typical test validity methods utilized with norm-referenced tests, Instructional Objectives Exchange consultants employ a content validity strategy; that is, "the judged congruence between test measures and the objective they are designed to assess." Each new IOX objective-based measure is then field tested to determine such information as homogeneity, test-retest reliability, etc.

Educators are encouraged to utilize the following suggestions regarding the employment of the objectives-based tests from the Exchange. The discussion is about the mechanics involved in administering the tests, and the possibilities that are available for evaluation and instructional purposes:

- (1) the cost of the IOX tests has remained economically feasible due to the decision to prepare these materials on pre-printed spirit masters from which educators can then duplicate approximately 200 to 300 copies;
- (2) the time required for the administration of each test is approximately five to ten minutes, and it is possible to schedule several related tests during the class

time; (3) the test collections are suitable for very young children and can even be self-scored; (4) scoring is very simple because the answers to most items can be marked with an X, and unless indicated otherwise, one point per correct response is recommended; and, (5) if a district decides to administer these tests to a large group, it is recommended that the possibility of using separate answer sheets and machine scoring be explored (46).

Reporting results to parents could be handled in several different ways. Millman (40, p. 226) has suggested that an entirely new criterion-referenced report card is desirable. Another suggestion for reporting results is to set up a profile sheet. The columns at the left could contain the numbers of the objectives and across the top of the sheet could be listed the percentage increments. Easy calculations are permitted by the IOX tests. Also, by using different colored marking pens the educator can easily construct line or bar graphs which graphically depict the student's percentage correct for the different objectives for pre- and post-tests. Another alternative would be to determine the mastery criterion levels; however, this is more difficult (45). It is not necessary to use the same passing percentage for each test. A high criterion level of 80 or 100 per cent might be considered adequate for tests of the most important skills while a lower passing level might be considered for tests containing less important skills. Further, since Instructional Objectives Exchange tests are available in two forms (A and B), it is possible to administer

different forms to each.

The teacher should consult the amplified objective before devising the needed instructional lessons. The objective is designed to inform the educator as to the exact boundaries of the behavior that is expected, and in addition, to advise what it is that is considered correct and incorrect learner responses.

At the present time there are available at the Instructional Objectives Exchange some thirty-five different collections of objectives which are suitable for grades kindergarten through twelve and cover a variety of different subjects. In most collections you receive an objective and six test items which are designed to reflect the obtainment of the objective. Two of the original thirty-five sets of objectives are different in that the focus of these objectives is on the affective concerns such as the self concepts of learners as well as their attitudes toward school. The other thirty-three collections represent cognitive outcomes.

Encouraging response to the Instructional Objectives Exchange has been given by the educators of America. In fact, 20,000 collections were ordered in the first 18 months since the creation of the Exchange. All monies received are being used to create new collections and to improve previously developed collections. Comments and suggestions from educators are solicited and utilized by the Exchange to constantly improve the quality of all materials (46).

Smith (56, p. 10) has recommended that educators could

use the extensive collections of the IOX as a starting point for the utilization of objectives and criterion-referenced tests in their own school districts. Millman (40, p. 228) concludes that the Instructional Objectives Exchange is one of the most ambitious firms in the creation and development of criterion-referenced materials.

CHAPTER BIBLIOGRAPHY

1. Airasian, P., and Madaus, G., "Criterion Referenced Testing in the Classroom," Measurement in Education, III, 1972, 1-8.
2. Baker, Eva, "Effects on Student Achievement of Behavioral and Non-behavioral Objectives," The Journal of Experimental Education, XXXVII, 1969, 5-8.
3. Baker, R. L., "Measurement Considerations in Instruction Product Development," paper presented at Conference on Problems in Objectives Based Measurement, Center for the Study of Evaluation, University of California, 1972.
4. Blaney, J. P., and McKie, D., "Knowledge of Conference Objectives and Effect Upon Learning," Adult Education Journal, XIX, 1969, 98-105.
5. Briggs, L. J., Stoker, H. W., and Scanlon, P., "Comparison of Performance on Objective-Referenced vs. Content Referenced Achievement Tests, Mimeographed Document, Florida State University, 1971.
6. Bryant, N., "The Effects of Performance Objectives on the Achievement Level of Selected Eighth-Grade Science Pupils in Four Predominantly Black Inner-City Schools," Unpublished doctoral dissertation, Indiana University, 1970.
7. Conlon, Betsy, "A Comparison of the Performance of Seventh-Grade Students With and Without Prior Knowledge of the Objectives of an Individualized Science Program, Unpublished dissertation, Florida State University, 1970.
8. Coulson, J. E., and Cogswell, J. F., "Effects of Individualized Instruction on Testing," Journal of Educational Measurement, II, 1965, 59-64.
9. Cox, Richard, "Evaluative Aspects of Criterion-Referenced Measures," paper presented at the 1970 meeting of the American Educational Research Association, Minneapolis, Minnesota, 1970.
10. Dalis, G. T., "Effect of Precise Objectives Upon Student Achievement in Health Education," The Journal of Experimental Education, XXXIX, 1970.

11. Davis, Frederick, "Criterion-Referenced Tests," paper presented at the annual meeting of the American Educational Research Association, New York City, New York, 1971.
12. Davis, F. B., "Criterion Referenced Measurement," AERA Conference Summaries, Princeton, New Jersey, 1972.
13. Doty, C. R., "The Effect of Practice and Prior Knowledge of Educational Objectives on Performance, unpublished doctoral dissertation, Ohio State University, 1968.
14. Duchastel, Philippe, "An Investigation of the Organizing Function of Instructional Objectives in Relation to Experience with Objective-Based Testing," Dissertation Abstracts, XXXIII, (September, 1972).
15. Ebel, Robert, "Content Standard Test Scores," Educational and Psychological Measurement, XXII, (Spring, 1962), 15-25.
16. Eisner, E. W., "Educational Objectives: Help or Hindrance," School Review, LXXV, 1967, 251-282.
17. Engel, Roberta, "An Experimental Study of the Effect of Stated Behavioral Objectives on Achievement in a Unit of Instruction on Negative and Rational Base Systems of Numeration," Unpublished master's thesis, University of Maryland, 1968.
18. Flanagan, J. C., "Unit, Scores, and Norms," Educational Measurement, Washington, D. C., American Council on Education, 1951, 695-763.
19. Flanagan, J. C., and Jung, S. M., "Evaluating a Comprehensive Educational System," Evaluative Research, Pittsburgh, Pennsylvania, American Institute for Research, 1970.
20. Fremer, John, "Applications of Criterion-Referencing to Schools," paper presented at Annual Conference of Educational Record Bureau, New York City, New York, 1973.
21. Glaser, Robert, "Instructional Technology and the Measurement of Learning Outcomes: Some Questions," American Psychologist, (August, 1963), 519-521.
22. Glaser, Robert, and Nitko, Anthony J., "Measurement in Learning and Instruction," Educational Measurement, edited by Robert Thorndike, Washington, D. C., American Council on Education, 1971, 625-670.

23. Green, Donald, "Racial and Ethnic Bias in Achievement Tests and What to Do About it," CTB/McGraw-Hill, 1972, 1-9.
24. Griffin, J. E., "The Relationship Between Behavioral Objectives and Measurement Instruments Used to Evaluate Student Progress in an Urban Adult Basic Education Program, Unpublished doctoral dissertation, Catholic University of America, 1971.
25. Hodges, Richard, "Some Assumptions about Behavioral Objectives as Related to Reading," Accountability and Reading Instruction: Critical Issues, edited by Robert Ruddell, Urbana, Illinois, National Council of Teachers of English, 1973.
26. Hsu, Tse-Chi, "Empirical Data on Criterion-Referenced Tests," paper presented at the annual meeting of the American Educational Research Association, New York City, New York, February, 1971.
27. Jenkins, Joseph, and Deno, Stanley, "Influence of Knowledge and Type of Objectives on Subject-Matter Learning," Journal of Educational Psychology, LXII, 1971, 67-70.
28. Johnson, Marjorie, and Kress, Roy, "Task Analysis for Criterion-Referenced Tests," The Reading Teacher, (January, 1971), XXIV, 355-359.
29. Keller, C. M., "Criterion Referenced Measurement: A Bibliography," Princeton, New Jersey, ERIC Clearinghouse on Tests, Measurement, and Evaluation, 1972.
30. Klein, Stephen, and Kosecoff, Jacqueline, Issues and Procedures in the Development of Criterion Referenced Tests, Washington, D. C., National Institute of Education, 1973.
31. Knipe, Walter, and Krahmer, Edward, "An Application of Criterion Referenced Testing," paper presented at the annual meeting of the American Educational Research Association, New Orleans, Louisiana, February, 1973.
32. Kurtz, Paul, "The Effect of Practice and Knowledge of Instructional Objectives on Learning from Prose," Dissertation Abstracts, XXXIII, 1972.
33. Lessenger, Leon, "Accountability in Education," National Committee for Support of the Public Schools, Washinton Government Printing Press, February, 1970.

34. Lindvall, C. M., and Cox, Richard, The IPI Evaluation Program, Chicago, Illinois, Rand McNally and Company, 1970.
35. Lipe, Dewey and Jung, Steven, "Manipulating Incentives to Enhance School Learning," Review of Educational Research, (October, 1971), XXI, 249-280.
36. Liveritte, Rudy, "A Philosophical Analysis of the Doctrine of Behavioral Objectives," Dissertation Abstracts, XXXIII, January, 1973.
37. Mager, R. F., Preparing Instructional Objectives, Palo Alto, California, Fearon Publishing Company, 1962.
38. Mager, R. F., and Clark, C., "Explorations in Student-Controlled Instruction," Psychological Reports, XIII, 1963, 71-76.
39. McNeil, John, "Concomitants of Using Behavioral Objectives in the Assessment of Teacher Effectiveness," The Journal of Experimental Education, XXXVI, 1967, 69-74.
40. Millman, Jason, "Reporting Student Progress: A Case for Criterion-Referenced Marking System," Phi Delta Kappan, (December, 1970), LII, 226-230.
41. Nitko, Anthony, "Criterion-Referenced Testing in the Context of Instruction," paper presented at the Educational Records Bureau, National Council of Measurement in Education, New York, October, 1970.
42. Otto, Walter, "Assessing Needs and Growth," Assessment Problems in Reading, edited by Walter MacGinitie, Delaware, International Reading Association, 1973, 16-17.
43. Piatt, R. G., "An Investigation of the Effect the Training of Teachers in Defining Writing, and Implementing Educational Behavioral Objectives has on Learner Outcomes for Students Enrolled in a Seventh-Grade Mathematics Program in the Public Schools, Unpublished Doctoral Dissertation, Lehigh University, 1969.
44. Popham, James, "The Instructional Objectives Exchange: New Support for Criterion-Referenced Instruction," Phi Delta Kappan, XXI, (November, 1970), 174-175.
45. Popham, James, "Defining Educational Objectives and Standards in Implementing the Stull Act," Stanford University Conference on Stull Act, October, 1972.
46. Popham, James, Instructional Objectives Exchange UCLA Graduate School of Education, (no date).

