Order Number 9423737

The effects of teaching practice review items and test-taking strategies on the ACT mathematics scores of second-year algebra students

McMann, Patricia Kovach, Ed.D.

Wayne State University, 1994

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

THE EFFECTS OF TEACHING PRACTICE REVIEW ITEMS AND TEST-TAKING STRATEGIES ON THE ACT MATHEMATICS SCORES OF SECOND-YEAR ALGEBRA STUDENTS

by

PATRICIA KOVACH MC MANN

DISSERTATION

Submittted to the Graduate School

of Wayne State University

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF EDUCATION

1994

MAJOR: ADMINISTRATION AND SUPERVISION - GENERAL

Approved by:

Edward Simpkins 3/3/94

Adviser Date

Jellilo Boyer

Burnis Hall

____

DEDICATION

I dedicate this study to my maternal grandparents, John Robba and Mary Tartal Robba and my paternal grandparents, Paul Kovach and Victoria Bator Kovach. They had the courage and the foresight to leave their native land of Czechoslovakia to come to America with the hope of building a better life for the generations to follow. In their wisdom, they understood the requisite of hard work and the need for a good education to make that dream a reality.

I also dedicate this study to my parents, Frank Kovach and Elizabeth Robba Kovach. To achieve any goal I set, they always encouraged me to believe in myself while not having to follow a traditional path. I thank them for instilling those same values of hard work and the importance of a good education in me.

Finally, this dissertation is dedicated to my husband Lawrence, and to my children, Larry and Ann. I could not have completed this goal without their tremendous love, patience, loyalty, and encouragement throughout these years of graduate study.

ACKNOWLEDGEMENTS

The author wishes to express her deep gratitude for the assistance and concern offered by the many people who have helped to make this dissertation possible.

Sincere thanks to my advisor, Dr. Edward Simpkins, for his counsel, support, guidance, and encouragement through out this project. Special thanks and appreciation to the other committee members for their expenditure of time:

To Dr. Joella Gipson-Simpson, my cognate advisor, for her suggestions and for providing me with the impetus to return to graduate school, and encouragement to see this goal to completion.

To Dr. Arthur Brown, for serving on my committee and his input on the paper.

To Dr. Burnis Hall, for his willingness to lend a constructive ear at any time and especially for serving on my defense committee.

To Dr. Donald Marcotte, for moderating for my defense.

A gigantic thank you goes to my colleagues, Margaret Butz, James Gilhool and Beth Ann Mace who participated in the study and helped to gather the data for this paper. A special word of appreciation and thank you to Margaret Butz, who helped with the writing and editing of the practice ACT questions. I appreciate their willingness to support me with this project.

Thank you can not begin to cover the kindness and support that I received from my dear friends in Pennsylvania, Dr. Stephen M. Priselac and Dr. Nancy J. Priselac. I appreciate Steve's expertise and knowledge; and all the endless hours he spent helping me categorize and analyze the statistical data from the study. Nancy was invaluable in helping me with her doctoral expertise and for her proof-reading skills in pouring over the paper countless times. Her guidance, counsel, and example through the years have been an inspiration to me.

Finally, the author wishes to express a thank you to her husband, Lawrence, and her son, Larry for the countless hours of help with computer formatting and critique of this paper.

I have been truly blessed to have been surrounded and supported by all these fine people.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

THE PROBLEM AND ITS SETTING

Overview

The roots of testing go back to 1854 when the Boston City School Trustees adopted a written examination to supplement traditional oral examinations which had become too unwieldy with the growth of the school population. Unfortunately with the testing, the schools got a message that they had not bargained for: a high failure rate, which indicated "deficiencies of instruction and perhaps excessive difficulty in the curriculum" and the genesis of test scores as a tool to make schools accountable to the public authority at the state level.[1] What followed was a growing dependence on standardized tests to serve as a report card on American education--an idea that still exists today.

After the turn of the century, there was wide-spread use of standardized testing for a variety of purposes--some of which are controversial by today's standards. Intelligence tests like the Stanford Binet were used to sort students into programs where they would be less likely to fail. During World War I, nearly two million recruits were given intelligence tests. It was discovered that "intelligence" and "non-intelligence" were equally distributed among all classes. With the discovery that intelligence was randomly

[1]Daniel and Lauren Resnick, "Standards, Curriculum and Performance: A Historical and Comparative Perspective," Educational Researcher (1987):5-20.

distributed among the rich as well as the poor, testing became a device that could be used for breaking through the class system and removing some of the social stratification built into our society. The second result of military testing was that large numbers of high school graduates scored alarmingly low, which sparked a debate about the failings of our schools.[2] As a result, school administrators looked for ways to diagnose school problems and justify subsequent solutions. Analysis of standardized test scores appeared to be an appropriate method to evaluate programs. After the war, educators turned to the hundreds of psychologists who had worked on the military intelligence tests for help in developing educational testing.

In the 1960's, standardized tests were used to measure the effects of federally funded compensatory education programs. At the state level, tests were used to assess the impact of the increasing amounts of money being earmarked by legislatures for public schools.

In the 1970's, standardized test scores provided the first systematically gathered evidence of declining academic achievement, but many dismissed it as misleading and unimportant. However, other evidence soon mounted to corroborate the decline, convincing even the most doubtful that reform was needed. Concern at the national level produced committees to study the problems and reports like A

---

[2]Ibid.

Nation at Risk were published. By the mid-1980's, with new reforms under way, standardized testing took on an increased importance. Policymakers and educators, some previously skeptical of standardized testing, embraced test scores as a way to identify problems and validate solutions.

Testing has been influential in shaping education in several ways. The history of testing showed that testing could be used to assess the effectiveness of teaching and learning. Results of tests could be used to compare schools, school districts, colleges, states, and even countries in determining effectiveness in educating students. When comparisons are made, tests can be used to hold individuals accountable for leadership, teaching, and learning.[3]

Tests can determine what curricula are taught. According to the research, there is validity to the old saying that "What gets tested gets taught." Items that are not tested in some manner tend to disappear from classroom instruction. Testing can be used to determine whether a student graduates from high school. The scores on an exit test for graduation measure whether the student has achieved the level of learning necessary to be certified as competent.[4]

Testing has been used to determine college admission. Such tests measure levels of learning, the information and

---

[3]Kenneth H. Ashworth, "Standardized Testing: A Defense," The College Board Review (Winter 1989-90):23-24.

[4]Ibid.

tools students have acquired and their skill in manipulating ideas and symbols. These tests can measure how quickly the student can comprehend and make essential connections in reading and analyzing material. Colleges and universities use standardized testing as one measure of determining the student's probability of success in pursuing a degree.

Purpose of the Study

The primary purpose of this study was to determine whether student scores on the mathematics portion of the American College Testing Program (ACT) test could be improved by practicing sample test items and studying general test-taking strategies.

There were two secondary purposes included in this study. One was to determine whether a student's gender has any significance in ACT performance and the other was to determine whether the previous number of years of mathematics study affected achievement levels.

