CALIFORNIA STATE UNIVERSITY, NORTHRIDGE

THE EFFECTS OF INSTRUCTION AND PRACTICE IN TEST-TAKING SKILLS ON THE STANDARDIZED READING TEST SCORES OF RANDOMLY SELECTED, CULTURALLY DIFFERENT, SECOND GRADE STUDENTS IN A TITLE 1 SCHOOL

A thesis submitted in partial satisfaction of the requirements for the degree of Master of Arts in Educational Psychology

by

Pearl Beverly Wenkert Kalechstein

May, 1974
The thesis of [name] is approved:

Committee Chairman

California State University, Northridge

May, 1974
ACKNOWLEDGMENTS

Sincere thanks are extended to the members of the master's committee: Dr. Robert Docter, Dr. Philip Hansen, and Dr. George Ferrell for their excellent advice, encouragement, and friendship.

Appreciation is offered to Mr. Philip Nash, Specialist in Research and Evaluation, Los Angeles Board of Education.

The author wishes to thank Mrs. Doyen, Mrs. Feldman, Ms. Mizota, and Mrs. Zakanych, second grade teachers; and Mr. Brajkovich, Principal; Mrs. Mehrens, Vice-Principal; and Mrs. Fletcher, Coordinator of the Vaughn Street Elementary School for their generous cooperation during this research.

Special appreciation is offered to Dr. Robert Docter for his friendship and advice throughout my graduate program.

Deep gratitude is expressed to my husband for his encouragement, advice, and patience.
TABLE OF CONTENTS

APPROVAL ................................................. ii
ACKNOWLEDGMENTS ........................................ iii
LIST OF TABLES ........................................ v
ABSTRACT ............................................... vi

Chapter

1. INTRODUCTION ........................................ 1
   Need for the Study .................................. 3
   Statement of the Problem ......................... 12
   Definition of Terms ................................ 13
   Limitations .......................................... 15

2. REVIEW OF THE LITERATURE ....................... 16
   Research on Practice and Coaching ............. 17
   Research on Test-wiseness ....................... 21
   Implications for the Present Investigation .... 32

3. PROCEDURES .......................................... 34
   Selection of Sample ................................ 34
   Collection of Data ................................ 35
   Materials .......................................... 37
   Treatment .......................................... 38

4. PRESENTATION AND ANALYSIS OF FINDINGS .... 41
   Statistical Analysis ................................ 41
   Discussion .......................................... 46

5. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS .. 48
   Conclusion .......................................... 50
   Educational Implications ......................... 50
   Recommendations ................................... 52

REFERENCES ........................................... 53

APPENDIX I: Experimental Material Administration
   Instructions ........................................ 60
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>STANDARD DEVIATIONS, CORRELATIONS OF PRETEST AND POSTTEST SCORES</td>
<td>42</td>
</tr>
<tr>
<td>2.</td>
<td>NON-ADJUSTED MEAN SCORES OF EXPERIMENTAL AND CONTROL GROUPS, PRETEST AND POSTTEST</td>
<td>43</td>
</tr>
<tr>
<td>3.</td>
<td>VARIANCE VALUES USED TO DETERMINE BETA</td>
<td>44</td>
</tr>
<tr>
<td>4.</td>
<td>MEAN PRETEST SCORES, MEAN POSTTEST SCORES, AND ADJUSTED MEAN POSTTEST SCORES</td>
<td>44</td>
</tr>
<tr>
<td>5.</td>
<td>ANALYSIS OF COVARIANCE</td>
<td>46</td>
</tr>
</tbody>
</table>
ABSTRACT

THE EFFECTS OF INSTRUCTION AND PRACTICE IN TEST-TAKING SKILLS ON THE STANDARDIZED READING TEST SCORES OF RANDOMLY SELECTED, CULTURALLY DIFFERENT, SECOND GRADE STUDENTS IN A TITLE I SCHOOL

by

Pearl Beverly Wenkert Kalechstein

Master of Arts in Educational Psychology

May, 1974

Evidence developed by noted test measurement experts (Vernon, 1954; Cronbach, 1970; Anastasi, 1969) indicated that practice and experience with tests tend to produce higher test scores, and tend to develop the components of test-wiseness. Research by Brim (1965), Goslin, et al. (1965), and Mercer (1973) indicated that culturally different children tend to lack the experiences necessary to develop test-wiseness.

The objective of this research was to investigate the effects of instruction and practice in test-taking techniques upon the standardized reading test scores of culturally different, second grade students from a Title I Public Elementary School. The hypothesis was that randomly selected, culturally different second grade students from Title I
schools who receive instruction and practice in test-taking techniques will achieve higher standardized reading test scores immediately following treatment than students who do not receive instruction and practice in test-taking techniques.

Forty-three culturally different, second grade students were randomly selected from four second grade classrooms in a Title I, public elementary school. The subjects were randomly assigned to either an experimental or a control condition.

The experimental group received approximately ten, thirty-minute periods of instruction and practice in test-taking skills. The materials used had a format similar to the Stanford Achievement Test, Primary I, Reading Form W. The control group received ten thirty-minute lessons with materials that did not have a format based on the Stanford Achievement Test. There was no deliberate instruction or practice in test-taking techniques with the control group.

Analysis of covariance showed that the experimental group performed significantly better than the control group when administered the Stanford Achievement Test. The hypothesis was supported.
Chapter 1

INTRODUCTION

Test-wiseness has been considered by measurement experts in education to be a potentially large source of test score variance (Thorndike, 1949; Vernon, 1962; Anastasi, 1969). Despite considerable interest in test-wiseness, there appeared to be no systematic study of either the importance of test-wiseness, nor the degree to which it could be taught or measured (Ebel, Millman, and Bishop, 1965).

In order to offer a theoretical framework which would serve as a basis for empirical investigation; Ebel, Millman and Bishop (1965) defined the term test-wiseness and developed an outline of test-wiseness principles. Their definition of test-wiseness was

... a subject's capacity to utilize the characteristics and formats of the test and/or the test-taking situation to receive a high score. Test-wiseness is logically independent of the examinee's knowledge of the subject matter for which the items are supposedly measures [p. 707].

Their outline of test-wiseness principles included time-using strategy, error-avoidance strategy, guessing strategy, deductive reasoning strategy, intent-consideration strategy, and cue-using strategy.
Both the outline and the definition have evidently served as an impetus for research on test-wiseness. Since 1965 several investigators have developed research whose purpose was to examine the effects of teaching test-wiseness to public-school students. Callenbach (1971) successfully taught test-taking skills to selected second-grade students. The studies on test-wiseness and their implications for the present investigation, will be examined in Chapter 2 of this study.

Measurement experts who have attempted to administer tests to children of low socioeconomic levels, have commented on their lack of test-wiseness (Smith, 1942; Eels, et al., 1951; Ortar, 1960). After testing Puerto Rican children in an area characterized by poverty, Anastasi and Cordova (1953) wrote:

Even more conspicuous as a reason for the inferior test performance of the present group is the children's attitude toward testing. A large number of factors, including lack of test sophistication, little or no motivation to excel in a competitive intellectual situation and a lack of interest in the relatively abstract and academic content of the test contributed to this attitude. The characteristic reaction to the testing was mild confusion, followed by amusement and indifference [Anastasi and Cordova, 1953, p. 15].

There has been no successful research that has attempted to teach test-taking skills to culturally different students from low socioeconomic neighborhoods. Nor is there a study which reports the effects of instruction and
practice in test-taking skills, upon the reading test scores of culturally different elementary school students at second grade level.

Need for the Study

From the earliest beginning of society men have measured the abilities of other men and recognized the existence of differences in abilities possessed by other individuals. All societies must train their members to perform the tasks necessary for the continuation and development of society. In addition, they must provide for the allocation of individuals to positions in the society [Goslin, 1967, p. 1].

One means or mechanism of allocation of individuals in literate societies seems to be written or oral achievement tests. Standardized tests of ability have been a major factor in this process in the United States (Goslin, 1967). The use of achievement tests can be traced back to antiquity (Coffman, 1969). The Chinese had a three thousand-year history of standardized, written tests whose purpose was to provide the government with capable civil servants through the use of competitive examinations (DuBois, 1966).

The earliest known measurement of academic achievement in the West was initiated in the year, 1200 A.D., when oral doctoral examinations were conducted at the University of Bologna in northern Italy (Linden, 1968). Written comprehensive essay examinations were introduced in Boston by Horace Mann in 1848; and modern achievement testing is generally considered to have begun with the publi-
cation of J. M. Rice's study of achievement in spelling in 1897. The development of standardized tests by E. L. Thorndike, Lewis Terman, and Truman Kelly in a wide variety of subject matter was to have a profound effect on American education (Coffman, 1969).