47. Popham, James, and Husek, T. R., "Implications of Criterion-Referenced Measurement," Journal of Educational Measurement, VI, (Spring, 1969), 1-9.
48. Popham, W. James, "An Approaching Peril: Cloud-Referenced Tests," Phi Delta Kappan, (May, 1974), 614-615.
49. Prescott, George, "Criterion-Referenced Test Interpretation in Reading," The Reading Teacher, (January, XXIV, 347-354.
50. Quie, Albert, "A New Approach to the Education of the Disadvantaged," paper presented to American Association of Colleges for Teacher Education, Chicago, Illinois, February, 1973.
51. Roberts, Tim, "Some Utilities of Criterion Testing," paper read at the Learning Disabilities Conference, Longmost, Colorado, November, 1973.
52. Roudabush, Glenn, "Item Selection for Criterion-Referenced Tests," paper presented at the American Educational Research Association meetings, New Orleans, Louisiana, February, 1973.
53. Rothkopf, E. Z., "The Concept of Mathemagenic Activities," Review of Educational Research, XL, 1970, 325-336.
54. Schneiderwent, M. O., "The Effects of Using Behavioral Objectives in the Instruction of Harvard Project Physics," Unpublished Doctoral Dissertation, University of Northern Colorado, 1970.
55. Shanner, W. M., "Evaluation of Mastery Level in Systems of Individualized Instruction," Systems in Education, Los Angeles, California, Daniel, Mann, Johnson and Mendenhall, 1968.
56. Smith, Charles, "Criterion-Referenced Assessment," paper presented at International Symposium on Educational Testing, the Hague, The Netherlands, July, 1973.
57. Stake, Robert, "1: Objectives, Priorities, and Other Judgment Data," Review of Educational Research, Washington, Education Research Association, XL, (April, 1970), 181-212.
58. Tiemann, P. W., "Student Use of Behaviorally-Stated Objectives to Augment Conventional and Programmed Revisions of Televised College Economics Lectures," Paper read at the annaul meeting of AERA, Chicago, 1968.

59. Trow, William Clark, "On Marks, Norms, and Proficiency Scores," Phi Delta Kappan, (December, 1966), 171-173.
60. Ward, J., "On the Concept of Criterion-Referenced Measurement," British Journal of Educational Psychology, (November, 1970), 314-323.
61. Westcott, Renée, "The Use of Media in Higher Education," Raleigh, North Carolina, State Department of Higher Education Facilities, May, 1970.

CHAPTER III

PROCEDURES FOR OBTAINING AND TREATING DATA

This study involved the incorporation of Instructional Objectives Exchange reading criterion-referenced materials into a basal reading program for fifth-grade students. The subjects were 121 fifth-grade students in one intermediate school. There was an experimental group composed of four classrooms and a control group which consisted of two classrooms. The two groups were equated for initial differences by the analysis-of-covariance on non-language I. Q. scores and reading comprehension and word-attack skills on a pre-test. This chapter presents in detail a description of the subjects, a description of the experimental methods and materials, procedures for collecting the data, the statistical treatment of the data, and a summary.

Description of the Subjects

Initially there was a total of 188 subjects involved in the study. They were selected from a large suburban school district. The six classrooms were randomly selected from nine fifth-grade sections which were temporarily assigned to an intermediate school while a new elementary school was being constructed. It was realized that not all of these students were eligible for the program and that their test scores could not be used in determining the results of this study.

Ineligibility was based on lack of I. Q. scores for certain transient students. The actual loss of eligible students during the course of the study was not as great as might be expected. Of the 140 eligible students who participated at the beginning of the program, 121 completed the program and it was the scores of these students which were subjected to statistical procedures to determine the results of the study. The ineligible students were allowed to participate in the program without using data generated by them in order to minimize classroom and teacher confusion. In this manner non-participating students were not subjected to undue peer pressures. Also, less extra and unnecessary work was generated for teachers who were cooperating by administering tests to students during the study. The experimental loss during the study can be attributed to a lack of pre- or post test scores, lengthy absenteeism, or moving from the school district.

The California Short Form Test of Mental Maturity had been administered by school officials prior to this study and these scores were obtained from the cumulative folders. The data for the present study were taken from the scores on Form X, Level II of the Stanford Diagnostic Reading Test.

Of the 121 students involved in the study 68 were in the experimental group. In this group there were twenty-three boys and forty-five girls. The control group, with a total of fifty-three subjects, had twenty-seven boys, and twenty-six girls. Statistical consideration was necessary since these classrooms were intact and was met through the use of analysis-of-covariance (9). It will be discussed further in this chapter, under the

topic, The Statistical Treatment of the Data

Description of Experimental Methods and Materials

This study required the two groups of subjects participate in an experimental study over a ten-week period during the spring semester of fifth-grade in 1974. The IOX materials were used in conjunction with the basal reader program and lessons were to be presented each day during the school week for the duration of the study.

Selection of the teachers for this study was limited due to existing class assignments. Efforts to avoid biased sampling results were made by the random selection of the participating classrooms to either the experimental or control group. The teachers involved in the study were female and held bachelor's degrees in education. One teacher, who taught a class of students who were in the experimental group, also held a master's degree. Teachers of students who were in the experimental group had a total of eighteen years of teaching experience, with a mean of 4.5 years. Teachers of the subjects who were in the control group had a total of sixteen years of teaching experience, with a mean of 8.0 years. There were two first-year teachers in the experimental group and one first-year teacher in the control group.

During the week of January 28, 1974, in-service training was conducted for both the experimental and control group teachers. The meetings were scheduled at different times during the same week in order that the two groups be instructed separately. No mention was made to the teachers about which groups would be

the experimental or the control. All equipment, such as student folders, and teaching kits were identical for both groups. During the in-service meetings all of the teachers were presented with educational literature. The book given to the experimental group dealt directly with criterion-referenced measurement and instruction; however, the control group teachers received a book which was not concerned directly with the study, but was useful in general classroom instruction. All of the teachers were requested not to discuss any materials or procedures with each other, but to defer questions that needed to be answered to special group meetings. All of the above mentioned items, such as the non-partial treatment given to both groups of teachers, were an attempt to prevent the "Hawthorne Effect" (2).

The first in-service meeting was conducted for the teachers of the control group. These teachers agreed to use just the materials included in their regular basal reader programs plus the lessons provided for instruction in this study. Procedures for the distribution, administration, and collection of materials were discussed. The book, Independent Learning in the Classroom (11), was presented to these teachers. Also, all of the materials required for administering the pre-test were distributed to the teachers. The directions for administering the Stanford Diagnostic Reading Test were discussed and the testing time schedule was decided.

The in-service program for the teachers of the experimental group was conducted on the following day. The educators in this group were also asked to continue the instruction as

suggested in the basal reader program and in addition they were requested to utilize the materials from the Instructional Objectives Exchange. Each week the teachers would receive an IOX folder which contained a separate instruction and information sheet. The teachers were asked to read over this sheet which contained the objectives for each test before presenting the criterion-referenced tests to the children. A sample teacher's instruction and information sheet may be examined in Appendix H. Further, the teachers were given information concerning the Instructional Objectives Exchange and were instructed to be concerned with the word-attack skills and reading comprehension collections only. Information given to the teachers regarding the word-attack skills collections of objectives-based tests was selected from the IOX book, Language Arts: Decoding Skills K-12 (4, p. 3). This book was designed to accompany the test collection of word-attack skills and contains additional material to "enhance the teacher's skill in accomplishing prespecified instructional objectives" (5). In the discussion with the teachers in regard to the lessons on word-attack skills, it was stressed that this collection consisted of objectives arranged into five sections which were entitled, "Discrimination, Sight Vocabulary, Recognition of Letters by Name, Recognition of Sound and Their Association with Letters, and Pronunciation." It was recommended that the directions on the tests could be read or paraphrased, and that items requiring an oral stimulus might be given by the teacher.

Information concerned with the use of the reading comprehension collection were selected from the IOX book,

Language Arts: Comprehension Skills K-12 (3, pp. 13-14).

The teachers were requested not to write any objectives in connection with a particular item on the tests even though this was suggested by the guide. They were instructed to utilize only the prepared IOX materials that were given to them. It was explained that the reading comprehension collection was organized into four sections which included, "Literal Comprehension, Interpretation, Vocabulary and Word Meaning, and Comprehension of Non-Written Materials." Samples of the word-attack skills and reading comprehension tests were given to the teachers for examination purposes. Samples of these materials may be examined in Appendix A and B. Appendix C is a copy of the Table of Contents for both of these collections.

Attention was directed to the difference between the two collections. In the reading comprehension collection there are lessons in which the instructions to students may be included within the item and may not be labelled as "direction." Any questions from the teachers were answered at this time. The procedures for the handling and distribution of these materials were then discussed. The teachers were told that the materials would be ready on Monday morning and would be collected after the reading period on Friday.

A short discussion of testing procedures was then accomplished. The teachers viewed a filmstrip entitled, Systematic Instructional Decision-Making. At the conclusion of the meeting the teachers of the experimental group were presented with the book, Planning an Instructional Sequence (8) which they were asked to read in connection with the study.

Procedures for Collecting the Data

Subjects were 121 fifth-grade students in six fifth-grade self-contained classrooms. Data from these subjects were collected over a ten-week period. All of the subjects' lessons were corrected and filed for reference. Each student was assigned an Arabic numeral which was listed on the label of his folder. The Arabic numeral, sex identification, name of the assigned teacher, along with pre- and post-test scores were recorded on keypunch worksheets and then punched on cards for processing at the North Texas State University Computer Center.

The Stanford Diagnostic Reading Test, Form X (6) was administered as the pre- and post-test by the classroom teacher. This is a group test which yields scores for reading "comprehension, Syllabication, Sound Discrimination, and Blending." These achievement tests were administered on Tuesday, Wednesday, and Thursday of the week of February 4, 1974.

The procedures for distributing the materials was essentially the same for both groups. The teaching kits were given to each teacher usually about 7:45 on Monday morning before school began at 8:30. At approximately 9:50 of each morning during the reading class, the folders were distributed to the students. The teacher also received a folder which included the time schedule, instructions for the daily lessons, the objectives and related information, and answer keys. The time allowed each day for the lessons was approximately fifteen to twenty minutes. The time schedule was developed for the purpose of allowing the educator to teach a certain lesson on a particular day. Answer keys for the assorted puzzles were

made available to the teacher. These puzzles were included in the student's folder and were to be utilized only if the student completed his lessons before the time limit. Two sample puzzles may be examined in Appendix D and E.

At this point the similarity between the two groups of folders ended. The instructions to the teachers of the control group were essentially the same each week; however, this depended upon the selection of materials included for the week's lessons. The material used by the control group was selected from commercially prepared mimeograph books which were purchased from a teacher's supply store. This type of material can usually be found in nearly every conventional classroom. A sample lesson is included in Appendix F. A list of these commercial materials and the name of the publisher is listed below in Table I.