To accomplish these objectives, a comparison of ACT scores on a Practice Mathematics Exam prepared by the American College Testing Program was made among eight sections of second-year algebra students. Students in four of the sections were taught test-taking strategies and the review practice test items, while students in the four control groups were taught second-year algebra without the use of ACT practice materials.

Significance of the Study

This study was significant in that it may establish a relationship between ACT test performance and test-taking preparedness in mathematics. It may be advantageous for students who have content knowledge to be able to increase their ACT score by the use of review items and test-taking strategies.

Secondly, this study was significant in that it has the potential to provide school districts with an avenue to improve their standardized test scores on the mathematics portion of the ACT. Recent legislation in the state of Michigan, Public Act 25, (1990) has mandated that all districts prepare an annual report for their public that explains/compares student outcomes on standardized tests like the ACT.

Thirdly, the study was significant in that it might establish a relationship between the number of years of mathematics taken and the student's pre-disposition to score higher on the ACT.

Fourthly, this study was significant because of its potential contribution in the area of career education and its relationship to the college-preparatory curriculum. If students realize their highest potential on the test, they may be better prepared to assess their potential for success in a specific occupational area.

The two major standardized tests used for college admission in the United States are the American College Testing Program (ACT) and the Scholastic Aptitude Test (SAT). Most Michigan colleges and universities require the ACT test rather than the SAT. The ACT is a national test that provides high schools and students with information that has been referenced on a national norm. Most school districts are concerned about how well their students score on it. The message that the scores send to parents and the community as a whole is that schools that have high ACT and SAT scores are better schools. Individual high school students and their parents are concerned that they achieve the score required for admission to the college or university of their choice.

Statement of the Problem

The American College Testing Assessment Program (ACT) contains four curriculum-based tests that measure academic achievement in the areas of English, mathematics, reading, and the natural sciences. These tests are based on and oriented towards major areas of secondary and postsecondary instructional programs. Performance on these content tests is an indicator of a student's academic development. Students are assessed on their ability to apply content knowledge and reasoning skills acquired in their coursework to materials similar to those they will encounter in a college setting. The four tests of the ACT are designed to measure the student's preparedness to profit from postsecondary education.

Our entire educational system was under fire because the United States fared so poorly on the international assessment tests in the early 1980's. One of the five areas that the report, A Nation At Risk, focused on was standards and expectations. Since 1983, many colleges and universities changed their entrance requirements. Along with increased admission standards, the National Collegiate Athletic Association (NCAA) raised its minimum standards for athletic eligibility. The guidelines in Proposition 48 require that all prospective college athletes have a minimum grade point of 2.0 on the four point scale, have completed a core of eleven academic courses, and have a specific minimum score on the college board exams.[5]

This factor has made the ACT increasingly important since the National Collegiate Athletic Association has ruled that all high school athletes must score at least 18 on a scale of 33 on the ACT to be eligible to receive an athletic scholarship at any college or university. Parents whose son or daughter has athletic talent and hopes to earn a scholarship must also be concerned about their student's academic preparation and now are applying pressure on schools to make sure that their child is academically prepared to compete.

In Michigan, Public Act 25 has modified the School Code to require administrators to report student achievement results to include the ACT/SAT scores annually to the community. There is increasing pressure to raise ACT scores since every high school can be compared quickly on this

---

[5]Emeral A. Crosby, "The 'At Risk' Decade", Phi Delta Kappa 74 (8) (April 1993):600.

national norm.

How students can best be prepared to achieve at their full potential on the ACT is a question the faces each high school. The most important prerequisite for optimum performance by students on the ACT test is a sound, comprehensive educational program. Because the ACT is based on curriculum in the four major areas of English, Mathematics, Reading, and the Natural Sciences, the best way for students to perform at their best is to apply themselves fully to the learning activities (courses) provided by their school curriculum. Students who plan to enroll in college should be taking college-preparatory classes.

Content knowledge alone will not necessarily guarantee success. Students' performance on the ACT tests may also be affected by test anxiety and the student's inability to work with the multiple-choice question format. The multiple-choice format of the ACT may be unfamiliar to the student. Some mathematics courses do not provide students with any opportunity to work with multiple-choice test questions. Mathematics instructors expect students to work out the solution to a problem and show all the steps as evidence of understanding the mathematical concept.

This study is designed to measure the effects of spaced practice on sample ACT mathematics questions and test-preparation suggestions for the mathematics section score of the ACT on students enrolled in a second-year algebra class.

Research Questions

This study of coaching for a standardized achievement test like the American College Testing Program (ACT) using practice test items and test-taking strategies was designed to answer the following questions:

*Research Question One: Will the use of practice test items similar to the mathematics questions given on the American College Testing (ACT) exam and test-taking suggestions significantly improve the ACT mathematics score of second-year algebra students?

*Research Question Two: Does gender have any significance in the performance of second-year algebra students on the mathematics section of the ACT?

*Research Question Three: Does the number of years of prior mathematics study affect achievement on the mathematics section of the ACT?

*Research Question Four: Does the level of student's prior achievement have any effect on the student's performance?

Figure 1. Research model outline used to study coaching effects on the mathematics section of American College Test (ACT).

Phase I PREPARATION OF MATERIALS

Phase II PRETESTING

Phase III IMPLEMENTATION OF STUDY

Phase IV POSTTESTING

Phase V DATA ANALYSIS

Phase VI REPORT OF DATA ANALYSIS

Definition of Terms

Achievement Test: An instrument used to measure the proficiency level of individuals in given areas of knowledge or skill.

ACT-American College Testing Program: An achievement testing program used as a reference for college admission that measures students potential in the four curriculum areas of English, mathematics, reading, and the natural sciences. ACT is a registered trademark of The American College Testing Program, Iowa City, Iowa.

Aptitude Test: An instrument used to predict performance in a future situation.

Coaching: Webster's definition of coaching is to instruct a person in a subject, or prepare a person for an examination by private tutoring. Preparation for tests can mean orientation to general test-taking skills, review of subject matter, drill and practice on sample test items, teaching specific strategies or tricks and anxiety reduction skills, or the development of concepts and competencies over time.

Control Group: The group in a research study that receives no treatment or the group that is treated "as usual".

Criterion-Reference Instrument: An instrument that specifies a particular goal, or criterion, for students to achieve.

Experimental Group: The group in a research study that receives the treatment (or method) of special interest in the study.

Norm-referenced Test: A test designed to provide a measure of performance interpreted in terms of the person's relative standing in some known group.

Practice Test: a test that has been modeled after another exam with similar questions that is used to build familiarity and skill in an area for a particular examination like the ACT/SAT.

Public Act 25: 1990 Legislation in Michigan that promised to improve education by affecting six major changes in the School Code--annual report, school improvement, accreditation, core curriculum, Intermediate School District extensions, and the hiring of non-certified personnel when certified personnel are not available.

Reliability: The degree to which scores obtained with an instrument are consistent measures of whatever the instrument measures.