Approximately each of the 50 million or more school children in the United States take an average of three standardized tests each year (Goslin, 1967). Two of the main uses of the resulting test scores are to group children according to their abilities into homogeneous groups for instructional purposes, and to identify pupils with special educational weaknesses and strengths (Goslin, 1967).

Data on the extent of testing in schools indicated a greater frequency of test administration in the elementary school than in the secondary level. A very high proportion of elementary schools in the United States used standardized achievement and intelligence batteries in grades Kindergarten through six (Goslin, 1967). Goslin's contention that the schools depend heavily upon standardized tests to screen children were supported by surveys of testing programs in public elementary schools (Womer, 1959; Goslin, Epstein, and Hallock, 1965).

Ebel (1966) stated that standardized tests encouraged and rewarded individual efforts; and that, through testing, opportunities were extended on the basis of
aptitude and merit, rather than on ancestry, influence, and social class. This tended to place standardized testing in a democratic context; however, the research by Goslin, Epstein and Hallock (1965) disputed the theory that opportunities for testing experience were equally available to all members of society. In their research, supported by the Russell Sage Foundation, they discovered that the extent of testing in the elementary school is positively related to the average-income level of families of children in the school; and to the number of full or part-time counselors in the school. Moreover, it was clearly observed:

1. More testing is done in urban and suburban schools than in small town and rural areas;

2. Schools having a high percentage of college preparation students and a low percentage of negro students give more tests than those schools having a low percentage of college preparation students and a high percentage of negro students;

3. The extent of testing is positively related to per pupil expenditure, and negatively related to the percentage of male dropouts;

4. Testing appears to be a white, middle class phenomenon (Goslin, 1967).

This research is supported in the investigation by Brim (1965). He discovered that both race and social class had a considerable effect on the experiences and attitudes of American adults concerning standardized intelligence.
tests. In his survey, it was disclosed that the race of the respondents was correlated to test-taking experience. Only 39 percent of the white respondents reported never having taken a test; while 57 percent of the negro respondents gave this answer. More negros (48 percent) than whites (32 percent) reported that their children never took a test (p. < .01). In relation to attitudes, white respondents were more likely to feel that I.Q. tests measured what was learned; whereas, the negro tended to feel I.Q. tests measured inborn or innate intelligence.

The Brim (1965) research also disclosed that 90 percent of the members of the highest socioeconomic classes reported having taken tests. In contrast, the members of the lowest socioeconomic classes reported only 35 percent as having taken a test. Moreover, in the lowest socioeconomic class, those who had taken tests reported that their experience was with one or two tests, while upper-class respondents were likely to have had experience in three or more testing contexts.

The effects of practice and experience with tests on test scores has received considerable attention in the United States and in Britain. The results of those studies will be discussed fully in Chapter Two of this investigation. However, it is pertinent to mention here that Vernon (1954), Anastasi (1969), and Cronbach, (1970) all agree that the average test score gain from taking a single previous parallel test is about five I.Q. points. In
summarizing over twenty studies on practice effects, Vernon (1954) states that the effects of practice on the performance of testees' seemed to persist over relatively long periods of time. Vernon says,

... practice does not merely help in understanding instruction quickly. Learning of appropriate sets and methods and, possibly, the reduction of anxiety and carelessness, may be more important [Vernon, 1954, p. 273].

Anastasi (1969) defines the test-sophisticated individual as an individual with extensive prior experience, who enjoys an advantage over an inexperienced, unsophisticated examinee (p. 570). She continues, "Most middle class American school children are not only fairly test wise, but are generally motivated to succeed ... in test situations [Anastasi, 1969, p. 570]."

It has been noted that practice and experience with tests tends to raise the test scores about 5 I.Q. points; and that it seems to lead to test sophistication. It was also noted that a member of the white high socioeconomic class tended to enjoy an advantage over a member of negro low socioeconomic class (Goslin, Epstein, and Hallock, 1965; Brim, 1965; Anastasi, 1969).

Eels, Davis, Havighurst, Herrick, and Tyler (1951) investigated the possibility that children from widely varying cultural and socioeconomic backgrounds may demonstrate different problem-solving styles on tests. The investigators administered sample batteries of standard
I.Q. tests to 5,000 white pupils of varying socioeconomic status, who were between the ages of 9 to 14 years. Eels, et al. (1951) sought to determine the relationship between I.Q. scores and cultural and economic status levels. They discovered that there was a difference in problem solving styles between children of different cultural and economic status levels:

1. High status students scored 8-13 points higher; however, there was a large overlapping in scores between the different status levels;

2. Differences between items by status showed large status differences in items containing relatively academic vocabulary;

3. Analysis of the items showed that children of high socioeconomic status tended to guess at items they didn't know; while children of low economic status tended to omit more items;

4. Analysis of the incorrect responses showed that children of high economic status had incorrect responses that were close to the answer, and were less variant; while children of low economic status tended to guess randomly, their incorrect responses showing greater variability (Eels, et al., 1951).

One of the conclusions that may be drawn from this study is that children from low socioeconomic backgrounds lack skill in guessing strategy. Another observation was drawn by Havighurst:
To the average lower-class child . . . a test is just another place to be punished . . . this child learns to accept the inevitable and get it over with as quickly as possible. Observation of the performance of lower class children on speed tests leads one to suspect that such children often work very rapidly through a test making responses at more or less random. Apparently they are convinced in advance they cannot do well on the test, and they find that by getting through the test rapidly they can shorten the period of discomfort which it produced [p. 212].

Mercer (1973) questioned whether the distinction between intelligence tests and achievement tests is a valid distinction. She suggested that both are equally sensitive to sociocultural factors and measure the same characteristics. She administered the Stanford Reading Achievement Tests and the WISC I.Q. Test to children in grades one through grades three. She determined that reading achievement scores cannot be distinguished from I.Q. test scores either by sociocultural components or by correlation with measures of I.Q. She also found I.Q. test scores were correlated .50 with a set of 17 sociocultural variables when all ethnic groups were combined; and the scores correlated .61 for Mexican-Americans, .52 for Blacks and .31 for Caucasians.

Mercer (1973) studied the process whereby a person is labeled as a mental retardate by formal organization in the community of Riverside. She found four times more Mexican-Americans and three times more blacks were being labeled as mental retardates than would be expected from their percentage in the general population. The rate of labeling
for English-speaking Caucasians was only about half the number that would be expected from their percentage in the general population. These disproportions were most pronounced among governmental agencies using the statistical model of "normal" and relying heavily on I.Q. tests in making diagnoses, especially in the public school. Ethnic disproportions did not occur among persons referred from privately-funded organizations, or those relying on a medical-pathological model of normal (p. 56).

She found persons from minority groups were more likely than English-speaking Caucasians to have been diagnosed as mentally retarded when a low I.Q. test score was the only symptom. She suggests that as I.Q. scores are significantly influenced by the sociocultural background of the person being evaluated, then the sociocultural background should be taken into account when the scores are being interpreted. She feels strongly that I.Q. tests should be abandoned in education placement and planning.

Standardized testing has become, however, an integral and significant element in the educational establishment (Coffman, 1969; Goslin, 1962). The utilitarian functions of testing; e.g., grouping and counseling individuals on the basis of their scores have been described and listed by many measurement experts (Goodenough, 1949; Cronbach, 1954; Anastasi, 1969). Despite criticism of commonly-used methods of validating achievement test item (Buros, 1967),
the concern about the effects of misuses of tests (Goslin, 1963), and the warning (Coffman, 1969) that standardized tests readily take on an aura of scientific precision far beyond that which its creators would claim, there seems to be little possibility that the use of tests will be abandoned. Goslin (1962) states:

Never before has a society so conscientiously sought to evaluate scientifically the intellectual abilities of its members and to provide each individual with opportunities consonant with his aptitude [p. 171].

The rationale for this study is based on the preceding evidence developed by noted test measurement experts (Anastasi, 1969; Cronbach, 1970; Vernon, 1954) that practice and experience with tests tends to produce higher test scores, and tends to develop the components of test-wiseness. Based on the evidence presented by professional educators (Brim, 1965; Goslin, Epstein, and Hallock, 1965; Mercer, 1971 and 1973), children of ethnic minorities and children of low socioeconomic status lack the experiences necessary to acquire the components of test wiseness.