Table I

SUPPLEMENTARY MATERIALS UTILIZED BY
THE CONTROL GROUP

| Publication | Publisher |
|--|-----------------------------|
| <u>Individualized Reading</u> | Hayes Publishing Company |
| <u>Look, Listen, and Learn</u> <u>Series - Comprehension</u> <u>Skills</u> | Milliken Publishing Company |
| <u>Look, Listen, and Learn</u> <u>Series - Syllabication</u> | Milliken Publishing Company |
| <u>Our Living Language,</u> <u>Grade 5, Book 2</u> | Milliken Publishing Company |
| <u>Phonics V</u> | Hayes Publishing Company |
| <u>Reference Tools and Study</u> <u>Skills</u> | Milliken Publishing Company |

Advertisements, in teacher supply catalogs for these books of duplicating masters, state that these materials are useful

in teaching "evaluation, interpretation, critical thinking, vocabulary, expression of ideas, word-attack skills, and communication."

The instructions for the teachers of the experimental group included a copy of the main "Objective, Sample Items, Amplified Objectives, Response Alternatives, and Criterion of Correction," (5) for each day's lesson. A sample of the teacher's instruction for one of the instructional lessons conducted on February 18, 1974 is included in Appendix G and H.

The basal reader for this study was the book, Images which was written by William Durr, Vivian Windley, and Anne McCourt. The reader was published by the Houghton-Mifflin Company and the copyright date was 1971. This book was a new adoption for the district and was being use for the first time. The children did not have the workbook that accompanies the reader, but the teacher had access to mimeograph masters of the workbook.

Statistical Treatment of the Data

Since, initially, there was no attempt to match the groups on reading ability or intelligence, it was felt that the best way to treat the data would be the analysis-of-covariance. This statistical procedure permits the equating of groups statistically rather than experimentally (9). An F-ratio was obtained to indicate any significant difference between the mean comprehension and word-attack scores of the two groups (experimental and control).

All hypotheses were tested by a one-way analysis-of-

covariance with I.Q. and pre-test scores as covariates.

Hypothesis I compares the reading comprehension scores of all of the experimental group with those of the control group. Hypothesis II compares the word-attack skills scores of all the experimental groups with those of the control groups. Hypothesis III, Part A, compares the reading comprehension scores of the experimental girls with those of the control girls. Hypothesis III, Part B, compares the reading comprehension scores of the boys in the experimental group with those of the boys in the control group. Hypothesis IV, Part A, compares the word-attack skills scores of the girls in the experimental group with those of the girls in the control group. Hypothesis IV, Part B, compares the word-attack skills scores of the experimental boys with those of the control boys.

The analysis of covariance yielded an F -ratio which was used for determining statistical significance at the .05 level. Tables in the statistic book by Ferguson (1) were consulted.

Summary

In this chapter, the fifth-grade subjects who participated in this study have been described. Information related to the experiment, such as a description of the experimental method and materials used, was reported. An explanation of the in-service meetings which were conducted for the participating teachers was discussed, and the procedures for collecting and analyzing the data was reported.

In the next chapter, the data will be presented under headings related to the hypotheses tested. Tables are included wherever needed.

CHAPTER IV

PRESENTATION AND ANALYSIS OF THE DATA

Presented in this chapter is an analysis of the data comparing two groups of fifth-grade children participating in an experimental study of the effectiveness of Instructional Objectives Exchange materials upon the learning of comprehension and word-attack skills in reading.

One hundred and twenty-one students were in the total group, with sixty-eight in the experimental group and fifty-three in the control group. Pre-existent intervening variables, such as may be observed in intact groups, were statistically controlled through the utilization of analysis-of-covariance, a statistical treatment which allowed the use of I. Q. and pre-test scores as the covariates. This statistical treatment is used to test significance of difference for mean gain. The appropriate table in Ferguson's textbook, Statistical Analysis in Psychology and Education, page 452, was consulted for the F-ratio required to reject the null hypothesis at the .05 level of significance.

It was expected that the children, both boys and girls, in the experimental group would score better on both reading comprehension and word-attack skills than the children, regardless of sex, in the control group. Results of the study are presented in the order of the hypotheses as stated in Chapter I.

Data Related to Hypothesis One

Research Hypothesis I was restated to read: There will be no significant difference in the mean gains of the experimental and control groups for reading comprehension. For this hypothesis achievement test data was calculated for 121 subjects. The mean scores of the experimental and control groups for reading comprehension are shown in Table II.

Table II

SUMMARY OF THE DATA RELATED TO THE MEAN SCORES OF THE READING COMPREHENSION POST-TEST

| Group | N | Mean I. Q. | Post-Test Mean Score | S.D. | Adjusted Mean |
|--------------|----|------------|----------------------|------|---------------|
| Control | 53 | 110.66 | 44.05 | 6.51 | 40.68 |
| Experimental | 68 | 101.47 | 39.46 | 5.99 | 42.08 |

Before adjustment the mean score of the experimental group was less than the mean score of the control group by 4.59. After adjustment was made for I. Q. and pre-test, the mean score of the experimental group was greater than the mean score of the control group by 1.40.

It was stated in Hypothesis I, that the students in the experimental group would achieve a significantly greater mean reading comprehension score on the Stanford Diagnostic Reading Test than would the students in the control group. Hypothesis I, in effect, is whether the above difference in the adjusted mean is significant. An analysis-of-covariance is presented in Table III on the following page between the experimental and

control groups. This comparison was made to determine if there was a significant difference between the mean reading comprehension scores of the two groups.

TABLE III

SUMMARY OF THE ANALYSIS-OF-COVARIANCE OF THE
READING COMPREHENSION POST-TEST SCORES

N = 121

| Source | Degrees of Freedom | Sum of Square | Mean Square | F Value | Probability |
|------------|--------------------|---------------|-------------|---------|-------------|
| Total | 118 | 1889.84 | --- | --- | --- |
| Within | 117 | 1848.66 | 15.80 | --- | --- |
| Difference | 1 | 41.18 | 41.18 | 2.60 | 0.1091 |

The obtained F-ratio of 2.60 was not acceptable at the .05 level of significance; therefore, Hypothesis I was not supported. This would indicate that the experimental group did not demonstrate a statistically greater mean increase in reading comprehension than did the control group.

Data Related to Hypothesis Two

Research Hypothesis II was restated to read: There will be no significant difference in the mean gains of the experimental and control groups for word-attack skills. For this hypothesis achievement test data was calculated for 121 subjects. The mean scores of the experimental and control groups for word-attack skills are shown in Table IV.

Before adjustment, the mean score of the experimental group was less than the mean score of the control group by 2.93.

TABLE IV

SUMMARY OF THE DATA RELATED TO THE MEAN SCORES
OF THE WORD-ATTACK SKILLS POST-TEST

| Group | N | Mean I. Q. | Post-Test Mean Score | S.D. | Adjusted Mean |
|--------------|----|---------------|-------------------------|-------|---------------|
| Control | 53 | 110.66 | 77.92 | 12.00 | 72.36 |
| Experimental | 68 | 101.47 | 74.99 | 10.46 | 79.32 |

It can be noted in the above table that after adjustment was made for I. Q. and pre-test, the mean score of the experimental group was greater than the mean score of the control group by 6.96.

It was stated in Hypothesis II that the students in the experimental group would achieve a significantly greater mean word-attack skills score on the Stanford Diagnostic Reading Test than would the students in the control group. Hypothesis II, in effect, is whether the above difference in the adjusted mean is significant. The results of the analysis-of-covariance in which the difference was tested is shown in Table V below.

TABLE V

SUMMARY OF THE ANALYSIS-OF-COVARIANCE OF THE
WORD-ATTACK SKILLS POST-TEST SCORES

N = 121

| Source | Degrees of Freedom | Sum of Square | Mean Square | F Value | Probability |
|------------|--------------------------|------------------|----------------|------------|-------------|
| Total | 118 | 6586.11 | --- | --- | --- |
| Within | 117 | 5500.18 | 47.01 | --- | --- |
| Difference | 1 | 1085.93 | 1085.93 | 23.10 | 0.0000 |

The obtained F-ratio of 23.10 was acceptable at the .01 level of significance; therefore, Hypothesis II was supported. This would indicate that the experimental group did demonstrate a statistically greater mean increase in word-attack skills than did the control group.

Data Related to Hypothesis Three, Part A

The mean scores of the girls in the experimental and control groups for reading comprehension are shown in Table VI. The null form of the research Hypothesis III, Part A was stated as follows: There will be no significant difference in the mean gains of the girls in the experimental group and the girls in the control group for reading comprehension.

TABLE VI

SUMMARY OF THE DATA RELATED TO THE MEAN SCORES OF THE READING COMPREHENSION SCORES FOR GIRLS IN THE EXPERIMENTAL AND CONTROL GROUPS

| Group | N | Mean I. Q. | Post-Test Mean Score | S.D. | Adjusted Mean |
|--------------|----|------------|----------------------|------|---------------|
| Control | 26 | 113.85 | 47.42 | 5.76 | 43.21 |
| Experimental | 45 | 100.76 | 39.93 | 4.99 | 42.37 |

Before adjustment the mean score of the experimental group was less than the mean score of the control group by 7.49. After adjustment was made for I. Q. and pre-test, the mean score of the experimental group was less than the mean score of the control group by 0.84.

It was stated that in Hypothesis III, Part A, the girls

in the experimental group would achieve a significantly greater mean reading comprehension score on the Stanford Diagnostic Reading Test than would the girls in the control group.

Hypothesis III, Part A, in effect, is whether the above difference in the adjusted mean is significant. The results of the analysis-of-covariance in which the difference was tested are shown in Table VII below.

TABLE VII

SUMMARY OF THE ANALYSIS-OF-COVARIANCE OF
THE READING COMPREHENSION SCORES FOR
GIRLS

N = 71

| Source | Degrees of Freedom | Sum of Square | Mean Square | F Value | Probability |
|------------|--------------------|---------------|-------------|---------|-------------|
| Total | 68 | 975.27 | --- | --- | --- |
| Within | 67 | 968.67 | 14.46 | --- | --- |
| Difference | 1 | 6.60 | 6.60 | 0.46 | 0.5016 |

The obtained F-ratio of 0.46 was not acceptable at the .05 level of significance; therefore, Hypothesis III, Part A was not supported. This would imply that the girls in the experimental group did not demonstrate a statistically significant greater mean increase in reading comprehension than did the girls in the control group.

Data Related to Hypothesis Three, Part B

Research Hypothesis III, Part B was restated to read: There will be no significant difference in the mean gains of the boys in the experimental group and the boys in the control group for reading comprehension. The mean scores of the boys

in the experimental group was less than the mean score of the control group by 2.29. After adjustment was made for I. Q. and pre-test, the mean score of the experimental group was greater than the mean score of the control group by 1.81. These results can be noted in Table VIII.