SAT-Scholastic Aptitude Test: An aptitude test that supposedly measures innate ability in the subject areas of mathematics and verbal (English) skills. The SAT is used to predict success in college and as a tool for college admission officers. SAT and The Scholastic Aptitude Test are registered trademarks of the College Entrance Examination Board (CEEB), New York, New York.

Teacher Identification Code: Teacher One taught the Eleventh Grade Honors Algebra II Control Group.

Teacher Two taught the eleventh grade honors Algebra II experimental group. Teacher Three taught the tenth and eleventh grade intermediate Algebra II control group. Teacher Four taught the tenth and eleventh grade intermediate Algebra II experimental group.

Validity: The degree to which correct inferences can be made based on results from an instrument; depends not only on the instrument itself, but also on the instrumentation process, and the characteristics of the group studied.

Assumptions

There are four assumptions underlying this study:

The first assumption was that students can be taught test-taking strategies and skills. Coffin makes the statement, "Increasing evidence shows the intensive preparation for the SAT has a positive impact on scores. ... Private counseling, coaching, and tutoring centers are springing up everywhere."[6]

The second assumption was that practice can lead to improved achievement. "Coaching does make a difference statistically was well as practically", according to Andrew Porter, a psychologist and principle associate in medicine at Harvard Medical School and Boston's Beth Israel Hospital. "Generally," says Porter, "the longest and most challenging courses that include homework, practice tests, and test-

[6]Gregory C. Coffin, "Computers Can Help Students Improve SAT Scores, " NASSP Bulletin, 72 (October 1988): 78.

taking strategies lead to the best results."[7]

The third assumption was that the reporting of ACT and SAT scores is important as one indicator of quality education since it is a score that is nationally recognized and is used as a measure for comparison in most documents that report student achievement. Administrators are interested in being able to report the best possible scores to their public. Section 1204a, Public Act 25, State of Michigan, requires that local districts prepare an annual educational report that must be made available to the State Board of Education and to the public at an open meeting. Included as one of the components of the annual reports is the reporting of student achievement that includes the mean score on college entrance exams like the ACT and SAT.[8]

The fourth assumption was that presenting test-taking strategies to improve students' test-taking abilities is a worthwhile skill to teach. Students must pass tests within and at the end of most courses in order to receive a passing grade for the class. Hymel and Guedry-Hymel discuss the necessity for the principal to take responsibility for inclusion of study skills (SS) and test-taking techniques (TTT) in their role as instructional leader. They propose a

---

[7]Andrew Porter, "External Standards and Good Teaching: The Pros and Cons of Telling Teachers What to Do," Educational Evaluation and Policy Analysis, 11 (1989):354.

[8]Michigan Department of Education, "Annual Report Carries $ Incentives and Noncompliance Penalty," Michigan Education Report (September 1990):2.

three-dimensional model that includes curriculum, instructional guidance, and assessment. Hymel and Guedry-Hymel feel that, "An underlying assumption of this model is that the instructional leadership function of the school administrator must include attention to SS & TTT, since they represent student behavior patterns that enhance the acquisition, retention, and transfer/application of learning across the cognitive, affective, and psychomotor domains."[9]

Chapter Summary

The concept of testing has been around for approximately one hundred fifty years and has provided the educational community with information concerning individual student progress or has served as a standard for judgement about academic excellence. This paper will discuss both issues historically, look at the current research of using review and practice tests as learning tools, and discuss standardized testing such as the Scholastic Aptitude Test (SAT) and American College Testing Program (ACT) being used as a score card to determine the quality of education in a given school or state or as a criterion for admission to and or predictor of success in college.

[9]Glenn M. Hymel and Linda Guedry-Hymel, "Promoting Study Skills and Test-Taking Techniques," NASSP Bulletin 71 (October 1987):97.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

The use of coaching and test-taking skills instruction have been issues during the last ten years as schools were compared internationally. The pressure is on for our educational system to be more accountable to constituents and to produce better results with students. Reports on all levels--national, state, and local, include information on the college testing services scores as a way of providing a benchmark to compare the quality of education. Use of college testing scores as a report card is an expanded purpose since originally the purpose of the college testing services was to provide a standard measure of ability to assist colleges in making admission decisions.

National Testing Implications

Over the past ten years, the efforts to improve American schools have focused on the use of standardized tests as measures of student achievement and as a predictor to make decisions about student placements, teacher competence, and school quality. Some recent polices have sought to "hold schools accountable" by using test scores to trigger rewards, sanctions, or initiate remedial actions.[10]

The evidence now available suggests that, for the most

[10]Linda Darling-Hammond, "The Implications of Testing Policy for Quality and Equality," Phi Delta Kappan 72 (November 1991):220-225.

part, these testing policies have not produced the positive effects that they were intended to bring about. Instead, they have had many negative consequences for the quality of American schools and for the equal opportunity to schooling opportunities for all students. The negative effects are stem partly from the nature of American tests and partly from the ways in which the tests have been used for educational decision making.[11]

In America 2000: An Educational Strategy, Secretary of Education, Lamar Alexander, stated that in order to have an effective education system, we must know how much each child knows. He suggested that parents have a right to know whether or not their child understands what is needed to be a scientist in the 21st century or what is needed to be a competitive worker for the world marketplace. Our testing does not provide information about the quality of education. Our tests tell us which students know the most about the questions asked and which students will do the best on future scholastic assignments. They do not tell us what students know in general. Current testing practices provide valid generalizations about how students measure up against one another.[12]

Caution is advisable when making inferences from test

---

[11]Ibid.

[12]U.S. Department of Education, "America 2000: An Educational Strategy," Government Printing Office (April 1991): 4-6.

scores. William Mehrens, Michigan State University, wrote,

> The only reasonable, direct inference you can make from a test score is the degree to which a student knows the content that the test samples. Any inference about why the student knows the content to that degree...is clearly a weaker inference...[13]

Former President Bush discussed a plan to establish mandated, standardized testing as means of improving education and as a means of holding schools accountable. The underlying assumptions for the plan were that uniform testing would improve educational instruction as a whole, and benefit both students and teachers in the process. A national test would measure the most important outcomes of learning and would become a standard for the public to measure the success or failure of the system. A new national test will not help teachers to teach or provide information on the effectiveness of education as long as the cost of testing prohibits the administration of anything but paper and pencil, machine-scoreable tests to students at the state or national level.[14]

Commercial publishers and non-school agencies produce norm-referenced, multiple-choice instruments that are designed to rank and sort students cheaply and efficiently. These

---

[13]William A. Mehrens, "National Tests and Local Curriculum: Match of Mismatch?," Educational Measurement: Issues and Practice 3 (March 1984):10.