Callenbach (1971) has demonstrated that individuals can be trained in the classroom to acquire the components of test wiseness; and after training in test wiseness, they can score significantly higher on standardized achievement tests than those individuals who have not been trained in the components of test wiseness. Upon Callenbach's recommendation and with his permission, this study uses the format and materials he developed for his study in the present
research which will train culturally different children in the components of test-wiseness.

This study attempted to show that randomly selected second grade children from a Title 1 school who are considered culturally different, can benefit from instruction and practice in test-taking skills in a classroom environment; and therefore it will substantiate the claim by measurement experts and educators that test-wiseness is a large source of test score variance.

Statement of the Problem

The objective of this research was to investigate the effects of instruction and practice in test-taking skills upon the standardized reading test scores of randomly selected, culturally different second-grade students who are from a Title 1 school.

The design incorporated measures to present data leading to an evaluation of the immediate effects of presenting instruction and practice in test-taking techniques to randomly selected second grade students who were of low socioeconomic status and of ethnic minorities.

Hypothesis

Randomly selected, culturally different second grade students from Title 1 schools who receive instruction and practice in test-taking techniques will achieve higher standardized reading test scores immediately following
treatment than students who do not receive instruction and practice in test-taking techniques.

The .05 level of significance was chosen to determine non-chance variations in the treatment.

**Definitions of Terms**

**Coaching.** Instruction and preparation in academic subject matter upon which a test is based (Good, C., p. 109).

**Culturally different.**
Although many students considered disadvantaged are white . . . a majority of Mexican-American, Negro, and Puerto Rican youth can be described as disadvantaged. . . . The students have been termed culturally different because of their lack of "American" (or majority, advantaged group) cultural values and attitudes, which make it difficult for counselors to communicate with them by traditional means. The roots of this problem lie in the experiences of the culturally disadvantaged in "homes which do not transmit the cultural patterns necessary for the types of learning characteristics of the schools, and the larger society [Bloom, et al., 1965, p. 4] [Deighton, L., 1971, p. 481].

**Culturally disadvantaged.** See culturally different.

**Ethnic.** Relating to races or large groups of people classed according to common traits and customs.

**Low socioeconomic.** See culturally different, poor, and socioeconomic.

**Minority.** A part of a population differing from other groups in some characteristics.

**Poor.** Wanting in material goods; lacking in the comforts of life, characterized by inefficiency and unproductiveness (Gove, P., Webster's Dictionary, 1966).
Poverty area. Those areas where there is an incidence of poverty at least one and one quarter times the national average (Good, C. V., 1973).

Practice effect. Difference in performance due to repetitive exposure to a specific test or parallel forms of the same test (Good, C. V., p. 206)

Socioeconomic. Of or pertaining to social economics also have a character or aspect both social and economic (2nd. ed., Webster's Dictionary).

Test-taking techniques. "Skills commonly needed by students to utilize the characteristics and format of the standardized test, such as time-using strategy, following directions, response marking, guessing strategy (Callenbach, 1971, p. 7)."

Test-wiseness. A student's capacity to utilize the characteristics and format of the standardized test, independent of the student's knowledge of the subject matter (Ebel, et al., 1965).

Title 1 Schools. The Elementary and Secondary Education Act of 1965 (ESEA) provided funds to local agencies and school districts for special programs of assistance to educationally disadvantaged children. The programs included preschool and preventive services, remedial instruction, health care, parent education, and teacher training. Since 1965, the largest part of ESEA federal assistance funds has been expended under Title 1 for aid to disadvantaged children (Deighton, p. 532).
Title 1 schools serve areas of high concentration of poverty and low academic achievement.

**Limitations**

The second-grade students in this study were randomly selected from four second-grade classrooms of the Vaughn Street School, a Title 1 school in the Los Angeles City School District. Generalizations from this sample to other school systems, grade levels, and socioeconomic areas would be inappropriate.

Instruction and practice materials used in this study were based on the general format of the Stanford Reading Test, Primary 1 (1964). Instruction and practice in test-taking techniques devoted to the format of the Stanford Reading Test tend to limit the generalizations that may be drawn for other test-training programs using different tests (Callenbach, 1971).

Practice materials used in this experiment were developed by Callenbach (1971), and were used with his permission.
Chapter 2

REVIEW OF THE LITERATURE

There has been considerable interest and research over the years regarding the effects of practice and coaching upon the scores of standardized tests (Thorndike, 1922; Anastasi, 1969; Cronbach, 1970). Although this research was not directed toward test-wiseness, Ebel, Millman and Bishop (1965) state, "Test wiseness seems to be responsible, in part, for the effects of practice and coaching on test scores." In an article analyzing research on practice and coaching, P. E. Vernon (1954) notes this relationship by revealing that students who are experienced examinees gain approximately half as much from practice and coaching as those who are less sophisticated. Vernon states:

Such familiarization probably improves performance partly by reducing anxiety and carelessness, partly by inducing the set of working quickly, taking careful account of instructions and not wasting time on the difficult ones [p. 276].

This chapter will include a summary of the research devoted to the effects of practice and coaching, and a review of the studies designed to examine the effects of instruction in test wiseness. The conclusion of the chapter will contain a discussion of the implications of these studies for this research.
Research on Practice and Coaching

The effects of practice and coaching were investigated with considerable interest as standardized tests began to play a large role, especially in Great Britain, in the selection processes for positions in civil service, and particularly in determining the type of secondary education children would have. The many studies on practice and coaching have been reviewed and analyzed by Vernon, 1954; Anastasi, 1969; and Cronbach, 1970.

Vernon (1954) notes some serious technical deficiencies in several of the studies:

1. Failure to distinguish clearly between practice on a test where examinees learn only from their own experience; and coaching, where they are given aid in improving their performance;

2. Failure to differentiate between practice and coaching on identical tests, parallel forms or similar materials;

3. Failure to report the scores in some standardized unit of measurement [p. 271].

Most writers, however, agree that the data is complete enough to justify making some definite conclusions.

Practice

Many studies have investigated the effects of repetitive use of the same test over varying periods of
time ranging from a few days to a few years (Cattell, 1931; McIntosh, 1944; Derner, et al., 1950; Levine and Angoff, 1954; Watts, 1958). All studies show a significant gain on the mean score after the subjects have been retested once. There were continued gains on mean scores after subsequent repetition of testing, that were not limited to the gains made on the initial repetition. For example, McIntosh (1944), used a standardized verbal I.Q. test for eleven year old children and obtained a mean score gain of 7.2 I.Q. points on the second trial; a gain of 10.7 I.Q. points on the fifth trial. Derner, et al. (1950) found I.Q. point increases on the Wechsler-Bellevue of 8.5 I.Q. points, 7.6 I.Q. points, and 6.2 I.Q. points; when subjects were retested after one week, four weeks and six months, respectively. Cronbach (1970) summarized the data and stated that practiced-uncoached groups gained about 6 I.Q. points when tested and retested with the same test instrument without any special explanation.

Gains in scores tend to be smaller when retesting with parallel forms of the same test (Thorndike, 1922; Adkins, 1937; Watts, et al., 1952; Munyan, 1947; Peel, 1951; Peel, 1952; Anastasi and Cordova, 1953; Quereshi, 1968). Watts, et al. (1952), in an experiment involving children from London, found the mean I.Q. score on the eighth parallel test was only 6 points higher than on the first parallel test.
Anastasi and Cordova (1953) administered the Cattell Culture Free Intelligence Test, Forms 2A and 2B to 176 Puerto Rican children from Spanish Harlem in New York City. It was a bilingual experiment in which one half the group received the test instruction in English during the first testing session (Form A); and they received the instruction in Spanish during the second session, (Form B). The order of the languages was reversed for the other half of the group. The authors used analysis of variance to evaluate the effects of language, session, order, and sex on test performance. They discovered that the most significant factor was the order that the tests were administered. There was a significant practice effect at the .01 level. The Anastasi and Cordova (1953) research suggested that children of low socioeconomic background, who are culturally different do tend to benefit from practice with parallel tests.