TABLE VIII
SUMMARY OF THE POST-TEST READING COMPREHENSION
OF BOYS IN THE EXPERIMENTAL
AND CONTROL GROUPS

| Group | N | Mean I. Q. | Post-Test Mean Score | S.D. | Adjusted Mean |
|--------------|----|------------|----------------------|------|---------------|
| Control | 27 | 107.59 | 40.81 | 5.53 | 38.93 |
| Experimental | 23 | 102.87 | 38.52 | 7.61 | 40.74 |

It was stated that in Hypothesis III, Part B, the boys in the experimental group would achieve a significantly greater mean reading comprehension score on the Stanford Diagnostic Reading Test than would the boys in the control group. Hypothesis III, Part B, in effect, is a test of whether the above difference in the adjusted mean is significant. The results of the analysis-of-covariance in which the difference was tested are shown in Table IX on the next page.

The obtained F-ratio of 2.70 as reported in Table IX was not acceptable at the .05 level of significance; therefore, Hypothesis III, Part B was not supported. This would indicate that the boys in the experimental group did not demonstrate a statistically greater mean increase in reading comprehension

than did the boys in the control group.

TABLE IX
SUMMARY OF THE ANALYSIS-OF-COVARIANCE OF THE
READING COMPREHENSION SCORES FOR BOYS

N = 50

| Source | Degrees of Freedom | Sum of Square | Mean Square | F Value | Probability |
|------------|--------------------|---------------|-------------|---------|-------------|
| Total | 47 | 626.97 | --- | --- | --- |
| Within | 46 | 592.26 | 12.88 | --- | --- |
| Difference | 1 | 34.71 | 34.71 | 2.70 | 0.1074 |

Data Related to Hypothesis Four, Part A

The null form of the research Hypothesis IV, Part A was stated as follows: There will be no significant difference in the mean gains of the girls in the experimental group and the girls in the control group for word-attack skills. The mean scores of the girls in the experimental and control groups for word-attack skills are shown in Table X.

TABLE X
SUMMARY OF THE DATA RELATED TO THE MEAN SCORES
OF THE WORD-ATTACK SKILLS FOR GIRLS

| Group | N | Mean I. Q. | Post-Test Mean Score | S.D. | Adjusted Mean |
|--------------|----|------------|----------------------|-------|---------------|
| Control | 26 | 113.85 | 79.46 | 11.69 | 71.64 |
| Experimental | 45 | 100.76 | 73.27 | 10.09 | 77.79 |

Before adjustment, the mean score of the experimental

group was less than the mean score of the control group by 6.20. After adjustment was made for I. Q. and pre-test, the mean score of the experimental group was greater than the mean score of the control group by 6.15.

It was stated that in Hypothesis IV, Part A, the girls in the experimental group would achieve a significantly greater mean word-attack skills score on the Stanford Diagnostic Reading Test than would the girls in the control group. Hypothesis IV, Part A, in effect, is whether the stated difference in the adjusted mean is significant. The results of the analysis-of-covariance in which the difference was tested is shown in Table XI below.

TABLE XI

SUMMARY OF THE ANALYSIS-OF-COVARIANCE OF THE WORD-ATTACK SKILLS SCORES FOR GIRLS

N = 71

| Source | Degrees of Freedom | Sum of Square | Mean Square | F Value | Probability |
|------------|--------------------|---------------|-------------|---------|-------------|
| Total | 68 | 3263.70 | --- | --- | --- |
| Within | 67 | 2861.83 | 42.71 | --- | --- |
| Difference | 1 | 401.87 | 401.87 | 9.41 | 0.0031 |

The obtained F-ratio of 9.41 was acceptable at the .01 level of significance; therefore, Hypothesis IV, Part A was supported. This would indicate that the girls in the experimental group did demonstrate a statistically greater mean increase in word-attack skills than did the girls in the control group.

Data Related to Hypothesis Four, Part B

Research Hypothesis IV, Part B was restated to read:

There will be no significant difference in the mean gains of the experimental and control group boys for word-attack skills. The mean scores of the boys in the experimental and control groups for word-attack skills are shown in Table XII below. It can be noted in this table that before adjustment the mean score of the boys in the experimental group was greater than the mean score of the boys in the control group by 1.91. After adjustment was made for I. Q. and pre-test, the mean score of the boys in the experimental group was greater than the mean score of the boys in the control group by 8.11.

TABLE XII

SUMMARY OF THE DATA RELATED TO THE MEAN SCORES OF THE WORD-ATTACK SKILLS SCORES FOR BOYS

| Group | N | Mean I. Q. | Post-Test Mean Score | S.D. | Adjusted Mean |
|--------------|----|------------|----------------------|-------|---------------|
| Control | 27 | 107.59 | 76.44 | 12.32 | 73.59 |
| Experimental | 23 | 102.87 | 78.35 | 10.59 | 81.70 |

It was stated that in Hypothesis IV, Part B, the boys in the experimental group would achieve a significantly greater mean word-attack skills score on the Stanford Diagnostic Reading Test than would the boys in the control group. Hypothesis IV, Part B, in effect, is whether the above difference in the adjusted mean is significant. The results of the analysis-of-covariance in which the difference was tested is shown in Table XIII on the following page.

TABLE XIII

SUMMARY OF THE ANALYSIS-OF-COVARIANCE OF THE
WORD-ATTACK POST-TEST SCORES FOR BOYS

N = 50

| Source | Degrees of Freedom | Sum of Square | Mean Square | F Value | Probability |
|------------|--------------------|---------------|-------------|---------|-------------|
| Total | 47 | 3158.73 | --- | --- | --- |
| Within | 46 | 2431.28 | 52.85 | --- | --- |
| Difference | 1 | 727.45 | 727.45 | 13.76 | 0.0006 |

In Table XIII the obtained F-ratio of 13.76 was acceptable at the .01 level of significance; therefore, Hypothesis IV, Part B was supported. This would indicate that the boys in the experimental group did demonstrate a statistically greater mean increase in word-attack skills than did the boys in the control group.

Summary of Findings

In this chapter have been presented the statistical data comparing the effects of Instructional Objectives Exchange criterion-referenced tests utilized in a basal reader program for fifth-grade students. An analysis-of-covariance was made on the effects of these IOX reading materials on the reading comprehension and word-attack skills of the experimental group as measured by the Stanford Diagnostic Reading Test. Two of the four hypotheses tested were accepted.

There were two hypotheses which dealt with the relationship of boys vs. boys and girls vs. girls in the mean gain in reading comprehension and word-attack skills. The results of

the tests of the hypotheses considered in this study can be summarized briefly in the following manner:

1. The experimental group did not have a significantly greater mean comprehension score on the Stanford Diagnostic Reading Test than did the control group. Therefore, Hypothesis I for this study was rejected.

2. The experimental group had a significantly greater mean word-attack skills score on the Stanford Diagnostic Reading Test than did the control group. Hypothesis II for this study was, therefore, accepted.

3. When a comparison was made between the mean comprehension scores of the boys and girls of the experimental group with those of the control group, no statistically significant difference was found. Hypothesis III, Parts A and B, was rejected.

4. When the mean word-attack skills scores of the boys and girls were compared on the Stanford Diagnostic Reading Test, a statistically significant difference was found. The difference favored the boys and girls of the experimental group. Hypothesis IV, Parts A and B, as stated in this study was accepted.

CHAPTER V

SUMMARY, FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

In spite of the numerous articles written and considerable research studies accomplished, educators are still uncertain as to how an individual child learns to read. Educators do agree that reading is a complex process not composed of a single skill, but consisting of numerous skills which are developed over a period of years (8). Reading deficiencies seem to be universal and educators are concerned about the formulation of programs which will "foster reading interest and abilities, prevent reading difficulties, and aid in the diagnosing and correcting problems that do occur." Yet, educators are still seeking answers that will allow them to develop adequate reading programs (12). Because of the limited amount of research effectuated in regard to the IOX reading collections of comprehension and word-attack skills, this study was conducted to determine the effectiveness of these materials used as supplementary lessons in a basal reader program. This chapter contains a summary of the methods and procedures used to obtain the data, a review of the findings, a discussion of the findings, conclusions, and recommendations.

Summary of Methods and Procedures

The purpose of this study was to investigate the relative

effectiveness of the Instructional Objectives Exchange criterion-referenced test collections of comprehension, and word-attack skills incorporated into a basal reader program for fifth grade students. Four classrooms, designated as the experimental group, and two classrooms, selected to be the control group, participated in this study. Within the experimental group were sixty-eight subjects, twenty-three boys and forty-five girls. The control group had fifty-three subjects, of which twenty-seven were boys and twenty-eight were girls.

The experimental and control groups were administered pre-tests in February, 1974, using the Stanford Diagnostic Reading Test, Form X, Level II. The two groups were equated by the analysis-of-covariance, using I. Q. and pre-test scores as the covariates. The experimental group was provided with reading criterion-referenced tests which were developed by the IOX while the control group received reading lessons which were purchased from commercial educational companies. Both groups continued their lessons from the basal reader and used the additional reading lessons as supplementary materials.

Following ten weeks of instruction, a post-test, using the Stanford Diagnostic Reading Test, Form X, Level II, was administered to the experimental and control groups. Comparisons of the mean score gains for reading comprehension and word-attack skills were made between the following:

1. The experimental and control groups
2. The girls in the experimental and control groups
3. The boys in the experimental and control groups.

All hypotheses were stated in null form and statistically tested for significance of difference. All data collected were punched on cards for processing at the North Texas State University Computer Center.

Review of the Findings

It was hypothesized that the group of students who were in the experimental group during the ten-week study period, would demonstrate significantly greater development in reading comprehension, and word-attack skills when compared to a control group of children also participating in the experimental study. The following hypotheses were formulated and investigated by statistical analysis for this study:

1. Students participating in the experimental group will achieve a significantly greater mean reading comprehension score as measured by the Stanford Diagnostic Reading Test than will students in the control group. There was no significant difference in the reading comprehension scores between the two groups. Hypothesis I was therefore rejected.

2. Students in the experimental group will achieve a significantly greater mean word-attack skills score as measured by the Stanford Diagnostic Reading Test than will students in the control group. A statistically significant difference at the .01 level of significance was found in favor of the experimental group. Hypothesis II was accepted as stated for this comparison.

3. Boys and girls in the experimental group will achieve a significantly greater mean reading comprehension score as measured by the Stanford Diagnostic Reading Test than will the

boys and girls in the control group. There was no significant difference in the reading comprehension scores. Hypothesis III, Part A and B, was therefore rejected.

4. Boys and girls who participated in the experimental group will achieve a significantly greater mean word-attack skills score as measured by the Stanford Diagnostic Reading Test than will the boys and girls in the control group. The statistical difference between the word-attack skills scores of the boys and girls in the experimental group and the boys and girls in the control group was significant at the .01 level of significance for boys and girls. The difference was in favor of the experimental group. Hypothesis IV, Part A and B, was accepted as stated for this comparison.