[14]Douglas Archbald and Fred M. Newmann, Beyond Standardized Testing: Assessing Authentic Academic Achievement in the Secondary School (Reston, VA.: National Association of Secondary School Principals, 1988), pp. 43-51.

instruments were not constructed to support or enhance instructions and they cannot measure many kinds of knowledge or performance skills that we expect or value for students in our technological world today.[15] Current research on performance and human learning indicates that most tests currently in use fail to measure students' ability to use higher-order thinking skills or to measure the student's ability to perform real-world tasks.[16]

In his study of schools in the early 1980's, John Goodlad noted that the trend was away from teaching students to think. He felt that the influence of basic skills tests and the importance given to the reporting of scores was contributing to this decline. Due to the pressure of state and district testing, Goodlad found that most students listened to or read short sections in textbooks, responded briefly to questions, and then took short answer or multiple choice quizzes. Rarely are American students asked to initiate anything, create their own projects, read or write an essay, or participate in analytical discussions.[17]

Ernest Boyer's research, during the same period, found that teachers were under great pressure to teach the skills that

---

[15]Alexandra K. Wigdor and Wendell R. Garner, eds., Ability Testing: Uses, Consequences, and Controversies (Washington, D.C.: National Academy Press, 1982), pp. 17-23.

[16]Lauren B. Resnick, Education and Learning to Think (Washington, D.C.: National Academy Press, 1987), pp.72-77.

[17]John I. Goodlad, A Place Called School: Prospects for the Future (New York: McGraw-Hill, 1984), p. 145.

were being tested. Boyer found that teachers were teaching the concepts on the nationally normed tests that were being reported to the public to the exclusion of other important skills that were not or could not be tested.[18]

The National and State Perspective of Accountability

In 1867, the United States Congress created the Department of Education. There was very little public interest in tracking educational progress until the late 1950's and early 1960's, when the launching of the Soviet satellite, Sputnik, and the civil rights movement caused concerns about our educational system. Reports about school reform shocked the nation again in the 1980's. The public would gather and use any available information to monitor the performance and determine the quality of the American educational system.[19]

A Nation at Risk was published by the National Commission on Excellence in Education in 1983. Terrel Bell, then Secretary of Education, introduced the first annual "wall chart", that compared states on a number of categories (e.g. students' SAT and ACT scores, graduation rates, teacher salaries, pupil/teacher ratios, expenditures,

---

[18]Ernest L. Boyer, High School: A Report on Secondary Education in America (New York: Harper & Row, 1983), pp.67-68.

[19]R. W. Selden, "Missing Data: A Progress Report from the States," Phi Delta Kappan 69 (July 1988):492-494.

etc.).[20] Ten years later, assessment and standards have not faded from the picture. In the April 1993 Phi Delta Kappa, Terrel Bell wrote about the "driving mechanism" to reward and motivate the attainment of higher national standards in education. In the years ahead, Bell believes we will have less testing but more effective assessment, tied to the national standards.[21]

The collection of performance data and the emphasis on accountability has come about because many states have increased aid to local school districts since 1981.[22] In an effort to improve the quality of education, many states have developed public report cards. California is one of the states that began to produce a report at both the state and local district level in 1985. California performance district reports include the following information: academic course enrollment; units required for graduation; academic achievement test results; drop-out rate; attendance rate; percentage of students taking the SAT and ACT; ACT, SAT, and Advanced Placement scores; performance of graduates

---

[20] J. Oakes, Educational Indicators: A Guide for Policymakers, (New Brunswick: Center for Policy Research in Education, 1986).

[21] Terrel H. Bell, "Reflections One Decade After A Nation At Risk," Phi Delta Kappan 74 (April 1993):596.

[22] U.S. Department of Education, Measuring Up: Questions and Answers about State Roles in Educational Accountability, Report of the OERI State Accountability Study Group, (Washington, D. C. Government Printing Offices, 1988b).

attending state colleges and universities; instructional time; distributions by sex and race/ethnicity of academic course enrollments, test results, and college-going rates; amount of homework; and student mobility.[23]

The issue of accountability and the use of testing as a basis for making judgements on the quality of education has included both national tests of the Scholastic Aptitude Test (SAT) and the American College Test (ACT) as a gauge or meter stick for comparison. Since the SAT and the ACT came into existence, they have been used as one of the available and reliable sources of information concerning student outcomes and the quality of education in the school district, the state, or the nation.

Michigan Legislation--Public Act 25 of 1990

In the State of Michigan, a landmark piece of educational legislation became part of the School Code on March 13, 1990, with then Governor James J. Blanchard's signature. Michigan State Representative James E. O'Neill, Jr. (D-Saginaw) introduced House Bill 4009 to the legislature in January 1989. The bill, known as Public Act 25, passed the House of Representatives in June 1989 and the Senate in February 1990.

Known as the "quality education package," Public Act 25 was hailed by the then Superintendent of Public Instruction,

---

[23] S. S. Kaagan and R. J. Coley, State Education Indicators: Measured Strides Missing Steps, (Princeton, N. J.: Center for Policy Research in Education, 1989).

Donald L. Bemis, as a "momentous piece of legislation that will drive comprehensive school reform in our state for this decade and beyond."[24]

Public Act 25 promised to impact and improve education through six major changes in the School Code:

1. Section 627 extends and describes what an intermediate school district (ISD) may do to serve its constituents.
2. Section 1204a requires local districts to prepare an annual educational report for each school in the school district and make it available to the State Board of Education and to the public at an open meeting.
3. Section 1233b allows school districts to hire non-certified individuals to teach computer science, a foreign language, mathematics, biology, chemistry, engineering, physics, or robotics in grades 9-12 under certain conditions.
4. Section 1277 asks districts to adopt and implement a three-to-five year school improvement plan for each school within the district.
5. Section 1278 asks districts to establish a core curriculum based on the district mission statement and make it available to all pupils in the district.

---

[24]Michigan Department of Education, "Quality Education Package Will Drive Reform," Michigan Education Report (September 1990):1.

6. Section 1280 asks that each school within a school district be accredited.[25]

The Annual Report

Each administrator and superintendent in the State of Michigan must be concerned with preparing the Annual Report as required by Section 1204a of Public Act 25. One of the key components of Public Act 25, the annual report, was designed to help citizens become more informed about their schools and to bring more accountability into the educational process. It is the only component of PA 25 that carries a financial penalty for a school district that does not comply. The report must address seven aspects of the educational program of each school within the district:

*School Improvement

*Core Curriculum

*Student Achievement

*Student Retention

*Accreditation Status

*Specialized Schools

*Parent-Teacher Conferences

Annual Education Reports must be prepared and reported on beginning with the 1990-91 school year.[26]

Student Achievement in Michigan

---

[25]Ibid.

[26]Michigan Department of Education, "Annual Report Carries $ Incentives and Noncompliance Penalty," Michigan Education Report (September 1990):2.