Vernon (1954) reviewed the results of about twenty studies and concluded that the average gain from a single previous parallel test produced about 5 I.Q. points. Further practice may produce some small gains, about 10 I.Q. points; but after five practice tests, there was no further improvement and the I.Q. measurements fluctuated or tended to decline.
Coaching

Research on the effects of coaching (Greene, 1928; Wiseman, 1954; Ortar, 1960) agree with the Anastasi (1969) summation of the literature, that large gains in test scores result from coaching on identical tests and smaller gains are obtained on parallel tests or similar material. Green (1928), as cited by Vernon (1954) designed research with three groups of children. One group was coached on the Stanford-Binet; one group was coached on similar tests; and one group was designated as a control group and was not coached. All the groups were tested with the Stanford-Binet at three weeks, again at three months, one year, and three years. The average rise of her control groups was 5.0 I.Q. points at three weeks; 2.6 I.Q. points at three months; 3.3 I.Q. points at one year; and 0.6 I.Q. points at three years. The average rise in score of the group coached on similar material was 7.9 I.Q. points at three weeks; 7.6 I.Q. points at three months; 5.6 I.Q. at one year; and 1.5 I.Q. points at three years. The group coached on the Stanford-Binet gained 29.1 I.Q. points in three weeks over the first test; 17.5 I.Q. points in three weeks; 12.6 I.Q. points in one year; and 4.3 I.Q. points in three years.

Ortar (1960) developed a three-part test, each part of the test being parallel. The research was designed to examine the value of practice and coaching on children who were culturally different. She administered and scored
the first part of the test. She used the second part of the test to coach the children and give them intensive practice with tests. The third part of the test was administered and scored; and it was the third part of the test that was used as an index of the subject's ability to profit from academic instruction. The mean scores of the third part of the test were significantly higher than the first test. Ortar concluded that the higher scores resulted in part from:

... better understanding of the principles underlying the test. Reason based on general principles is often alien to some of the subjects because it had not been demanded of them at all (cultural pattern) or because it was not absorbed when taught (social class influence)... [Ortar, 1960, p. 183].

Cronbach (1970) summarized the literature on coaching and concluded that coached groups gained 5-6 points after being told about tests and having had numerous items explained by teachers.

Research on Test-wiseness

Callenbach (1971) cited an unpublished Doctoral dissertation by Kreit (1967), that professed to examine the effects of teaching test-wiseness to naive third grade students. Kreit administered several different intelligence tests to four experimental classes, and to four control classes. The experimental group was administered four I.Q. tests; the control population was administered two tests. Kreit (1967) reported that the experimental group
increased their scores significantly on the fourth test (p < .05). However, there was no significant difference in the gain made by the control and the experimental groups. Kreit concluded that the mean difference in scores between the experimental and control groups was about 3.4 I.Q. points but many have been as high as 6.1 I.Q. points.

The results of the Kreit study support Cronbach's (1970) summary of data on the practice effect: the average gain from experiencing four to eight tests is approximately 6 points. Kreit's study does suggest that test-naive third grade students will benefit from practice experience with standardized tests.

Moore, Schutz and Baker (1966) organized research whose main purpose was to show that test-taking strategy can be taught with self-instructional materials. They developed a programmed text, designed to teach eighth grade students strategies for answering speed and power tests scored with and without a correction for guessing. Sample test directions, and test items were included in the text to give the students experience and guided practice.

The populations were made up of four eighth grade classrooms. Two classrooms were chosen at random for experimental treatments, and the remaining two classrooms were designated as the control population. The "Paragraph Meaning Test" of the Stanford Achievement Test administered four weeks prior to the treatment was used as a control in a covariance analysis.
The students were required to take four tests under four different conditions: 1) The first condition was a power test with no penalty for guessing; 2) the second condition was a speed test with no penalty for guessing; 3) the third condition was a power test with a penalty for guessing; 4) the fourth condition was a speed test with a penalty for guessing. The test consisted of typical questions included in college-level I.Q. tests.

The investigators found that the experimental group under the conditions of tests 1, 3, and 4 marked significantly more questions than did the control groups under the same conditions (p < .01). They found no difference in the number of questions answered under the condition of test 2. The control groups marked a significantly larger number of items on test 3 which was a power test with a penalty for guessing.

Moore, Schutz and Baker (1966) did find that self-instructional materials can help develop test-wiseness; however, their use of the "Paragraph Meaning Test" scores as a covariate in analyzing the results of the criterion test is not explained. Second, the performance score should have included the number and accuracy of responses, rather than just the number of responses.

Wahlstrom and Boersma (1968) hypothesized that test-taking skills could be taught by classroom instruction and that such instruction would enable a test-wise examinee to
obtain a higher score than an equally knowledgeable examinee who lacks test-wiseness.

The subjects consisted of one hundred and seventeen students in four ninth-grade classes from two schools. One-sixth of the subjects in each school were randomly assigned to the control groups (C-1 and C-11), the experimental groups (X-1 and X-11), and the placebo groups (P-1 and P-11).

The authors developed an "Outline of Test-Wiseness Principles," and "Elaboration of Selected Principles," based on the analysis of test-wiseness published by Ebel, Millman and Bishop (1965), which were designed to meet the language level of their ninth-grade experimental groups. Criterion test items were written by test specialists particularly for ninth-grade social studies students. The item pool was composed of 108 items from previous tests. Fifty-four items were randomly selected to construct test 1-G from the available 108 items. The remaining test items were incorporated into test 2-G. Two additional tests were developed which contained modified items from tests 1-C and 2-G and were called test 1-P and test 2-P. Tests 1-G and 2-G contained "good" items and Test 1-P and test 2-P contained items considered "poor." That is, when developing tests 1 and 2-P the test specialists incorporated "faults" common to multiple-choice items, into the test (Wahlstrom and Boersma, p. 415).
The "Verbal Reasoning Test" of the Differential Aptitude Tests was administered to all the subjects in their homerooms at the start of the school day. Raw scores from the test were used to ensure that the six groups were not significantly different from one another in their general ability.

Pretests 1-G and 1-P were administered to the intact classes immediately after the Verbal Reasoning Test. During the following two hours, classes were held as usual, with the exception that no instruction in social studies was given to any of the groups. Approximately three hours after taking the pretest, the groups were separated into three different rooms, where they received the following treatment:

1. Control groups C-1 and C-11 watched television for the entire 100 minutes of the treatment time.

2. Experimental groups X-1 and X-2 were presented with a lesson on test-wiseness by an experimenter E-1, for twenty-five minutes. During the second 25-minute period, another experimenter, E-2, supervised study of the "Outline of Test-Wiseness Principles." During the third 25-minute period experimenter E-1 discussed the principles of test-wiseness. For the last 25-minute period, a paper "Elaboration of Selected Principles" was distributed by experimenter E-2 and studied by the students.

3. Placebo groups P-1 and P-2 were supervised in
reading and discussing pamphlets on general occupational information, for the first 25 minutes. They received a lecture on occupational information for the following 25-minute period by experimenter E-1; during the third 25-minute period they were supervised by experimenter E-2; and for the remaining period of time, experimenter E-1 discussed occupational roles with the students.

After the groups had received the treatments described above, the students were returned to their homerooms, where tests 2-G and 2-P were given. All subjects were allowed 35 minutes to complete the criterion tests.

Correlated t-tests for all groups were used to test whether pretest and posttest scores differed significantly. There were no significant differences between the pretest and posttest scores of the control and placebo groups. The posttest mean score of the X-1 group was significantly higher than its pretest (p < .01). No significant difference was observed between 1-P and 2-G mean scores for the X-11 group, which the authors suggest is attributable to the "good" items in 2-G minimizing the effects test-wiseness.

The investigators conclude that the results of their study indicate that test-wiseness can be taught in the classroom and does lead to improved achievement on test performance. However, several factors tend to minimize the utility of this research. First, the investigators chose
to evaluate the effect of their instruction in test-wiseness principles with teacher-made social studies tests, rather than standardized tests. Second, the additional variable they chose to include by using "good" and "poor" quality items tended to place the research in an artificial setting, and clouded the results of their study.

Tinney (1968) developed research to compare the effect of training in test-taking skills on the reading test scores of fifth-grade students of high and low socioeconomic levels. He selected a sample of 105 students from two classrooms in each of two public schools. The schools were selected on the basis of census tract information to represent high and low socioeconomic levels. The pupils selected in each school were randomly assigned by sex to experimental and control groups.

The experimental groups received five consecutive daily lessons of 30-minutes duration. The sessions were devoted to the skills involved in efficient use of time, importance and proper use of guessing, an organized system of eliminating answers, proper use of a separate answer sheet, a basic approach to answering test questions from stories, and the typical format and purpose of comprehension test questions related to finding information, seeing relationships, making interpretations and appreciation of the story.