Discussion of the Findings

Utilizing the IOX criterion-referenced tests for word-attack skills resulted in general improvement in this area by a majority of the experimental groups. This finding is consistent with the research and reports of Emans (5), Agnew (1), and Templin (13), which show that word-attack skills are needed to help provide visual and auditory clues in word recognition. Reports by Clymer (3), and Otterman (11) differed from the findings in this study in that they concluded that many of the phonic generalizations utilized in the schools are of limited value and that some aspects of word-attack are very difficult for the slow-learner.

The lack of significant results in regard to the

comprehension skill may be traced to the utilization of the standardized reading test which tends to measure student's comprehension ability in terms of literal comprehension only. With norm-referenced tests it is difficult to assess the student's ability in the various comprehension areas (7). Early research studies concerning comprehension were descriptive in nature, and very little progress was accomplished in isolating and discovering the "independent variables present in the comprehension process" (2). Feder (6) in a more recent study, identified "three gradations of comprehension: information, information, inference, and appreciation." It is therefore possible that the mental processes involved in the comprehension of students was not completely measured by the Stanford Diagnostic Reading Test.

Educators have been aware for a long period of time that sex differences affect reading achievement. Research studies have been conducted to determine the sex differences in learning (4). Knoski (9), McNeil (10), and Weintraub (14) have conducted studies which indicate the disparity between boys and girls in reading achievement. An examination of the scores of the girls vs. girls and the boys vs. boys was to determine which reading materials produced any significant gains. Another reason for the comparison was to recommend the utilization of materials which will help both sexes perform satisfactorily.

Conclusions

Based on the data presented in this experimental study and within the limitations of this study, the following

conclusions seem to be justified:

1. In this study the findings indicate that the use of IOX criterion-referenced tests as practice materials in contrast to other published materials is likely to result in a more significant increase in knowledge of word-attack skills by fifth grade boys and girls.

2. The findings do not support the superiority of the IOX materials similarly used for comprehension skills. Two explanations are suggested here: (a) comprehension skills are inherently more complex than word-attack skills and any differences due to different instructional treatments would take longer to appear than the time covered in this study, or (b) that norm-referenced tests such as the Stanford Diagnostic Reading Test used in this study are not sufficiently sensitive to detect growth in the numerous comprehension skills contained in the IOX materials.

Recommendations

The following recommendations are derived from the analysis of the data collected in this study and the findings of related research:

1. Should this study be repeated, the number of weeks should be increased in an attempt to provide a suitable length of presentation time for the IOX reading comprehension materials.

2. A study to determine whether or not the subject's sex is a statistically significant factor could be accomplished by a comparison of the scores of the boys vs. the girls in the two groups.

3. If a similar experimental study is conducted, consideration should be given to the type of grouping used. There is always the possibility of extreme variability within intact groups which may render the statistical analysis insignificant.

4. In-service training should be instituted for all teachers to assist them in utilizing IOX objectives-based materials.

5. Further experimentation should be done comparing the effects of IOX materials in the other grades of the elementary school, to determine the optimum grade level at which such materials might be taught.

CHAPTER BIBLIOGRAPHY

1. Agnew, Donald C., "The Effect of Varied Amounts of Phonetic Training on Primary Reading," Duke University Research Studies in Education, No. 5, Durham, North Carolina, Duke University Press, 1939.
2. Chester, Robert D., "The Psychology of Reading," The Journal of Educational Research, LXVII, (May-June, 1974), 404.
3. Clymer, Theodore, "The Utility of Phonic Generalizations in the Primary Grades," The Reading Teacher, XVI, (January, 1963), 252-260.
4. Criscuolo, Nicholas, "Sex Influences on Reading," The Reading Teacher, XXXI, (May, 1968), 764.
5. Emans, Robert, "The Usefulness of Phonic Generalizations Above the Primary Grades," The Reading Teacher, XX, (February, 1967), 419-425.
6. Feder, D. D., "Comprehension Maturity Tests--A New Technique in Mental Measurement," Journal of Educational Psychology, XXIX, (November, 1938), 597-606.
7. Guszak, Frank J., "Teacher Questioning and Reading," The Reading Teacher, XXI, (December, 1967), 227-234.
8. Hackney, Ben H., "Reading Achievement and Word Recognition Skills," The Reading Teacher, XXI, (March, 1968), 515.
9. Konski, Virginia, "An Investigation into Differences Between Boys and Girls in Selected Reading Readiness Areas and in Reading Achievement," Unpublished doctoral dissertation, University of Missouri, 1951.
10. McNeil, J. D., "Programmed Instruction Versus Classroom Procedures in Teaching Boys to Read," American Educational Research Journal, I, 1964, 113-119.
11. Otterman, Lois M., "The Value of Teaching Prefixes and Word Roots," Journal of Educational Research, XLVIII, (April, 1951), 611-616.

12. Shepherd, George, "Reading Research and the Individual Child," The Reading Teacher, XXI, (January, 1968), 335.
13. Templin, Mildred C., "Phonic Knowledge and Its Relation to the Spelling and Reading Achievement of Fourth Grade Pupils," Journal of Educational Research, LXVII, (February, 1954), 441-454.
14. Weintraub, S., "What Research Says to the Reading Teacher," The Reading Teacher, XX, 1966, 155-165.

IDENTIFYING WORDS WITH SILENT LETTERS.

WORD ATTACK (Test 33, Form A)

Name _____

an



objectives-based test

DIRECTIONS.


Put an "X" next to the word in each row that has a letter which is not pronounced.
The example is done for you.

EXAMPLE: a) right b) rats c) ranch

-
- | | | | |
|-----|------------------------------------|----------------------------------|-----------------------------------|
| 1. | <input type="checkbox"/> a) comb | <input type="checkbox"/> b) bum | <input type="checkbox"/> c) brush |
| 2. | <input type="checkbox"/> a) new | <input type="checkbox"/> b) knit | <input type="checkbox"/> c) fish |
| 3. | <input type="checkbox"/> a) king | <input type="checkbox"/> b) no | <input type="checkbox"/> c) know |
| 4. | <input type="checkbox"/> a) dumb | <input type="checkbox"/> b) run | <input type="checkbox"/> c) rub |
| 5. | <input type="checkbox"/> a) skin | <input type="checkbox"/> b) sin | <input type="checkbox"/> c) sign |
| 6. | <input type="checkbox"/> a) night | <input type="checkbox"/> b) nod | <input type="checkbox"/> c) noon |
| 7. | <input type="checkbox"/> a) plain | <input type="checkbox"/> b) pack | <input type="checkbox"/> c) star |
| 8. | <input type="checkbox"/> a) writer | <input type="checkbox"/> b) wash | <input type="checkbox"/> c) war |
| 9. | <input type="checkbox"/> a) two | <input type="checkbox"/> b) top | <input type="checkbox"/> c) trip |
| 10. | <input type="checkbox"/> a) kit | <input type="checkbox"/> b) knot | <input type="checkbox"/> c) ton |
-

DERIVING MEANING OF AN UNFAMILIAR WORD GIVEN MORE THAN ONE
POSSIBLE DEFINITION
CONTEXT CLUES (Test 9, Form A)

Name _____

an  objectives-based test**DIRECTIONS.**

Read the sentences. Then mark an "X" next to the best definition for the underlined word. The example is done for you.

EXAMPLE: "Pull him up. We're losing slack in the port line," the captain called. "Put it on the surge and get him out of the water."

- Surge means: ___ a) swelling, rolling wave
 ___ b) sudden rise or fall
 X c) part of a machine for raising heavy weights

1. The company president thought the problem was important and should be settled immediately. He put it on the docket to be discussed at tomorrow's meeting.

- Docket means: ___ a) written summary of a document
 ___ b) calendar of business matters
 ___ c) list of things postponed

2. The mountain road was serpentine. It almost made Mother ill to drive on it because she slid from one side of the car to the other.

- Serpentine means: ___ a) dull and greenish
 ___ b) winding and turning
 ___ c) sly

3. She not only did what they asked of her; she did more. No wonder they decided to jump her into a better job.

- Jump means: ___ a) escape or run from
 ___ b) spring into the air
 ___ c) put ahead or advance

4. Every night he counted his store. Was it because he thought he might have been robbed, or was it that he enjoyed looking at the glittering stuff?

- Store means: ___ a) treasure owned by a person
 ___ b) place where a person buys things
 ___ c) food put away for an emergency

5. He hammered at the idea until I said that I agreed with him.

- Hammered means: ___ a) beat into a shape or flattened
 ___ b) forced upon by constant repetition
 ___ c) struck, hit

APPENDIX C

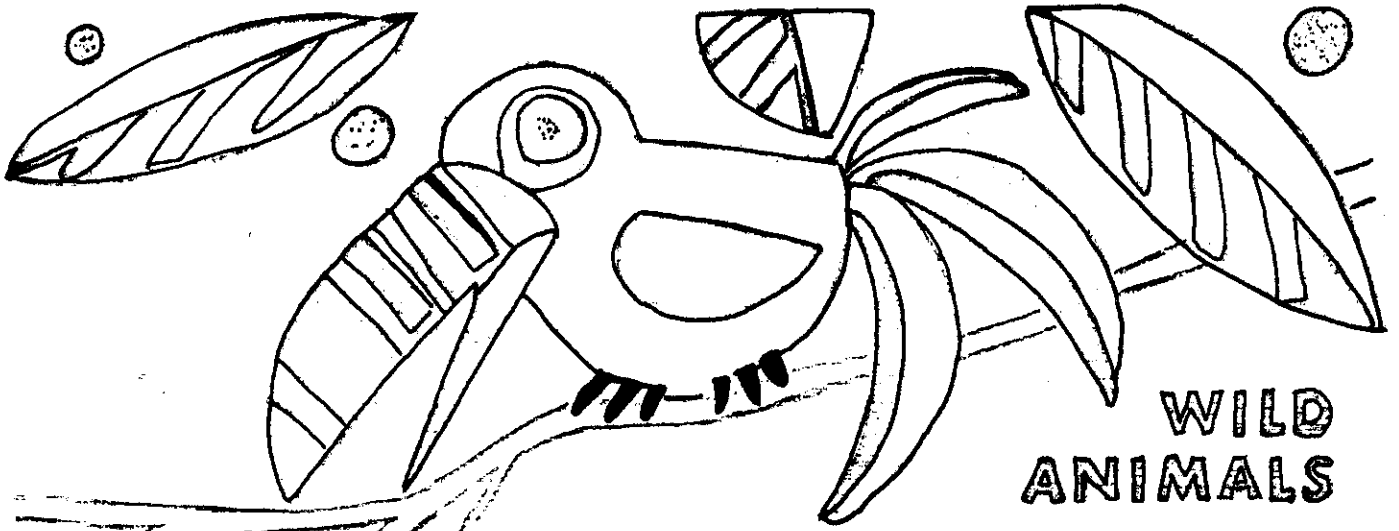
Table of Contents -- Reading: Word Attack Skills (Grades K-6)