According to Public Act 25, the section of the annual report on student achievement should contain an aggregate of student achievement based on the results of any locally administered student competency tests, statewide assessment tests, or nationally normed achievement tests. The results for both the current and previous year must be reported. Where possible, the data should show a three-year comparison. Examples for the types of test scores that may be used in the report are:

* Local school district competency tests that are used for promotion or graduation decisions.
* Michigan Educational Assessment Program (MEAP) results for mathematics and reading administered to students in grades four, seven, and ten and for science in grades five, eight, and eleven.
* Nationally normed achievement tests, including commercially available tests given to all students in grades K-12.
* College entrance tests such as the American College Testing Program (ACT) and the College Board's Scholastic Aptitude Test (SAT) and also to include the Preliminary College Testing of the P-ACT and the PSAT
* Portfolio assessment.[27]

Implications in High Stakes Testing

All too often testing is portrayed as uniquely good or

---

[27] Ibid.

uniquely bad; but the truth is that testing is a two sided coin--good and bad. On the plus side, tests can influence curriculum, teaching, and learning in some ways that are desirable. Tests can help to focus instruction and give students and teachers certain goals to attain. Research shows that changing the content of an important exam can be a powerful force to induce changes in the curriculum. Both of these outcomes can have positive implications for education and student learning.[28]

The flip side of the coin is the possibility that an important examination narrows the curriculum and encourages extraordinary, even exclusive, attention by teachers and students to the material covered on the exam. The amount of instructional and study time given to various topics in the curriculum is likely to be in direct proportion to their appearance on the exam. Therefore, valuable educational objectives and experiences may be omitted from the classroom because there is no easy way to test them properly. Teachers and students can spend an inordinate amount of time on strategies and practices whose only purpose is to improve test performance. Taken to the extreme, test performance can become regarded by students, parents, and teachers as the main, if not the sole purpose of education.[29]

---

[28]W. James Popham, "The Merits of Measurement-Driven Curriculum," Phi Delta Kappan 68 (May 1987):679-682.

[29]Ralph W. Tyler, "The Impact of External Testing Programs," in Warren G. Findley, ed., The Impact and

Another implication of testing that should be considered is called "test score pollution". Thomas Haladyna and Nancy Haas of Arizona State University and Susan Nolen of the University of Washington define "test score pollution" as a breakdown in the validity of a test such that the inferences that one would like to make cannot be made. "Test score pollution" occurs when test scores rise or fall without any change in the underlying construct that the scores are related to. The authors lay out a continuum of test preparation activities running from the ethical to the highly unethical. They believe that ethical practices include training in general test-wiseness, checking answer sheets to see that they are properly filled in, and increasing student motivation to perform on the test through appeals to students, parents, and teachers. Gerald Bracey feels that one might argue with the first of these suggestions in the case of norm-referenced tests, because it is highly unlikely that the norming group received such preparation.[30] Haladyna, Haas, and Nolen consider unethical preparation activities to include developing a curriculum based on the content of a test, preparing and teaching objectives based on test content, presenting similar items to practice before the test, and using any kind of

---

Improvement Of School Testing Programs (Chicago: University of Chicago Press, 1963), pp.193-210.

[30]Gerald W. Bracey, "Testing: Some Cautionary Tales," Phi Delta Kappan 72 (November 1991): 225.

commercial or computer score-boosting packages. Highly unethical behavior would include dismissing the low-achieving students on testing day, presenting items from the test verbatim during preparation, and manipulating the test setting so that students would do poorly initially and look better when they took the posttest at the end of the course. Haladyna, Haas, and Nolen write,

> Until there is serious reform in the way schools prepare students to take standardized achievement tests, test results will continue to misrepresent American public education and its accomplishments. However, as long as test scores remain the single most important index of educational effectiveness, such reform is unlikely to take place.[31]

Ethics: Teaching to the Test

One major concern about standardized testing is that when test scores are used to make important decisions, teachers may be tempted to teach directly to the test. Teaching to the test is not a new concern. With the great emphasis on international comparisons and teacher accountability, it is more likely to happen today than in the past. There is a point at which teaching to the test can be appropriate; but legitimate teaching to the test can cross an ill-defined line and become inappropriate teaching of the test.[32]

All educators are not in agreement as to where the fine line is crossed and which activities are appropriate. One

---

[31]Ibid.

[32]L. A. Shepard and A. E. Kreitzer, "The Texas Teacher Test," Educational Researcher 16 (June 1987):22-31.

way to look at the situation is to describe activities on a continuum. Mehrens and Kaminski suggest the following continuum points:

1. Giving general instruction on district objectives without referring to the objectives that the standardized tests measure

2. Teaching test-taking skills

3. Providing instruction on objectives when objectives may have been determined by looking at the objectives that a variety of standardized tests measure (The objectives taught may or may not contain objectives on teaching test-taking skills.)

4. Providing instruction based on objectives (skills and sub-skills) that specifically match those on the standardized test to be administered

5. Providing instruction on specifically matched objectives (skills and subskills) where the practice or instructions follows the same format as the test questions

6. Providing practice or instruction on a published parallel form of the same test

7. Providing practice or instruction on the test itself.[33]

Mehrens and Kaminski suggest that: Point 1 is always

---

[33]William A. Mehrens and J. Kaminski, "Methods for Improving Standardized Test Scores: Fruitful, Fruitless, or Fraudulent?, Educational Measurement: Issues and Practices 8 (January 1989):16-20.

ethical and Points 6 and 7 are never ethical; Point 2 is typically considered ethical. The point at which you cross the continuum is somewhere between Points 3 and 5 depending on what inference you want to make from the test scores.[34]

Ligon and Jones suggest that an appropriate activity for preparing students for standardized testing is: "one which contributes to students' performing on the test near their true achievement levels, and one which contributes more to their scores than would an equal amount of regular classroom instruction."[35]

Matter offers this thought about preparing students: "Ideally, test preparation activities should not be additional activities imposed upon teachers. Rather they should be incorporated into the regular, ongoing instructional activities whenever possible."[36]

ACT Standards/Suggestions Concerning Practice Ethics

The American College Testing Program appears to have a different view of the continuum with regard to the ethical use of questions from parallel test forms. The American College Testing Program appears not to agree with Mehrens

---

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

[35] G. D. Ligon and P. Jones, "Preparing Students for Standardized Testing: One District's Perspective" (Paper presented at the annual meeting of the American Research Association, April 1, 1982), p. 1.

[36] M. K. Matter, "Legitimate Ways to Prepare Students for Testing: Being Up Front to Protect Your Behind," In J. Hall and F. Wolmut, eds., National Association of Test Directors 1986 Symposia, (Oklahoma City: Oklahoma City Public Schools, 1986), p.10.

and Kaminski's view that it is never ethical to practice on a parallel form of the test. In fact, ACT encourages and promotes familiarity with the content of the test and provides an older form of the test for practice. In the brochure, Preparing for the ACT Assessment, available to every student in the country, is the following statement on page 2.