Each lesson was planned to incorporate review of
the preceding lesson, teaching of a new lesson skill, worksheet practice and a final lesson review. The worksheets were developed by using stories and questions from Form B of the New Developmental Reading Test and the worksheets incorporated the format of that test.

The control groups received five drawing lessons.

Tinney measured the effects of the training on reading achievement test scores of the experimental and control groups by administering, as a pretest and posttest, Form A of the New Development Reading Test, Intermediate Level. Differences in final reading scores were analyzed by analysis of covariance, using the pretest reading score as a covariate.

The resulting data showed that the high socioeconomic control group made approximately the same gain as the high socioeconomic experimental group. The low socioeconomic groups showed negative changes for both the experimental and control groups. Tinney found that the low socioeconomic groups were reading approximately two and one half years below the high socioeconomic groups on the pretest and approximately three years below on the posttest. However, the difference between the experimental and control groups were not significant at either the .01 or the .05 levels of significance.

Callenbach (1971) developed a study whose purpose was to investigate the effects of instruction and practice in test wisdom on the standardized reading test scores of second-grade students. The study was designed to measure the reading test scores of the subjects before the
treatment, immediately after treatment, and four months after treatment.

The subjects were selected from two second-grade classrooms of a public elementary school. All the available students were given the Stanford Reading Test, Primary 1, Form X (1964). On the basis of their total raw score, the subjects were ranked and matched into two groups. The two groups were randomly assigned to the experimental condition and the control condition. Callenbach presented pretest data on the means and standard deviation of the experimental group (117.75; 22.58) and on the control group (117.37; 23.38); in order to demonstrate that the two groups had been equated. He also presented a table showing the paternal occupations of the students in each group to indicate that their backgrounds were also matched. Their backgrounds were middle and upper socio-economic levels.

Callenbach selected four test-taking skills as the basis of instruction for his experimental group:

1. Comprehension of oral direction;
2. Correct marking responses;
3. Time-using strategy;
4. Guessing skills.

The materials developed for the experimental group were similar in format to the Stanford Reading Test, Primary 1 (1964). The practice sheets consisted of eight
lessons using geometric figures and arithmetic problems.

The control group also had eight lessons of the same time duration as the experimental group with similar geometric and arithmetic problems; but these lessons were presented without instruction in test-taking skills and the lessons did not have the format that was similar to the format of the Stanford Reading Test, Primary I.

The experimental and control groups met twice a week, for four weeks. Each session took about one half hour. Makeup lessons were given to all students who were absent. To control for teacher effect, the investigator presented the materials to both groups.

The immediate effect of the treatment was measured with the Stanford Reading Test, Primary I, Form W. The delayed effect of the treatment on the standardized reading test scores of both the experimental and control groups was measured with the Stanford Reading Test, Primary II, Form X (1964).

The statistical analysis was based on the test for significance of difference between means of equivalent groups. The immediate effects of treatment were measured through an analysis of improvement in test scores between the experimental and the control groups. The delayed effects were also measured by analyzing the difference between the scores of the experimental and control groups.

The data from this study supported Callenbach's
hypothesis that subjects who received instruction and practice in test-taking techniques achieved higher scores than subjects who did not receive instruction and practice in test-taking techniques. The experimental group scored significantly higher (p < .025) than the control group in the immediate posttest. The experimental group scored significantly higher (p < .01) than the control group in the delayed posttest. Callenbach included all appropriate tables in his study.

The data added to the literature on gains on standardized test scores due to retesting with parallel forms of the same test (Vernon, 1954; Cronbach, 1970). The analysis of the scores showed that each group, both control and experimental, scored significantly higher (p < .0005). This gain was probably due to the practice effect derived from the pretest experience.

The significant difference between the scores of the control and experimental groups on the delayed posttest also indicated that the gain in score was maintained over a period of four months, and that the experimental group widened its advantage over the control group.

By designing abstract mathematical and geometric instructional materials independent and unrelated in content to the academic information measured by the Stanford Reading Test; Callenbach's study leads to the possibility that test-wiseness is a cognitive skill which can be taught in the classroom (Ebel, et al., 1958, p. 1511).
Implications for the Present Investigation

The review of the literature on practice and coaching effects on test scores, and analysis of the research pertaining to test-wiseness suggested the following implications:

1. First, it is likely that culturally different students from low socioeconomic backgrounds will benefit from practice and instruction in test-taking skills. The Anastasi and Cordova study (1953), and the Ortar study (1960), demonstrated that children from culturally different, low socioeconomic levels benefit from practice and coaching on tests.

2. An effort must be made in this research to control the practice effects of the pretest, and limit its influence on the posttest. The large gain made by the control group in Tinney's study (1968), and the significant gains in scores made by the control group in Callenbach's study (1971) indicated that the extent of the practice effect is considerable. Vernon (1954) stated that the effects of any type of practice or coaching are limited.

3. The Callenbach study supplied evidence that the practice sessions and materials must simulate real testing conditions. Vernon (1954) suggested that there was little transfer if practice and coaching activities and materials did not closely parallel real testing conditions.

4. The Callenbach study indicated that second-
grade students can benefit from practice and instruction in test-taking skills.

5. The Tinney research (1968) indicated that naive examinees should be instructed on a limited number of test-taking skills, rather than incorporate all the test-taking skills into a treatment (Callenbach, 1971).
Chapter 3

PROCEDURES

This research was designed to investigate the effect of instruction and practice in test-taking skills upon the standardized reading test scores of randomly selected, culturally different second-grade students from a Title 1 school.

Selection of Sample

To provide for statistical equivalence between groups, and to control for classroom teacher variable, all second-grade students from Vaughn Street Elementary School were listed, and a sample was randomly selected for inclusion in the experiment, using the table of random numbers (Ary, et al., 1972, p. 366).

The subjects were alternately assigned to one of two groups until there was a total of fifty subjects in the experiment, or twenty five subjects in each group. The two groups were randomly assigned to either the experimental or control condition by the flip of a coin.

Based upon the advice of the principal of the school, each group was divided into two subgroups. There were then a total of four subgroups; two subgroups in
the experimental condition, and two subgroups in the control condition. The suggestion to subdivide the experimental and the control groups for purposes of instruction into a total of four subgroups, was based on the prevailing ratio of adult to child at this school, which was approximately one adult to ten students.

Collection of Data

Measures Used

The pretest used was the Cooperative Primary Reading Test, Form 12A. It was administered to the subjects by the Los Angeles City School System in May, 1973, approximately ten months prior to the experimental procedure. The pretest data were used for two reasons:

1. The pretest was administered approximately ten months prior to the posttest to control the practice effects of a pretest, and to limit its influence on the posttest. The large gain made by the control group in Tinney's study (1968), and the significant gains in score made by the control group in Callenbach's study (1971), indicated that the extent of the practice effect is considerable.

2. The present research was designed to control the sensitization effect of a pretest. A pretest just prior to the treatment could cause the experimental groups to respond to the treatment with greater awareness and sensitivity than would otherwise occur if they had not had a pretest.
The control groups were labeled A and AA; the experimental groups were labeled B and BB. Control group A was composed of thirteen subjects; control group AA was composed of twelve subjects. Experimental group B was composed of thirteen subjects; experimental group BB was composed of twelve children. Control groups A and AA received identical control treatment; and experimental groups B and BB received identical experimental materials and treatment.

The design selected for this research was designated and labeled "Design 5. Randomized Groups, Pretest-Posttest Design," by Ary, et al. (1972, p. 244). This design incorporated randomly selected subjects assigned to an experimental and control condition. Treatment specific to the research was introduced to the experimental subjects; while another irrelevant treatment was designated for the control group.

A pretest was administered to the experimental and control groups ten months prior to treatment. Following treatment, a posttest was administered to both groups. The average difference between the pretest and the posttest scores was found for each group. The average difference scores were compared in order to ascertain whether the experimental treatment produced a greater change than the control situation (Ary, et al., p. 245).

The analysis of covariance was used, since it is, . a method for analyzing differences between experimental groups on the dependent variable after
taking into account any initial differences between the groups on pretest measures or any other relevant independent variables [Ary, et al., p. 244].

The beta Formula was used to adjust posttest scores of the experimental and control groups to allow for initial differences between the experimental and control groups on pretest scores.

\[
\text{beta within } = \frac{E x y w}{E x^2 w}
\]

\[
\text{Adjusted Posttest } \quad \overline{Y_e} = \overline{Y_e} - bw (\overline{X_e} - \overline{X_t})
\]

\[
\text{Experimental Formula}
\]

\[
\text{Adjusted Posttest } \quad \overline{Y_c} = \overline{Y_c} - bw (\overline{X_c} - \overline{X_t})
\]

\[
\text{Control Formula}
\]

\textbf{Materials}.