| <u>Test Number</u> | <u>Test Title</u> | <u>Test Number</u> | <u>Test Title</u> |
|--------------------|--|--------------------|---|
| 1. | Matching Similar Letter Combinations | 20. | Identifying Words Containing Beginning Consonant Blends |
| 2. | Matching Similar Letter Orderings | 21. | Differentiating Long from Short Vowels |
| 3. | Matching Letter Combinations | 22. | Selecting Similar Sounds of R-Controlled Vowels |
| 4. | Initial Consonant Sounds (Pictures) | 23. | Matching Sounds of Diphthongs |
| 5. | Matching Rhyming Words (Pictures) | 24. | Pronouncing Unfamiliar Letter Combinations of the CVC Pattern |
| 6. | Matching Final Consonant Sounds (Pictures) | 25. | Matching Vowel Sounds in CVCe Letter Combinations |
| 7. | Matching Medial Vowel Sounds (Pictures) | 26. | Matching Vowel Sounds to Words with Two Consecutive Vowels |
| 8. | Combining Sounds to Form Words | 27. | Identifying Consonant Digraphs in Words |
| 9. | Matching Beginning Sounds to Written Single Letters (Pictures) | 28. | Selecting Correct Word Forms |
| 10. | Matching Final Consonant Sounds to Written Single Letters (Pictures) | 29. | Using Words Having Irregular Plural Forms |
| 11. | Matching Beginning Sounds to Written Consonant Blends (Pictures) | 30. | Choosing Homonyms to Complete Sentences |
| 12. | Matching Rhyming Words | 31. | Pronouncing Basic Sight Words (List II) |
| 13. | Matching Sounds of Short Vowels (Pictures) | 32. | Finding Beginning Three-Letter Consonant Blends in Spoken Words |
| 14. | Matching Sounds of Simple Consonant Digraphs (Pictures) | 33. | Identifying Words with Silent Letters |
| 15. | Recognizing Compound Words | 34. | Distinguishing Meaning of Accented Homographs |
| 16. | Identifying Meaning of Simple Contractions | 35. | Determining Number of Syllables |
| 17. | Identifying Base Words | 36. | Selecting Words with Affixes |
| 18. | Pronouncing Basic Sight Words (List I) | 37. | Pronouncing Unfamiliar Letter Combinations Using Phonic Generalizations |
| 19. | Selecting Pairs of Similar Consonant Variants | 38. | Pronouncing Basic Sight Words (List III) |

Table of Contents -- Reading Comprehension Skills

| <u>Test Number</u> | <u>Test Title</u> | <u>Test Number</u> | <u>Test Title</u> |
|--------------------|---|-------------------------------|---|
| MAIN IDEA | | CONTEXT CLUES | |
| 1. | Finding Common Attributes | 1. | Subject Pronoun Referents -- Matching Pictures to Pronouns |
| 2. | Identifying Main Ideas of Pictures | 2. | Subject Pronoun Referents -- Matching Nouns to Pronouns |
| 3. | Identifying the Most General Statement | 3. | Subject and Object Pronoun Referents |
| 4. | Finding the Moral | 4. | Subject, Object, and Demonstrative Pronoun Referents |
| 5. | Finding the Factual Generalization | 5. | Indefinite, Demonstrative, and Possessive Pronoun Referents |
| CONCLUSIONS | | 6. | Selecting Examples and Synonyms to Match Contextual Definitions |
| 1. | Identifying the Probable Outcomes of a Pictured Situation | 7. | Deriving Meaning of Unfamiliar Letter Combinations |
| 2. | Simple Logical Reasoning | 8. | Deriving the Meaning of an Unfamiliar Word |
| 3. | Identifying the Cause of a Simple Event | 9. | Deriving Meaning of an Unfamiliar Word Given More than One Possible Definition |
| 4. | Identifying the Cause of a Detailed Event | PUNCTUATION | |
| 5. | Identifying Situations from Information | 1. | Matching Punctuation with Intonation |
| 6. | Identifying Causal Relationships Based on Agreement and Difference | 2. | Identifying Sentences with Ellipse, Titles, Restrictive and Non-Restrictive Clauses, and Emphasis |
| 7. | Identifying Possible Outcomes | 3. | Determining Sentence Meaning from Comma, Italics, Quotation and other Marks |
| 8. | Drawing Conclusions about Occupation, Character, and Locality | SYNTACTICAL STRUCTURES | |
| 9. | Making Deductions from Factual Information | 1. | Understanding Explicitly-Stated Content |
| 10. | Identifying the Writer's Point of View | 2. | Selecting Opposites of Negative Statements |
| SEQUENCE | | 3. | Interpreting Passive and Active Voice |
| 1. | Determining First and Last in Temporal Sequence | 4. | Interpreting Complex Sentences |
| 2. | Ordering Events and Instructions | AFFIXES | |
| 3. | Determining Sequence on the Basis of Logical Dependency | 1. | Prefixes |
| 4. | Determining Sequence from Tense and Words that Signal Order | 2. | Suffixes |
| 5. | Arranging Sentences in Narrative Order | | |
| 6. | Rearranging Sentences from Flash-Back or Flash-Ahead Sequence to Chronological Sequence | | |
| 7. | Identifying the Dramatic Functions of Paragraphs | | |

APPENDIX D



WILD
ANIMALS

1 LEAF - AF + OPAL - AL + CARD - C =

2 R + CABBAGE - C + SIT - AGES =

3 DENT + BEER - BENT =

4 S + NAB + RAKE - BAR =

5 BEET + WAR - WET =

6 SLIDE + E - SEED + TOP + AN - PAT =

7 BALL + I + GATE - BE + OR =

8 FEET + BOX - BEET =

9 T + SIGH + EARN - NASH =

10 WON + LAP - NAP + F =

11 BAND + GOES + OR - SOON =

12 MOON + SEAT - TAN =

13 SKIP + FUN + RAKE - FIRE - PA =

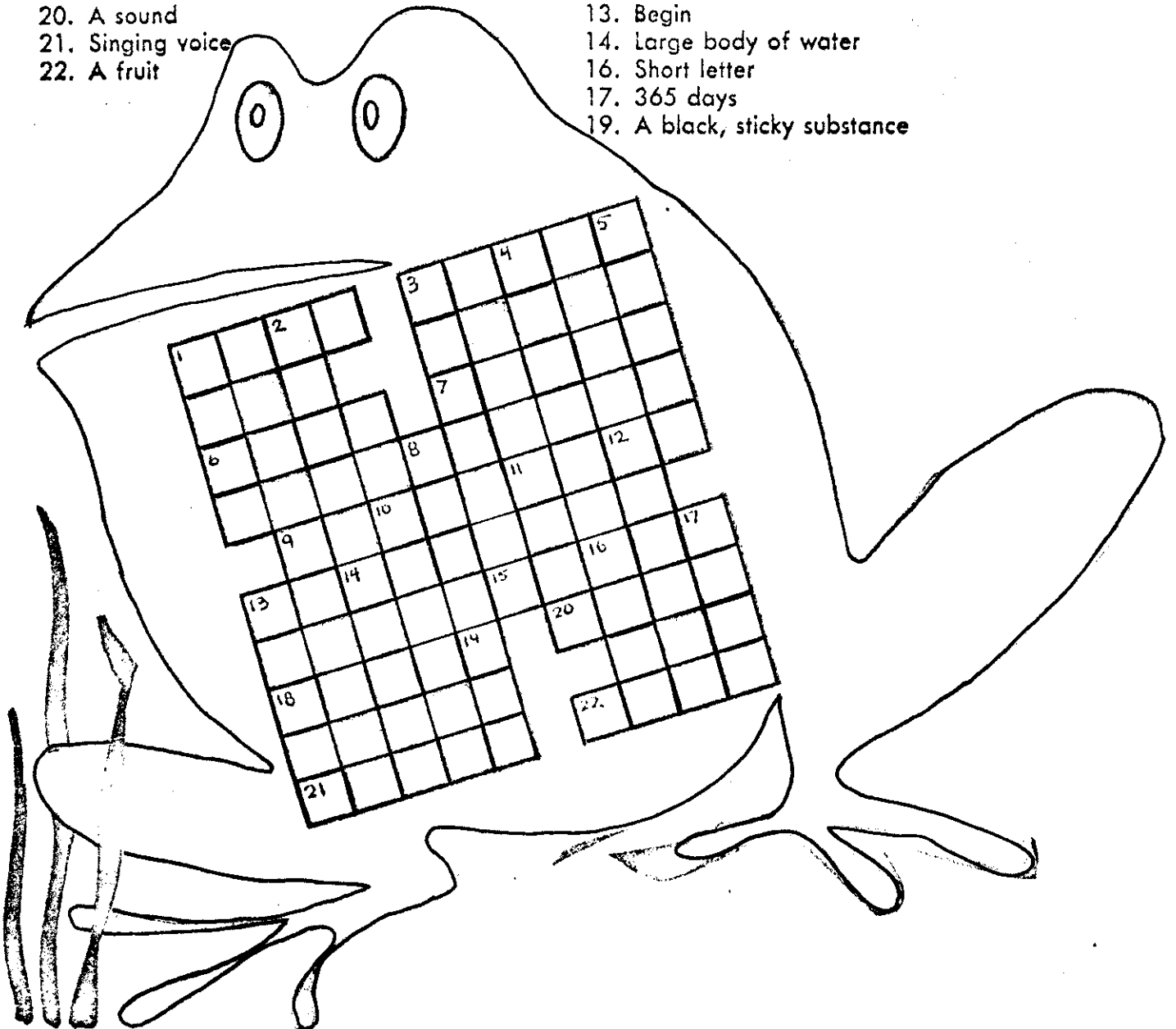
APPENDIX E

ACROSS

1. Stout cord
3. Goes with egg
6. Holds ice cream
7. Waterway
8. Not out
10. Not off
11. Turn loose
13. Long-legged bird
15. What we buy things with
18. F.B.I. man
20. A sound
21. Singing voice
22. A fruit

DOWN

1. Big stone
2. Small horse
3. Farm building
4. Put a lid on
5. Doctor's helper
8. Writing fluid
9. On or near
10. Else
11. 12 inches
12. Level
13. Begin
14. Large body of water
16. Short letter
17. 365 days
19. A black, sticky substance



PLEASE NOTE:

Page 80, Appendix F, "Reference Tools and Study Skills", copyright 1972 by Milliken Publishing Co., not microfilmed at request of author. Available for consultation at North Texas State University Library.

UNIVERSITY MICROFILMS

Name _____ Date _____

READ FOR DETAILS

Read each selection below carefully. Then cover the selection with a piece of paper and answer the questions **without looking back** in the paragraph. Remember, you are reading for details.

I. Rudyard Kipling was born in Bombay, India, in 1865. Although he went to school in England, he spent most of his life in India. He married an American girl and lived in Vermont for 4 years. He won the Nobel Prize for literature in 1907. Using India as a background, he wrote stories, poems, and novels. Among his most famous works are *Kim*, *Captains Courageous*, *The Light That Failed*, *Just So Stories*, *Jungle Book*, and *Plain Tales from the Hills*. He died in 1931.