A Message to Students

> The best indicator of how well you will do in college is a measure of how well you can perform the skills necessary for college coursework. The ACT Assessment--chances are, you and your classmates call it simply "the ACT"--measures these skills in four major curriculum areas: English, mathematics, reading and science reasoning. These areas are tested because they include the major areas of instruction in most high school and college programs.
>
> This booklet, which is provided free of charge, is intended to help you do your best on the ACT. It summarizes general test-taking strategies, describes the content of each of the tests, provides specific tips for each, and lets you know what you can expect on the test day. Included in this booklet are a practice test--a "retired" form of the ACT Assessment that was administered to students on a national test date--and a sample answer sheet and scoring instructions.
>
> Read this booklet carefully and take the practice test well before the test day so you will be familiar with the ACT, what it measures, and the strategies you can use to do your best on it.[37]

See Appendix A for a copy of the entire page two from *Preparing for the ACT Assessment*.

History of Scholastic Aptitude Test (SAT) and American College Testing (ACT)

---

[37]American College Testing Program, "Preparing for the ACT Assessment,"(Iowa City: Iowa, 1991),p. 2.

The two tests used to determine college admissions are the American College Test, known as the ACT, and the Scholastic Aptitude Test, commonly called the SAT. The origin of the two major national tests each came about at a different time and for a different reason. The SAT has been used for almost seventy years, while the ACT has been in existence for about half that time.

The SAT was first administered in 1926 to approximately 8000 students. At that time, each university had its own test and the college admission process was cumbersome for a student who was applying to more than one university. The main purpose of the SAT was to simplify the college admissions process for both the student and the college. This goal remains unchanged today. The word "aptitude" suggests innate ability (prior achievement level) in an area rather than knowledge and skills that can be obtained from in-and out-of-school experiences. The original test measured two areas--verbal skills and mathematics and was designed to predict whether a student had the ability to be successful in college. Today we know that both mathematical concepts and verbal skills can be learned and that reasoning and problem-solving skills can be developed. The SAT organization acknowledges that the word "aptitude" is not technically correct, therefore, the name of the test is being changed to SAT-I. The name is not the only change being made by SAT. The format of the test is being reviewed in light of the recommendations made by the National Council of Teachers of Mathematics (NCTM) Curriculum and Evaluation Standards. Over the last three years, the College Board and the Educational Testing Service (ETS) have been involved in a major research effort to investigate the following possible changes to the current SAT: to make the test more closely related to the current mathematics curriculum; to begin to move away from an exclusively multiple-choice format; to increase the usefulness of scores derived from the test; and to reduce the impact of speed on students' performance.[38]

The American College Testing (ACT) program is more recent than the SAT having been in existence for a little over thirty years. The test was introduced by E. F. Lindquist. Lindquist believed that a college-entrance exam should measure, as closely as possible, the student's ability to do the kinds of tasks that would be required in college and beyond. The ACT was to be a measure of achievement of knowledge and skills in the areas of English, mathematics, social studies, and natural sciences. Lindquist believed that the area of mathematics, like each of the other subjects, should focus on the outcomes of secondary education that are necessary for successful performance in college classes. His belief was that the ACT should be an achievement test designed to measure developed or acquired skills and should consist of tasks that corresponded to

[38]James S. Braswell, "Changes in the SAT in 1994", The Mathematics Teacher 85:(1) (January 1992):16-21.

recognized high school learning experiences. Like the SAT, the ACT has undergone changes in the last few years. In 1989, the College Board introduced the enhanced ACT test. Changes were made to include the social sciences as part of a reading section and to stratify the mathematics portion score. In mathematics, the student is tested over the six areas of: pre-algebra, elementary algebra, intermediate algebra, coordinate geometry, plane geometry, and trigonometry. A sub-score is provided to the student for each area. The ACT requires that students work with the six subject areas over three skill level of thinking: basic skills, application, and analysis.[39] Both testing programs have considered the use and implications for use of calculators on future tests.

Admission to College

The use of standardized tests as a criterion for admission to college is believed to be discriminatory to minorities and the disadvantaged. Both the ACT and the SAT have been cited for questions that give an unfair advantage to the white, middle class. Coaching minority students has met with limited success. Test preparation clinics run by the NAACP found that the clinic helped to improve scores of the blacks but not minorities in general.[40] Samuel Jordan

---

[39]A. Candace Noble and Kenneth B. Mullen, "The ACT Assessment in the 1990's," 85:(1) (January 1992):22-25.

[40]Beverly Cole, "College Admissions and Coaching", Negro Educational Review 38 (April-July 1987):125-35.

reported in a conference presentation that 80% of black colleges that responded to a questionnaire required standardized test scores for entering freshman but most schools in this group (92%) used those scores for placement purposes only.[41] One black college found that a change was affected by the high failure rate of its students when students were given test preparation help through a faculty support system, use of correlated practice items, and an advanced testing seminar.[42] Cuyahoga Community College, Ohio, in conjunction with Cleveland Public Schools, and Links, Inc., a national organization of black women dedicated to civic and educational activities, implemented a program to help improve test-taking skills with inner city students in Cleveland. No significant improvement in test scores was demonstrated but the individual attention given to students did allow for personal improvement plans.[43]

A 1982 study, done by Harold Urman, investigated the effects of test-wiseness skills in ethnically-diverse groups of elementary students. The purpose was to improve achievement in verbal and mathematics skills and to determine if a child's race was a significant factor. Third

---

[41]Samuel Jordan, Jr., Assessment of Standardized Tests Scores and The Black College Environment. Paper presented at Southern Conference on Afro-American Studies. (Jackson, Mississippi, March 27-28, 1987).

[42]Ibid., p.5.

[43]Major L. Harris and Rae Rohfeld, "SAT/ACT Preparation Program: A Team Approach." National Council on Community Surveys and Continuing Education (1983).

and fifth grade students made up a cross-section of White, Black and Hispanics were given test-wiseness training. A pre and post test using the Stanford Achievement Test showed significant increases in the scores for both verbal and mathematics achievement. However, there was no significance found between races. This study showed that teachers should incorporate test-wiseness training into daily classroom activities.[44]

Criticism of Testing

Testing frequently comes under fire for not really being a fair picture of the true abilities of students and for not testing what is relevant to what is taught in schools. Some research conducted has shown that there may be some validity to the criticism. Slack and Porter charged that test scores on the most widely used test, the Scholastic Aptitude Test, could be substantially influenced by coaching programs and that the test was not very good at predicting college grades.[45] Jencks and Crouse concluded that tests such as the Scholastic Aptitude Test did not measure what we normally call "aptitude" any better than do conventional achievement tests. SAT scores seemed to be as dependent as

---

[44]Harold Neal Urman, "Ethnic Differences and the Effects of Testwiseness Training on Verbal and Math Achievement." Ph. D. dissertation, University of Southern California, 1982.