\textbf{Experimental Group Materials}

The materials used were constructed by Dr. Carl Callenbach (1971), in partial fulfillment of the requirements for his degree of Doctor of Education. The materials were used with Dr. Callenbach's permission. The materials were based on the format of the Stanford Achievement Test, Primary 1 Reading Test (1965). The experimental materials consisted of geometric and arithmetic problems, rather than
reading problems; in order to keep the instruction and practice materials independent of the subject matter measured by the Stanford Achievement Test, Primary 1, Reading Test (Callenbach, 1971).

The following test-taking techniques were selected from Millman, et al. (1965) by Callenbach (1971), to be included in the lesson plans:

1. Following orally-administered directions;
2. Marking responses;
3. Time-using strategy; and
4. Guessing strategy.

A sample of the instructions was included in the Appendix.

Control Group Materials

To control for possible Hawthorne effect, the control lessons were of the same duration as the experimental group lessons, and were developed from the same arithmetic and geometric materials as the experimental group lessons. However, they did not have the same format, and the materials were presented without instruction in test-taking techniques.

Treatment

The experimental and control treatments were conducted for five weeks, with two lessons of approximately thirty minutes for each group, each week. Absenteeism
occurred in all groups; however, no makeup lessons were offered. The investigator presented the materials to all groups in order to control for teacher effect. The groups met in the Vaughn Street Elementary School Library. The first two weeks of treatment took place on February 21, February 26, February 28, and March 5, 1974. The schedule for the first two weeks was:

- Control Group A 8:45 - 9:15
- Experimental Group B 9:20 - 9:50
- Experimental Group BB 10:15 - 10:45
- Control Group AA 10:50 - 11:20

The schedule was rotated the following three weeks: March 7, March 12, March 14, March 19, March 21, and March 26, 1974. The schedule for the last three weeks was:

- Experimental Group BB 8:45 - 9:15
- Control Group AA 9:20 - 9:50
- Control Group A 10:15 - 10:45
- Experimental B 10:50 - 11:20

Posttest

The Word Reading and Paragraph Meaning sections of The Stanford Achievement Test, Primary 1, Reading Test, Form W, was administered by the investigator to the experimental and the control groups. The test is appropriate for administration to both first and selected second-grade students (Kelly, et al., 1966, p. 1). The "Technical Supplement" (1966) presents detailed information regarding the
standardization, reliability and validity of this test.

The Word Reading and Paragraph Meaning sections of the Stanford Achievement Test, Primary 1, Form W, was administered by the investigator in the Vaughn Street Library to the experimental and control groups during the week immediately following the experiment. The testing schedule was as follows for the Word Reading Section on March 26, 1974:

<table>
<thead>
<tr>
<th>Group</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental Group BB</td>
<td>8:45 - 9:15</td>
</tr>
<tr>
<td>Control Group AA</td>
<td>9:20 - 9:50</td>
</tr>
<tr>
<td>Control Group A</td>
<td>10:15 - 10:45</td>
</tr>
<tr>
<td>Experimental Group B</td>
<td>10:50 - 11:20</td>
</tr>
</tbody>
</table>

The schedule was as follows for the Paragraph Meaning Section on March 28, 1974:

<table>
<thead>
<tr>
<th>Group</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental Group BB</td>
<td>8:45 - 9:15</td>
</tr>
<tr>
<td>Control Group AA</td>
<td>9:20 - 9:50</td>
</tr>
<tr>
<td>Control Group A</td>
<td>10:15 - 10:45</td>
</tr>
<tr>
<td>Experimental B</td>
<td>10:50 - 11:20</td>
</tr>
</tbody>
</table>

Since five students moved, and since there was no pretest data on two students, seven students were eliminated from the sample.

The data which were collected and analyzed are presented in Chapter 4, with a discussion of findings.
Chapter 4

PRESENTATION AND ANALYSIS OF FINDINGS

Statistical Analysis

The analysis of covariance was used in this study since it "permits a compensation to be made in the analysis of the data for the lack of equivalency in groups initially. Precision is attained without having the problems of matching [Ary, et al., 1972, p. 275]."

This statistic was used to determine if the post-test results on the Stanford Achievement Test, Primary 1, Reading Test, Form W, differed significantly between the experimental and control group; with the achievement on the pretest, the Cooperative Primary Test, Reading Form 12A, being used as a covariant, or controlling variable.

Table 1 presents the standard deviations of the experimental and control groups on both the pretests and posttests, and the correlations between the two tests. The differences in the standard deviations between the experimental and control groups were insignificant (p > .05) which would indicate that the two groups had the necessary homogeneity of variance. The larger standard deviations for the experimental group on the pretest and posttest would suggest greater variability in scores, but the small

41
sample size of the groups contributed to the apparent differences in the standard deviations.

Table 1

STANDARD DEVIATIONS, CORRELATIONS OF PRETEST AND POSTTEST SCORES

<table>
<thead>
<tr>
<th>Group</th>
<th>Experimental</th>
<th>Control</th>
<th>Significance of Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD Pretest</td>
<td>9.49</td>
<td>6.79</td>
<td>None*</td>
</tr>
<tr>
<td>SD Posttest</td>
<td>16.81</td>
<td>14.09</td>
<td>None*</td>
</tr>
</tbody>
</table>

Correlation Between Pre- and Posttest** .715 .678 None***

* The F test of differences between Standard Deviations to determine homogeneity of variance. \( p > .05 \) (Guilford, P., 1950, p. 233).

** Pearson Product Moment \( r \)

*** The z test of differences between \( r \)’s \( p > .05 \) (Guilford, 1950, p. 224)

The correlations of .715 and .678 were not significantly different \( (p > .05) \); however, both correlations were of sufficient magnitude to indicate that the pretest scores on the Cooperative Primary Test, Reading Form 12 A, made a satisfactory covariant with the Stanford Achievement Test, Primary 1, Form W Reading Test with which to measure the
initial level of pupil achievement before any experimental conditions were imposed.

Table 2

NON-ADJUSTED MEAN SCORES OF EXPERIMENTAL AND CONTROL GROUPS, PRETEST AND POSTTEST

<table>
<thead>
<tr>
<th>Groups</th>
<th>N</th>
<th>ΣX</th>
<th>ΣX²</th>
<th>X</th>
<th>ΣY</th>
<th>ΣY²</th>
<th>Y</th>
<th>ΣXY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>23</td>
<td>333</td>
<td>6801</td>
<td>14.48</td>
<td>1075</td>
<td>56461</td>
<td>46.74</td>
<td>18074</td>
</tr>
<tr>
<td>Control</td>
<td>20</td>
<td>229</td>
<td>3499</td>
<td>11.45</td>
<td>603</td>
<td>21953</td>
<td>30.15</td>
<td>8137</td>
</tr>
<tr>
<td>Total</td>
<td>43</td>
<td>562</td>
<td>10300</td>
<td>13.07</td>
<td>1678</td>
<td>78414</td>
<td>39.02</td>
<td>26211</td>
</tr>
</tbody>
</table>

The experimental group had a mean score of 14.48, and the control group had a mean score of 11.45 on the pretest. The experimental group had a mean score of 46.74 and the control group had a mean score of 30.15 on the posttest. The scores represent mean scores that have not been adjusted by the beta Formula.
Table 3

VARIANCE VALUES USED TO
DETERMINE BETA

<table>
<thead>
<tr>
<th>Group</th>
<th>$\Sigma x^2$</th>
<th>$\Sigma y^2$</th>
<th>$\Sigma xy$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>1979.74</td>
<td>6216.43</td>
<td>2509.87</td>
</tr>
<tr>
<td>Control</td>
<td>876.95</td>
<td>3772.55</td>
<td>1232.65</td>
</tr>
<tr>
<td>Within</td>
<td>2856.69</td>
<td>9988.98</td>
<td>3742.52</td>
</tr>
<tr>
<td>Between</td>
<td>98.10</td>
<td>2944.00</td>
<td>531.41</td>
</tr>
<tr>
<td>Total</td>
<td>2954.79</td>
<td>12932.98</td>
<td>4279.93</td>
</tr>
</tbody>
</table>