1. Where was Kipling born? _____
2. Did he ever live in America? _____
3. Where did he go to school? _____
4. What award did he receive? _____
5. Name a book he wrote. _____

II. Alberta lived in Harlem. Her home was a cold-water flat on the third floor. She shared a bedroom with two sisters, but her brother had a room of his own. "The lucky devil!", thought Alberta. She hated always being under the inquisitive eyes of her two younger sisters who were continually asking her, "What are you doing?" or pestering her to let them use her lipstick or perfume. In one more month Alberta would be 18, and Mom had promised she could move to a place of her own. How she waited for that day!

Complete the following sentences by underlining the correct word.

1. Alberta lived in (Manhattan, Harlem, Brooklyn).
2. Alberta had (2 brothers and 1 sister, 2 sisters and 1 brother, 1 sister and 1 brother).
3. Alberta lived on the (ground, third, fifth) floor.
4. She lived in a (house, hotel, flat).
5. Alberta wanted a (room, book, friend) of her own.
6. When she was (18, 20, 21) she would move out.
7. Alberta was continually being pestered by her (brothers, mother, sisters).

III. It was a dull, gray November day. No clouds dotted the dreary sky. The recent snow, having melted, left dirty patches and muddy spots. The blustery wind blew dead, brown leaves into nooks where they settled like lonely children. Bits of newspaper, discarded paper cups, and other debris swirled in the air. Few people were about. Sandy's mood fit the day perfectly.

Write **T** for true statements; **F** for false statements.

1. It was a dull, spring day. _____
2. The sky was dotted with clouds. _____
3. A strong wind blew leaves and papers about. _____
4. It had snowed recently. _____
5. There were few people on the street. _____
6. Although the day was dull and gray, Sandy was happy _____
7. This story gives you a gloomy feeling. _____

APPENDIX G

(Excerpt from Reading: Word Attack Skills K-6 Test Manual)

TEST # 6. MATCHING FINAL CONSONANT SOUNDS (PICTURES)

Objective: Given sets of three pictures of objects, one of which is a sample, the student will be able to mark the alternative picture whose name has the same final consonant sound as the sample.

Sample Item:

Directions. Look at the picture in the box. Say the name of the picture softly to yourself. Then mark an "X" through the picture in each row whose name ends like the name of the picture in the box.

Example:



a)



b)



Amplified Objective:

Testing Situation.

1. The student will be given sets of three pictures of objects, one of which is a sample.
2. The sample picture will have a one-syllable name familiar to the student and a final consonant sound which is not a blend.
3. The student will mark the alternative picture whose name has the same final consonant sound as the sample.

Response Alternatives.

1. All alternative pictures will have one-syllable names familiar to the student. No pictures whose name ends with a consonant blend will be used.
2. All alternative pictures will correspond with the sample in one of the following aspects: initial consonant sound, final consonant sound, or approximate medial vowel.

Criterion of Correctness. The correct answer will be a picture whose name ends with the same consonant sound as the sample.

APPENDIX H

(Excerpt from Reading: Comprehension K-6 Test Manual)

TEST # 4. FINDING THE MORAL

Objective: The student can select the statement which expresses the moral of a one paragraph selection.

Sample Item:

Directions. Read each paragraph. Then put an "X" next to the answer which states the moral of the story.

Example: John and Sam were both hunters. They awakened one morning to find rabbits in their traps. John said that he would use the rabbit to feed his family. Sam said, "John, you are foolish. If you let the rabbit go, he will lead you to the rabbit hole. Then you can shoot several and have a lot of meat." John didn't like the idea and told Sam to carry out his plan by himself. As he cooked his rabbit, John watched Sam follow his rabbit into the woods. Later in the day, Sam returned. He had no rabbit at all and no food to feed his family that night.

- a) Having a family to feed is a problem.
- b) Rabbits are difficult to trap.
- c) Having some is better than risking having none.

Amplified Objective:

Testing Situation.

1. The student will be given a one paragraph selection and several statements.
2. The paragraph may contain simple, compound and complex sentences in both active and passive voice.
3. The paragraph should not exceed eight sentences in length.
4. Paragraphs will be narrative and may or may not include dialogue.
5. The student will select the one statement which expresses the moral of the paragraph.

Response Alternatives.

1. There will be three alternatives for each item.
2. Distractors may be of the following types:
 - a. a moral which can not be substantiated by the information in the selection.
 - b. a moral which applies to information in the passage which is irrelevant to the main concern of the passage and thus to the moral.
 - c. a statement or paraphrase from the paragraph usually not having universal application.

Criterion of Correctness. The correct answer will be the most encompassing moral which can be substantiated by the information given in the paragraph. It must identify the main idea or point of the paragraph and will go beyond the specifics of the fictional situation to the universal implications of the human behavior presented.

APPENDIX I

STUDENTS AVAILABLE, THOSE ELIGIBLE, THOSE LOST,
AND THOSE WHO COMPLETED

| Group Number | Total | | Eligible | | Exper. Mortality | | N Complete | |
|--------------------|-------|----|----------|----|------------------|---|------------|----|
| | B | G | B | G | B | G | B | G |
| Control | | | | | | | | |
| 11 | 14 | 18 | 14 | 14 | 0 | 2 | 14 | 12 |
| 12 | 17 | 16 | 13 | 16 | 0 | 2 | 13 | 14 |
| TOTAL | 65 | | 57 | | 4 | | 53 | |
| Exper. | | | | | | | | |
| 21 | 16 | 14 | 12 | 9 | 1 | 1 | 11 | 8 |
| 22 | 11 | 17 | 6 | 14 | 3 | 3 | 3 | 11 |
| 23 | 15 | 18 | 5 | 15 | 1 | 3 | 4 | 12 |
| 24 | 13 | 19 | 8 | 14 | 3 | 0 | 5 | 14 |
| TOTAL | 123 | | 83 | | 15 | | 68 | |
| GRAND TOTAL | 188 | | 140 | | 19 | | 121 | |

APPENDIX J

PRE-TEST - POST-TEST READING COMPREHENSION SCORES
FOR CONTROL AND EXPERIMENTAL GROUPS

| Reading Comprehension | Group | | Mean I. Q. | Pre-Test Unadjusted Mean | Post-Test Achieve. Mean |
|-----------------------|---------|--------|------------|--------------------------|-------------------------|
| | Control | Exper. | | | |
| Control Total | N = 53 | --- | 110.66 | 40.55 | 44.06 |
| Exper. Total | --- | N = 68 | 101.47 | 32.04 | 39.46 |
| Girls Control | N = 26 | --- | 113.85 | 43.15 | 47.42 |
| Girls Exper. | --- | N = 45 | 100.76 | 31.60 | 39.93 |
| Boys Control | N = 27 | --- | 107.59 | 38.04 | 40.81 |
| Boys Exper. | --- | N = 23 | 102.87 | 32.91 | 38.52 |

PRE-TEST - POST -TEST WORD-ATTACK SKILLS SCORES
FOR CONTROL AND EXPERIMENTAL GROUPS

| Word-Attack Skills | Group | | Mean I. Q. | Pre-Test Unadjusted Mean | Post-Test Achieve. Mean |
|--------------------|---------|--------|------------|--------------------------|-------------------------|
| | Control | Exper. | | | |
| Control Total | N = 53 | --- | 110.66 | 74.68 | 77.92 |
| Exper. Total | --- | N = 68 | 101.47 | 60.74 | 74.99 |
| Girls Control | N = 26 | --- | 113.85 | 75.50 | 79.46 |
| Girls Exper. | --- | N = 45 | 100.76 | 58.27 | 73.27 |
| Boys Control | N = 27 | --- | 107.59 | 73.89 | 76.44 |
| Boys Exper. | --- | N = 23 | 102.87 | 65.57 | 78.35 |

BIBLIOGRAPHY

Books

- Buros, Oscar K., The Seventh Mental Measurements Yearbook, Highland Park, New Jersey, The Gryphon Press, 1972.
- Ferguson, George, Statistical Analysis in Psychology and Education, New York, McGraw-Hill Book Company, 1971.
- Fox, David, The Research Process in Education, New York, Rinehart and Winston, Inc., 1969.
- Good, Carter, Dictionary of Education, Third Edition, New York, McGraw-Hill Book Company, 1973.
- Hodges, Richard, "Some Assumptions About Behavioral Objectives as Related to Reading," Accountability and Reading Instruction: Critical Issues, edited by Robert Ruddell, Urbana, Illinois, National Council of Teachers of English, 1973.
- Lindvall, C. M., and Cox, Richard, The IPI Evaluation Program Chicago, Illinois, Rand McNally and Company, 1970.
- Mager, R. F., Preparing Instructional Objectives, Palo Alto, California, Fearon Publishing Company, 1962.
- Olson, Arthur, "Reading: The Approaches We Use," Reading and the Elementary School Child: Selected Readings on Programs and Practices, edited by Virgil Howes and Helen Darrow, New York, Macmillan Company, 1968.
- Popham, W. James, and Baker, Eva, Planning an Instructional Sequence, Englewood Cliffs, New Jersey, Prentice-Hall Inc., 1970.
- Shanner, W., Systems in Education, Los Angeles, California, Daniel, Mann, Johnson, and Mendenhall, 1968.
- Williams, Lois, Independent Learning in the Classroom, Washington, D. C., American Association of Elementary-Kindergarten-Nursery Education, NEA Center, 1969.
- Zintz, Miles, V., The Reading Process: The Teacher and the Learner, Dubuque, Iowa, William C. Brown Company, 1970.

Articles

- Airasian, F., and Madaus, G., "Criterion Referenced Testing in the Classroom," Measurement in Education, III (March, 1972), 1-8.
- Baker, Eva, "Effects on Student Achievement of Behavioral and Non-behavioral Objectives," The Journal of Experimental Education, XXXVII (Summer, 1969), 5-8.
- Blaney, J. P., and McKie, D., "Knowledge of Conference Objectives and Effect Upon Learning," Adult Education Journal, XIX (Summer, 1969), 98-105.
- Chester, Robert, D., "The Psychology of Reading," The Journal of Educational Research, LXVII (May-June, 1974), 404.
- Coulson, J. E., and Cogswell, J. F., "Effects of Individualized Instruction on Testing," Journal of Educational Measurement, II (June, 1965), 59-64.
- Clymer, Theodore, "The Utility of Phonic Generalizations in the Primary Grades," The Reading Teacher, XVI (January, 1963), 252-260.
- Criscuolo, Nicholas, "Sex Influences on Reading," The Reading Teacher, XXXI (May, 1968), 764.
- Dalis, G. T., "Effect of Precise Objectives Upon Student Achievement in Health Education," The Journal of Experimental Education, XXXIX (Fall, 1970).
- Ebel, Robert, "Content Standard Test Scores," Educational and Psychological Measurement, XXII (Spring, 1962), 15-25.
- Eisner, E. W., "Educational Objectives: Help or Hindrance," School Review, LXXV (Autumn, 1967), 260.
- Emans, Robert, "The Usefulness of Phonic Generalizations Above the Primary Grades," The Reading Teacher, XX (February, 1967), 419-425.
- Feder, D. D., "Comprehension Maturity Tests--A New Technique in Mental Measurement," Journal of Educational Psychology, XXIX (November, 1938), 597-606.
- Glaser, Robert, "Instructional Technology and the Measurement of Learning Outcomes: Some Questions," American Psychologist, (August, 1963), 519-521.
- Gusak, Frank J., "Teacher Questioning and Reading," The Reading Teacher, XXI (December, 1967), 227-234.
- Hackney, Ben H., "Reading Achievement and Word Recognition Skills," The Reading Teacher, XXI (March, 1968), 515.