[45]W. V. Slack and D. Porter, "The Scholastic Aptitude Test: A Critical Appraisal," Harvard Educational Review 50 (1980): 154-155.

conventional achievement tests are on the home environment of the student and the quality of the school. Jencks and Crouse confirmed the findings of Slack and Porter in that SAT scores were no better than achievement scores in predicting success in college of adult life.[46] More recently, Crouse and Trusheim of the University of Delaware reported in May 1991 that high school grades and class rank consistently correlate better than SAT scores with freshman grades and with college graduation rates.[47]

Coaching Effects

Since it does not seem feasible that testing will soon fade from the horizon, school administrators and teachers are interested in having their schools show acceptable scores. One indicator that the scope of concern for better scores is widespread, came from a survey which estimated that one-third of the private and public schools in the Northwest United States offered some sort of SAT preparation course.[48]

Can special preparation (especially over a relatively short period of time) have a significant impact on test scores, beyond the effects of regular schooling and/or

---

[46]C. Jencks and J. Crouse, "Should We Relabel the SAT...Or Replace It?", New Directions for Testing and Measurement: Measurement, Guidance, and Program Improvement, NO. 13 (San Francisco: Jossey-Bass, 1982), pp. 132-136.

[47]James Crouse and Dale Trusheim, "Much Ado About The SAT," Phi Delta Kappan 72 (October 1991):254.

[48]D. L. Alderman and D. E. Powers, "The Effects of Special Preparation on SAT Verbal Scores," American Educational Research Journal 17 (1980):239-253.

simply retaking the exam? This issue has been studied and debated for many years and it has importance for several reasons. Firstly, some test takers would have an unfair advantage over others if extra preparation is effective and not readily available to all. Secondly, if test preparation is not effective, then the time and money put into test preparation might be better spent in worthwhile academic pursuits. Thirdly, the question could be raised as to whether the test is a true indicator of general academic ability if short-term preparation that is geared mainly to test taking skills is effective in increasing scores.

Many early studies had been conducted analyzing the effects of coaching on the Scholastic Aptitude Test and other standardized exams without reaching any consensus. One early report summarizing the studies of a number of British experts was done by Vernon. Vernon reported that the average effect of coaching and practice was to increase aptitude scores by nearly .6 standard deviations, equivalent to nine (9) points on an IQ scale. He pointed out that such an effect could be achieved in a remarkably short time, usually between three and nine hours. Therefore, Vernon recommended that all students be coached.[49]

Later reviews, on the other hand, state that the SAT and similar tests are largely resistant to the effects of drill and practice. The trustees of the College Entrance

---

[49]P. E. Vernon, "Practice and Coaching Effects in Intelligence Tests," Educational Forum 18 (1954):269-280.

Examination Board (CEEB) summarized the results of seven of these studies conducted prior to 1965 in the following way:

> Despite variable factors from one study to another the net result across all studies is that score gains directly attributable to coaching amount, on the average, to fewer than 10 points... The magnitude of the gains resulting from coaching vary slightly but they are always small regardless of the coaching method used or the differences in students coached.[50]

Since the average increase that could be expected from coaching was ten points on the SAT score-scale of 200 to 800 points, they (CEEB) viewed coaching programs as a waste of time and money.

In 1978, the CEEB softened its stand somewhat to say that if a student was taking a mathematics course, a review of mathematical concepts would be useful. Their advice about coaching in general, however, had not changed much:

> The verbal and mathematical abilities measured by the SAT were developed over years of study and practice. Drilling or last-minute cramming probably will not do much to prepare you for the test.[51]

Research in the area of coaching has reported mixed gains and conclusions. The subject area that has been coached, whether it was verbal skills or mathematics, provided different results. The majority of the gains were

---

[50]College Entrance Examination Board, Effects of Coaching on Scholastic Aptitude Test Scores, (New York, NY: ERIC Document Reproduction Service No. ED 169, 1968), p 4.

[51]College Entrance Examination Board, Taking the SAT: A Guide to the Scholastic Aptitude Test and the Test of Standard Written English, (New York, NY: ERIC Document Reproduction Service No. ED 203, 1978), p.3.

in mathematical skills. One possible reason for the inconsistencies in conclusions is that different reviewers looked at different parts of the literature on coaching. None of the reviews came close to looking at the entire range of relevant studies. Another possible explanation for the differences is that the reviewers had not all analyzed study results with objective statistical tools. Researchers who use less formal methods of analysis often see what they expect to see in collection results.

A well designed study carried out by Evans and Pike yielded sizable coaching effects. The study was carried out by Educational Testing Service researchers who were familiar with the SAT items pool, and who developed special coaching materials for specific item types included in the pool. The evidence indicated that test preparation could be especially effective on the mathematics portion of the SAT exam. The gains in the verbal section were not as great.[52]

A study done at Harvard University found that coaching only raised scores 10-15 points on the SAT and the conclusion was that commercial test-taking courses were not valuable. However, the same study indicated that short-term coaching did not have a significant effect on verbal scores, but it did significantly raise the mathematics scores.[53]

---

[52]F. R. Evans and L. W. Pike, "The Effects of Instruction for Three Mathematics Item Formats," Journal of Educational Measurement 10 (1973):257-272.

[53]Frederick L. Smyth, "Commercial Coaching and SAT Scores: The Effects on College Preparatory Students in

Most of the studies reviewed by the College Entrance Exam Board (CEEB) and Pike examined the impact of preparation offered by public and private secondary schools. Many students spend money to enroll at commercial coaching schools. It was estimated some 50,000 students spend approximately $10,000,000 annually on commercial coaching for all standardized examinations, not just the SAT and ACT. In 1979, the Federal Trade Commission found reasonably strong gains of at least 25 points for each section of the SAT for students enrolled in the commercial study program offered by The Stanley H. Kaplan Educational Center.[54]

Independent studies done by the National Education Association and the National Academy of Sciences conclude that long-term coaching can have a meaningful impact on scores. Porter reported that an analysis of 31 studies on coaching in the May 1980 Harvard Educational Review found that students gained approximately 30 points per each section of mathematics and science. The longest and most challenging courses that include homework, practice tests and test-taking strategies lead to results that are

---

Private Schools," Journal of College Admissions 123 (Spring 1989):7.

[54]Federal Trade Commission, Bureau of Consumer Protection, Effects of Coaching on Standardized Admission Examinations: Revised Statistical Analyses of Data Gathered by Boston Regional Office of the Federal Trade Commission, (Washington, D. C.: Federal Trade Commission, Bureau of Consumer Protection, NTIS No. PB-296, 1979), p. 196.

statistically significant.[55]

In a Changing Times article, McCormick noted that Stanley Kaplan claimed to have raised the scores for some of the students who had taken his course a total of 250 to 300 points combined out of the 1600 possible on the SAT. Kaplan indicated that typical average gains are more modest, usually in the range of 140 to 150 points overall. The officials of both the ACT and SAT still downplay the role of commercial coaching courses. Fred Moreno, assistant director of public affairs for the College Board in New York City, contended that the score gains from coaching was minimal. Students who retake the SAT without any help will usually score 26 total points higher on the base of 1600. Students who have been coached an average of 60 hours gain on the average of 40 points, according to College Board figures. The College Board does not consider the difference to be statistically significant.[56]

The fact that there are simple test gains from one occasion to another cannot be attributed to the effects of coaching alone. An individual's test scores may vary from one test administration to another due to practice with taking tests, measurement error, and real growth in abilities, irregardless of any intervening test preparation. Although it is difficult to assess the effect of test

---

[55]Ibid.