Table 3 indicates the values necessary to determine a correction value using the beta Formula:

$$\text{beta within} = bw = \frac{E x y w}{E x^2 w}$$

Table 4

MEAN PRETEST SCORES, MEAN POSTTEST SCORES
AND ADJUSTED MEAN POSTTEST SCORES

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean Pretest</th>
<th>Mean Posttest</th>
<th>Mean Adjusted Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>14.48</td>
<td>46.74</td>
<td>44.89</td>
</tr>
<tr>
<td>Control</td>
<td>11.45</td>
<td>30.15</td>
<td>32.27</td>
</tr>
</tbody>
</table>

In Table 4 the experimental group had a posttest mean of 46.75; and the control group had a posttest mean of
30.15. When the means were corrected for the pretest scores using the beta Formula, the adjusted mean posttest score of the experimental group was 44.89; and the adjusted mean posttest score of the control group was 32.27. The adjusted means of both groups indicate what they would have been if both groups had equal means and variances on the pretest scores. The adjusted mean score of the experimental group was 12.62 points higher than the control group which gives strong support to the hypothesis that culturally different students can be taught test-taking techniques.

\[
\begin{align*}
\text{Adjusted Posttest} & \quad \bar{Y}_e = \bar{Y}_e - bw (\bar{X}_e - \bar{X}_t) = \\
\text{Experimental} & \quad = 46.74 - 1.31 (14.48 - 13.07) \\
& \quad = 44.89 \\
\text{Adjusted Posttest} & \quad \bar{Y}_c = \bar{Y}_c - bw (\bar{X}_c - \bar{X}_t) \\
\text{Control} & \quad = 30.15 - 1.31 (11.45 - 13.07) \\
& \quad = 32.27
\end{align*}
\]
Table 5

ANALYSIS OF COVARIANCE

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>1647.68</td>
<td>1</td>
<td>1647.68</td>
<td>12.96*</td>
</tr>
<tr>
<td>Within Groups</td>
<td>5085.94</td>
<td>40</td>
<td>127.15</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6733.62</td>
<td>41</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .01.

The F ratio of 12.96 was significant at the .01 level, one degree of freedom, between groups, and 40 degrees of freedom, within groups.

The significant F ratio indicated that mean scores of the experimental group on posttest Y was significantly higher than mean scores of the control group on posttest Y, with both groups controlled by their initial achievement on pretest X. (Posttest Y being the Stanford Achievement Test, Primary 1, Reading, Form W; Pretest X, being the Cooperative Primary, Reading Test Form 12 A.)

Discussion

The hypothesis of this study suggested that randomly selected, culturally different, second grade students from a Title 1 school who received instruction and practice
in test-taking techniques, would receive higher standardized reading test scores than second grade students who did not receive instruction and practice in test-taking techniques. The data supports the hypothesis. The scores of the experimental group were significantly higher ($p < .01$) than the control group scores in the posttest.

The conclusions drawn from this research tend to support the conclusions and data drawn by Callenbach (1971). His study indicated that second grade students benefit from practice and instruction in test-taking skills. The present study indicated that culturally different second grade students from a Title I school do benefit from instruction in test-taking skills.
Chapter 5

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The purpose of this research was to analyze the effects of instruction and practice in test-taking techniques on the standardized reading test scores of randomly selected, culturally-different, second-grade students, in a Title I School.

All the students in the four second-grade classrooms at the Vaughn Street Elementary School, Los Angeles City School District, were listed; and a sample was randomly selected for inclusion in the experiment, using the table of random numbers (Ary, et al., 1972, p. 366).

The subjects were alternately assigned to one of two groups until there was a total of fifty subjects in the experiment, or twenty-five subjects in each group. The two groups were randomly assigned to either the experimental or control condition. The groups were further divided into subgroups; the control groups being labeled A and AA, the experimental groups being labeled B and BB, in order to maintain the pupil-to-adult ratio which prevailed at that school. The control groups A and AA were treated as one group statistically and the experimental groups were treated as one group statistically.
The experimental and control groups were subjected to ten, 30-minute lessons, which took place in the school library, for five weeks. The experimental lessons were composed of instruction and practice in the following test-taking techniques: orally-administered direction-marking responses, time-using strategy, and guessing strategy. The control lessons were composed of similar mathematical and geometric material; but they did not have a format based on the Stanford Reading Tests, Primary 1, as did the experimental material, nor were instructions on test-wiseness skills presented.

Randomly selected, culturally different second grade students from a Title 1 School who receive instruction and practice in test-taking skills will achieve higher standardized reading test scores than randomly selected second grade students from a Title 1 School who do not receive instruction and practice in test-taking skills.

Analysis of covariance was used to determine if the posttest results on the Stanford Achievement Test, Primary 1, Reading Test, Form W, differed significantly between the experimental and control group; with the achievement on the initial pretest, the Cooperative Primary Test, Reading Form 12A, being used as a covariant or controlling variable.

The beta Formula was used to adjust posttest scores of the experimental and control groups to allow for initial
differences between the experimental and control groups on pretest scores.

Conclusion

The analysis of the effects of instruction and practice in test-taking skills upon the standardized reading test scores of randomly-selected second-grade students from a Title I School indicated that the experimental group achieved significantly higher test scores ($p < .01$) than the control group.

The data from this research tend to confirm the hypothesis that instruction and practice in test-taking skills will improve the scores on standardized reading tests of randomly-selected, culturally-different, second-grade students from a Title I School.

Educational Implications

Elementary schools are depending upon the standardized test with great frequency for the purposes of diagnosing individual differences; and for the purpose of grouping children according to their abilities (Goslin, et al., 1965). Measurement experts and educators have suggested that test-wiseness is a possible source of variance in the standardized test score (Thorndike, 1949; Ebel and Damrin, 1960; Anastasi, 1969). The data developed in this study tend to support the suggestion that test-wiseness is a
source of variance in standardized test scores. In order to control the test score variance for which test-wiseness may be responsible, practice and instruction in test-taking skills should be taught in the elementary school classrooms.

The research by Eels (1951), Brim (1965), Anastasi and Cordova (1953), Ortar (1960) all tend to demonstrate that culturally different children tend to lack test-taking skills. The research by Mercer (1973) indicates that culturally different children are being labeled as retardates on the basis of one standardized I.Q. test. It is clear that great caution should be used by educators in the evaluation of the standardized test scores of culturally-different children.

The educational implications which are suggested by this research are:

1. Test-taking skills should be taught in the classrooms in order to equalize the opportunities of all children who are to be tested with a standardized test.

2. Great caution must be exercised by educators when interpreting the test scores of culturally-different children.
Recommendations

The following suggestions for further research are:

1. The results of this research were based on a population of 43 culturally different students. It is suggested that further research on test-wisness skills to culturally different subjects employ a larger number of subjects in their sample.

2. The practice materials in the experimental treatment were based on the format of the Stanford Achievement Test, Reading 1, Form W. It is suggested that further research on test-wisness skills use practice materials based on other primary reading achievement test.

3. It has been suggested by Vernon (1954) and Anastasi (1969) that test-wisness tends to alleviate test-anxiety. It is suggested that further research on test-wisness employ an instrument to measure the extent of test anxiety of students who have been taught test-taking skills, as compared to students who have not been taught test-taking skills.

4. It has been demonstrated by Callenbach (1971) and Kalechstein (1974) that test-taking skills can be taught in the elementary classroom. It is suggested that materials and programs be developed to accomplish the objective of teaching all elementary school children test-taking skills.
REFERENCES


Eels, K., David, H., Havighurst, R., Herrick, U., & Tyler, R. Intelligence and cultural differences, University of Chicago Press, 1951.


Greene, K. B. The influence of specialized training on tests of general intelligence. 27th Yearbook of the National Society for the Study of Education, 1928,


McIntosh, D. The effects of practice in intelligence test results. British Journal of Educational Psychology, 1944, 14, 44-45.


Smith, S. Language and non-verbal test performance of racial groups in Honolulu before and after a fourteen year interval. Journal of General Psychology, 1942, 26, 51-93.


Thorndike, R. L. The effect of interval between test and retest on the constancy of the I.Q., 1933, 24, 543-549.


Introduction to Experimental Lessons

The following eight lessons were constructed to provide instruction and practice in 1) following orally administered directions; 2) response marking; 3) using time efficiently; and 4) guessing.

The administration directions accompanying the practice and/or testsheets are detailed and self-explanatory and are intended to insure a test-taking atmosphere during the use of the materials.

Teacher-pupil discussions preceding or following practice, or whenever the investigator deemed discussion necessary, were used to instruct and reinforce test-taking skills.