- Rothkopf, E. Z., "The Concept of Matemagenic Activities," Review of Educational Research, XL (June, 1970), 325-336.
- Shepherd, George, "Reading Research and the Individual Child," The Reading Teacher, XXI (January, 1968), 335.
- Stake, Robert, "1: Objectives, Priorities, and Other Judgement Data," Review of Educational Research, Washington, D. C., Education Research Association, XL (April, 1970), 181-212.
- Templin, Mildred, "Phonic Knowledge and Its Relation to the Spelling and Reading Achievement of Fourth Grade Pupils," Journal of Educational Research, LXVII (February, 1954), 441-454.
- Trow, William Clark, "On Marks, Norms, and Proficiency Scores," Phi Delta Kappan, (December, 1966), 171-173.
- Ward, J., "On the Concept of Criterion-Referenced Measurement," British Journal of Educational Psychology, (November, 1970), 314-323.
- Weintraub, S., "What Research Says to the Reading Teacher," The Reading Teacher, XX (November, 1966), 155-165.

Reports

- Davis, F. B., "Criterion Referenced Measurement," AERA Conference Summaries, Princeton, New Jersey, ERIC Clearinghouse on Tests, Measurement, and Evaluation, 1972.
- Briggs, L. J., Stoker, H. W., and Scanlon, P., "Comparison of Performance on Objective-Referenced Vs. Content Referenced Achievement Tests," Mimeographed Document, Florida State University, 1971.
- Green, Donald, "Racial and Ethnic Bias in Achievement Tests and What to Do about It," CTB/McGraw-Hill, 1972, 1-9.
- Keller, C. M., "Criterion Referenced Measurement: A Bibliography," Princeton, New Jersey, ERIC Clearinghouse on Tests, Measurement, and Evaluation, 1972.

Publications of Learned Organizations

- Agnew, Donald C., "The Effect of Varied Amounts of Phonetic Training on Primary Reading," Duke University Research Studies in Education, No. 5, Durham, North Carolina, Duke University Press, 1939.

- Instructional Objectives Exchange, "Promoting Educational Excellence," (no date).
- Jenkins, Joseph, and Deno, Stanley, "Influence of Knowledge and Type of Objectives on Subject-Matter Learning," Journal of Educational Psychology, LII, 1971, 67-70.
- Johnson, Marjorie, and Kress, Roy, "Task Analysis for Criterion-Referenced Tests," The Reading Teacher, XXIV (January, 1971), 355-359.
- Lipe, Dewey and Jung, Steven, "Manipulating Incentives to Enhance School Learning," Review of Educational Research, XXI (October, 1971), 249-280.
- Mager, R. F., and Clark, C., "Explorations in Student-Controlled Instruction," Psychological Reports, XIII, 1963, 71-76.
- McNeil, J. D., "Programmed Instruction Versus Classroom Procedures in Teaching Boys to Read," American Educational Research Journal, I, 1964, 113-119.
- Millman, Jason, "Reporting Student Progress: A Case for Criterion-Referenced Marking System," Phi Delta Kappan, LII (December, 1970), 226-230.
- Otterman, Lois, "The Value of Teaching Prefixes and Word Roots," Journal of Educational Research, XLVIII (April, 1951), 611-616.
- Popham, James, and Husek, T. R., "Implications of Criterion-Referenced Measurement," Journal of Educational Measurement, VI (Spring, 1969), 1-9.
- Popham, W. James, "An Approaching Peril: Cloud-Referenced Tests," Phi Delta Kappan, (May, 1974), 614-615.
- _____, "Description and Use of IOX Objectives-Based Test Collection," 1973.
- _____, "Instructional Objectives Exchange, UCLA Graduate School of Education, (no date).
- _____, "The Instructional Objectives Exchange: New Support for Criterion-Referenced Instruction," Phi Delta Kappan, XXI (November, 1970), 174-175.
- Prescott, George, "Criterion-Referenced Test Interpretation in Reading," The Reading Teacher, XXIV (January, 1971), 347-354.

- Flanagan, J. C., "Unit, Scores, and Norms," Educational Measurement, Washington, D. C., American Council on Education, 1951.
- Flanagan, J. C., and Jung, S. M., "Evaluating a Comprehensive Educational System," Evaluative Research, Pittsburgh, Pennsylvania, American Institute for Research, 1970.
- Glaser, Robert, and Nitko, Anthony J., "Measurement in Learning and Instruction," Educational Measurement, edited by Robert Thorndike, Washington, D. C., American Council on Education, 1971.
- Klein, Stephen, and Kosecoff, Jacqueline, Issues and Procedures in the Development of Criterion Referenced Tests, Washington, D. C., National Institute of Education, 1973.
- Lessenger, Leon, "Accountability in Education," National Committee for Support of the Public Schools, Washington, D. C., Government Printing Press, February, 1970.
- Otto, Walter, "Assessing Needs and Growth," Assessment Problems in Reading, edited by Walter MacGinitie, Newark, Delaware, International Reading Association, 1973.
- Popham, James, "Defining Educational Objectives and Standards in Implementing the Stull Act," Stanford University Conference on the Stull Act, 1972.
- Westcott, Renée, The Use of Media in Higher Education, Raleigh, North Carolina, State Department of Higher Educational Facilities, May, 1970.

Manuals

- Karlsen, Bjorn, Madden, Richard, and Gardner, Eric, Manual for Administering and Interpreting Stanford Diagnostic Test, New York, Harcourt, Brace & World, 1966, 27-30.
- Instructional Objectives Exchange, Language Arts: Comprehension Skills Grades K-12, Los Angeles, California, 1972, 1-107.
- _____, Language Arts: Decoding Skills Grades K-12, Los Angeles, California, 1972, 1-142.
- _____, IOX Test Manual Reading Comprehension Skills Collection, Los Angeles, (no date).

Sullivan, Elizabeth, Clark, Willis, and Tiegs, Earnest,
California Short Form Test of Mental Maturity,
 Revised Edition, Monterey, California, California Test
 Bureau, 1963.

Speeches

Baker, R. L., "Measurement Considerations in Instruction
 Product Development," paper presented at Conference
 on Problems in Objectives Based Measurement, Center
 for the Study of Evaluation, University of California,
 1972.

Cox, Richard, "Evaluative Aspects of Criterion-Referenced
 Measures," paper presented at the 1970 meeting of the
 American Educational Research Association, Minneapolis,
 Minnesota, 1970.

Davis, Frederick, "Criterion-Referenced Tests," paper
 presented at the annual meeting of the American
 Educational Research Association, New York City,
 New York, 1971.

Fremer, John, "Applications of Criterion-Referencing to
 Schools," paper presented at Annual Conference
 of Educational Records Bureau, New York City, New
 York, 1973.

Hsu, Tse-Chi, "Empirical Data on Criterion-Referenced
 Tests," paper presented at the annual meeting of the
 American Educational Research Association, New York
 City, New York, February, 1971.

Knipe, Walter, and Krahmer, Edward, "An Application of
 Criterion Referenced Testing," paper presented at the
 annual meeting of the American Education Research
 Association, New Orleans, Louisiana, February, 1973

Nitko, Anthony, "Criterion-Referenced Testing in the
 Context of Instruction," paper presented at the
 Educational Records Bureau, National Council on
 Measurement in Education, New York, October, 1970.

Quie, Albert, "A New Approach to the Education of the
 Disadvantaged," paper presented to American Association
 of Colleges for Teacher Education, Chicago, Illinois,
 February, 1973.

Roberts, Tim, "Some Utilities of Criterion Testing," paper
 read at the Learning Disabilities Conference,
 Longmost, Colorado, November, 1972.

Roudabush, Glenn, "Item Selection for Criterion-Referenced Tests," paper presented at the American Educational Research Association Meetings, New Orleans, Louisiana, February, 1973.

Smith, Charles, "Criterion-Referenced Assessment," paper presented at International Symposium on Educational Testing, The Hague, The Netherlands, July, 1973.

Tiemann, P. W., "Student Use of Behaviorally-Stated Objectives to Augment Conventional and Programmed Revisions of Televised College Economics Lectures," paper read at the annual meeting of AERA, Chicago, 1968.

Filmstrips

Systematic Instructional Decision-Making, Los Angeles, Vimcet Association, 1967.

Unpublished Materials

Bryant, N., "The Effects of Performance Objectives on the Achievement Level of Selected Eighth Grade Science Pupils in Four Predominantly Black Inner-City Schools," unpublished doctoral dissertation, Indiana University, 1970.

Doty, C. R., "The Effect of Practice and Prior Knowledge of Educational Objectives on Performance, unpublished doctoral dissertation, Ohio State University, 1968.

Duchastel, Philippe, "An Investigation of the Organizing Function of Instructional Objectives in Relation to Experience with Objective-Based Testing," Dissertation Abstracts, XXXIV, Florida State University, 1973.

Engel, Roberta, "An Experimental Study of the Effect of Stated Behavioral Objectives on Achievement in a Unit of Instruction on Negative and Rational Base Systems of Numeration," unpublished master's thesis, University of Maryland, 1968.

Conlon, Betsey, "A Comparison of the Performance of Seventh-Grade Students with and without Prior Knowledge of the Objectives of an Individualized Science Program, unpublished doctoral dissertation, Florida State University, 1970.

- Griffin, J. E., "The Relationship Between Behavioral Objectives and Measurement Instruments Used to Evaluate Student Progress in an Urban Adult Basic Education Program," unpublished doctoral dissertation, Catholic University of America, 1971.
- Konski, Virginia, "An Investigation into Differences Between Boys and Girls in Selected Reading Readiness Areas and in Reading Achievement," unpublished doctoral dissertation, University of Missouri, 1951.
- Kurtz, Paul, "The Effect of Practice and Knowledge of Instructional Objectives on Learning from Prose," Dissertation Abstracts, XXXIII, 1972.
- Liveritte, Rudy, "A Philosophical Analysis of the Doctrine of Behavioral Objectives," Dissertation Abstracts, XXXIII, January, 1973.
- Piatt, R. G., "An investigation of the effect the train of Teachers in Defining, Writing, and Implementing Educational Behavioral Objectives has on Learner Outcomes for Students Enrolled in a Seventh-Grade Mathematics Program in the Public Schools, unpublished doctoral dissertation, Lehigh University, 1969.
- Schneiderwent, M. O., "The Effects of Using Behavioral Objectives in the Instruction of Harvard Project Physics," unpublished doctoral dissertation, University of Northern Colorado, 1970.