Discussions focusing on the necessity of following orally administered directions and making responses properly stressed:

1) That standardized tests contain numerous directions that are read orally to the pupils;
2) That in order to answer the questions correctly, the pupils must know the proper way to respond;
3) That in order to answer the questions, the pupils must understand the directions;
4) That different parts of standardized tests require different kinds of responses; and
5) That if the pupil does not understand the directions, or proper way to mark responses, he should immediately ask his teacher for clarification.

Discussions focusing on using time efficiently and guessing stressed:

1) That pupils should complete answers which they are sure of first;
2) That pupils should then do the more difficult and time-consuming questions;
3) That pupils should guess only after an honest attempt has been made to answer the questions;
4) That pupils should respond to all questions; and
5) That pupils should check answers during any remaining time to assess correctness and to avoid careless mistakes.
Experimental Lesson I

READ TO THE PUPILS:

For the next four weeks we are going to be doing some things that will help you learn to do better on tests like the one you took last week. (Hold up a copy of the Stanford Reading Test.) I will teach you some things you can do to make better scores on these tests, and we will practice them together.

Each time we meet together we will be working on things in a test booklet. (Hold up the Experimental Booklets.) Today I am going to give each of you a booklet and crayon; do not write on the booklet or open it until I tell you to do so.

Pass out the crayons and booklets.

Do not open your booklet until I tell you to do so.

Look at the front page where it says "My Name is __________." Write your name on the line.

Now listen carefully to what I tell you. Open your booklet to page one. Find the top box with the word SAMPLE in it. Put your finger on the box.

Now I will tell you what is in the box. You see a triangle, a circle, another triangle, and a square. I am going to tell you what to do in this box. I want you to color the circle . . . Color the circle.

Check to see that the pupils have colored the proper figure.

Now I am going to tell you what to do in the other boxes on this page. You must listen carefully. I will say what I want you to do two times. Listen carefully both times.

Look at box 1. In box one there is a circle, a triangle, a square, and another square. Color the triangle . . . Color the triangle.

Look at box 2. In box 2 there is a triangle, a square, a circle, and another circle. Color the first circle . . . Color the first circle.

Look at box 3. In box 3 there is a square, a triangle, a circle, and a square. Place a cross (X) in the second square . . . Place a cross (X) in the second square.
Look at box 4. In box 4 there is a triangle, another triangle, a circle, and another circle. Draw a line under the first triangle. Draw a line under the first triangle.

Look at box 5. In box 5 there is a square, another square, a circle, and a triangle. Draw a line under the second square. Draw a line under the second square.

Look at box 6. In box 6 there is a triangle, a circle, another circle, and a triangle. Color the first triangle. Color the first triangle.

Stop.

Discussion of answers. See Introduction to Experimental Lessons.

Now turn to the next page of the booklet. Turn to page 2. Here are some more boxes. Find the top box with the word SAMPLE in it. Put your finger on the box.

What do you see inside the box? That's right. A triangle, a circle, another triangle, and a square.

Now listen carefully to what I tell you to do. If you made mistakes before, listen very carefully.

Color the square. Color the square.

Check to see that the pupils have colored the proper figure.

Now look at box 1. Place a cross (X) inside the first square. Place a cross (X) inside the first square.

Now look at box 2. Draw a line under the triangle. Draw a line under the triangle.

Look at box 3. Place a cross (X) inside the second square. Place a cross (X) inside the second square.

Look at box 4. Color the two triangles. Color the two triangles.

Look at box 5. Draw a line under the circle. Draw a line under the circle.
Look at box 1. Place a cross (X) inside the first triangle. Place a cross (X) inside the first triangle.

Discussion of answers. See Introduction to Experimental Lessons.

Collect materials.

Experimental Lesson II

READ TO THE PUPILS:

I am going to give you your test booklets again. Be sure you have your own booklet. I will also pass out a crayon to each of you.

Now listen carefully to what I tell you. The last time I was here we worked with some circles, triangles, and squares, and we learned that it is very important to listen carefully to test directions. Today we are going to practice some more. Open your booklet to page 3. Find the top box with the word SAMPLE in it. Put your finger on the box.

Check to see that the pupils have located the proper box.

Now I will tell you what is in the box. There is a baseball and a bat, a fish, and a banana. I will tell you what to do in this box. Draw a circle around the banana. Draw a circle around the banana.

Check to see that all pupils have circled the proper picture.

Now I am going to tell you what to do in the other boxes on this page. You must listen carefully. I will say what I want you to do two times. Listen carefully both times.

Look at box 1. In box 1 is a pair of sandwiches, a dog, a sun. Mark a cross (X) on the sun. Mark a cross (X) on the sun.

Look at box 2. In box 2 is a pair of socks, some buttons, a table. Draw a line under the things your mother might sew on your shirt. Draw a line under the things your mother might sew on your shirt.
Look at box 3. In box 3 is a goat, a doll, a boat. Draw a circle around the thing that travels on water. Draw a circle around the thing that travels on water.

Look at box 4. In box 4 is a pair of mittens, some dishes, a duck. Draw a line under the things you wear when the weather is cold. Draw a line under the things you wear when the weather is cold.

Look at box 5. In box 5 is a goat, a fox, and a deer. Mark a cross (X) on the animal with the largest tail. Mark a cross (X) on the animal with the largest tail.

Look at box 6. In box 6 is a pretzel, a piece of meat, and a pair of mittens. Draw a circle around the thing next to the mittens. Draw a circle around the thing next to the mittens.

Stop.

Discussion of answers. See Introduction to Experimental Lessons.

Now turn to the next page. Turn to page 4 of the test booklet. Here are some more boxes. I will tell you what to do. Listen carefully to my directions. Try to get them all right.

Look at box 7. Draw a line under the baseball and bat. Draw a line under the baseball and bat.

Look at box 8. Draw a circle around the thing that crows. Draw a circle around the thing that crows.

Look at box 9. Draw a line under the comb. Draw a line under the comb.

Look at box 10. Mark a cross (X) on the key. Mark a cross (X) on the key.

Look at box 11. Draw a circle around the thing next to the dog. Draw a circle around the thing next to the dog.

Look at box 12. Draw a line under the thing used by football players. Draw a line under the thing used by football players.
Look at box 13. Draw a line under the thing that tells time . . . Draw a line under the thing that tells time.

Stop.

Discussion of answers. See Introduction to Experimental Lessons.

Collect materials.

Experimental Lesson III

SAY TO THE PUPILS:

The last two times I was here with you we talked about how important it is that we listen carefully to directions when taking tests. We practiced listening. Today we are also going to listen carefully, but we are going to do some new things too.

I am going to give each of you your test booklet and a pencil. Be sure that you have your own booklet. Do not open it until I tell you to do so.

Pass out the pencils and the booklets.

Now open your booklet to page 5. Look at the top of the page. In the box is a number problem. The number problem says "1 plus 2 is what number; 1 plus 2 is what number." $3 - 1 = 2$

What is the number? That's right. Three. Now see that the "3" has been crossed out in the row of numbers under the number problem. The right answer has been crossed out. In order for the answer to be right, you must cross out the "3."

Now I want you to finish the number problems on this page. You may not be able to answer all of the problems. Do not worry about this. We will talk about the number problems after you have finished this page. Do not turn to the next page. Stop when you have finished this page.

Begin.

Walk around the room observing the pupils. When nearly all of the pupils have completed the page, say
STOP.

Discussion of answers. See Introduction to Experimental Lessons.

Now turn to page 7. Here are some more number problems. Place your finger on the box with SAMPLE written in it. Help me do the sample. The number problem says "3 plus 2 is what number; 3 plus 2 is what number."

That's right. Cross out the "5." Now finish the page. Try very hard to remember what we have just talked about.

Begin.

Walk around the room observing the pupils. When nearly all of the pupils have completed the page, say

STOP.

Discussion of answers. See Introduction to Experimental Lessons.

Collect materials.

Experimental Lesson IV

READ TO THE PUPILS:

I am going to give each of you your test booklet and a pencil. Do not open the booklet until I tell you to do so.

Pass out the pencils and booklets.

Today we are going to practice some more number problems. Open your booklet to page 8. Look at the top box. Read the number problem with me: "4 plus 2 is what number; 4 plus 2 is what number."

Will somebody tell me the answer? That's right. Six. Now look at the numbers beside the circles. You see a "13," an "8," a "6," and "104." Look at the cross (X) marked in the circle beside the "6." The cross (X) means that "6" is the right answer.