[56]Kathleen McCormick, "Cramming for College;" Changing Times 41:(9), (September 1987):61.

practice, without the confounding of all the other factors, it appears that simply repeating a test like the SAT increases scores. The College Board did a study in 1991 that included all students who took the SAT as juniors in the spring of 1990 and again as seniors in the fall of the same year. The College Board found an increase of approximately 15 points on the verbal portion and 12 points on the mathematics section based on the SAT range of 200-800 points for each section.[57]

Commercial Coaching

Smyth did a study to assess the effects of commercial coaching on a group of seniors from eight private, nonboarding college-preparatory high schools in suburban Baltimore, MD, and Washington, D.C. Smyth compared the students' PSAT score to the best score the student achieved on any subsequent SAT. Data responses from 438 students were included in the study, 200 of whom had taken some kind of formal training for the SAT. Smyth found that the group who had preparation gained six more points on the verbal portion and thirty-two more points on the mathematics. T tests showed the probability for math increase to be $\rho <$ .00. While the increase in verbal points was insignificant, the coached students did show a significant gain in the

[57]College Board, APT Guide for 1991-92 for High Schools and Colleges, (New York: College Board Publications, 1991).

mathematics.[58]

The study also showed other key factors that seemed to contribute to the gains of students in both groups. Students who had lower PSAT scores produced higher gains in both verbal and mathematics on the SAT. Uncoached students who took the SAT the second or third time showed increases that matched the coached students on the verbal scores (the mathematics gains were better for coached students). Those gains were not matched by students who took the SAT only once.[59]

The gains of the different preparation companies are summarized in the bar graphs of Figure 1 and Figure 2 in Appendix G. None of the four companies (Academic Testing, "Kaplan", Princeton Review, & "Study Works") showed any significant increases in verbal scores between the students who received coaching and those that did not. All companies had significant gains in mathematics. The company that seemed to have the most consistent gains in both mathematics and verbal scores was Kaplan.[60]

Smyth's study found that the mathematics scores in the nonprep students came closer to significance than the verbal scores of either the prep or the non-prep groups. This

---

[58]Fred L. Smyth, "Commercial Coaching and SAT Scores: The Effects on College Preparatory Students in Private Schools," The Journal of College Admissions 123 (Spring 1989):4-5.

[59]Ibid.

[60]Ibid., p. 8

variance suggests that possibly mathematics concepts are more directly related to differences in the curriculum and methods of instruction in various schools than are the verbal.[61] Messick observed that the fact that mathematics has the potential for greater gain should not be surprising given the greater curriculum relatedness of the SAT-M (mathematics) when compared with the SAT-V (verbal).[62]

Coaching Effects Study--University of Michigan

R. Bangert-Drowns, J. Kulik, and C. Kulik, from University of Michigan, did an analysis of over one hundred different coaching studies in 1984. The primary purpose of the project was to evaluate the effectiveness of coaching for aptitude tests; however, Bangert-Drowns, Kulik, and Kulik also did an analysis of coaching on achievement tests and practice techniques in general. The studies investigated how the analyses differed from one another in experimental design and other key features. Categorical variables were created to classify the studies according to those features. Glass's (1981) index of effect sizes was used to transform effects measured on different tests to a common scale. Glass's index gives the number of standard-deviation units

---

[61] Ibid.

[62] S. Messick, "Issues of Effectiveness and Equity in the Coaching Controversy: Implications for Educational and Testing Practice," Educational Psychologist 17 (1982):71.

that separate the group averages that are being compared.[63]

Coaching for Aptitude Tests

Bangert-Drowns, Kulik, and Kulik found in an analysis of 38 studies that there were two major factors that affected the size of coaching gains on aptitude tests. The first factor was the test on which the student was being coached. Coaching programs had different effects on the SAT than they did on other aptitude tests. The second factor that affected study results was the experimental design used for the original study. Some coaching studies used posttest-only designs, whereas others used pretest-posttest designs. In the posttest-only studies, no pretesting was done, and the difference in posttest scores of the coached experimental group and the uncoached control group was used as the effect of coaching. Pretest-posttest studies were more elaborate in design. In these studies, a pretest was given to both groups before the beginning of the coaching program, and then the posttest was given to both groups at the completion of the coaching program. The difference in gains for the experimental and control groups was taken to represent the effect of coaching. The analysis found that pretest-posttest studies sometimes yield larger estimates of effect size than do posttest-only studies because the pretest may sensitize members of the experimental group to the treatment.

---

[63]G. V. Glass, "Primary, Secondary, and Meta-analysis of Research," Educational Researcher 5 (1976):3-8.

The coaching studies for the SAT were analyzed separately from the other aptitude tests. All 14 SAT studies used the pretest-posttest design which yields the maximum estimate of size of an effect. However, the coaching effects were small. Improvement from pretest to posttest averaged .36 standard deviations for the experimental group and .21 standard deviation for the control group. The .15 difference between the two gains is approximately equivalent to 15 points on the SAT scale of 200-800. Although group results were small, some individual student gains were impressive.

There were 24 total studies done on aptitude tests other than the SAT. In the 17 studies that used a pretest-posttest design, improvement in the experimental group averaged .76 standard deviations and .25 standard deviations for the control group. The difference of .51 standard deviations is the effect attributable to coaching alone. The seven studies that did not use a pretest yielded a significantly lower estimate of the size of coaching effects. The difference between the coached and uncoached groups in these studies was an average of .27 standard deviations. The overall average estimated effect of coaching on aptitude tests other than the SAT was .43 standard deviations or the equivalent of approximately six points gain in Intelligence Quotient (IQ).[64]

---

[64] J. A. Kulik, R. L. Bangert-Drowns, and C-L. C. Kulik, "Effectiveness of Coaching for Aptitude Tests,"

Coaching for Achievement Tests

Thirty coaching programs for achievement tests using both the pretest-posttest design and the post-test only design were evaluated by Bangert-Drowns, Kulik, and Kulik. The average effect estimated from pretest-posttest studies based on the difference of gains of the experimental and control groups was .32 standard deviations. The average effect estimated from the post-test only studies was .18 standard deviations. On the basis of all 30 studies, the average effect of a coaching program was to raise performance by .25 standard deviations of approximately two or three months on a grade-equivalent scale. The only factor that seemed to influence the amount of gain on the achievement test was the length of the coaching program. There was a direct relationship with the longer program providing larger gains.[65]

In the most recent comprehensive review, Becker analyzed a total of 48 studies taken from earlier meta-analyses. Becker surveyed all pretest-posttest studies including the ones that did not have a comparison group by the use of alternative measures of the effects of coaching. Becker considered several factors simultaneously, and asked about the relative contribution to the coaching effect

---

Psychological Bulletin 95(2) (1984):182-185.

[65]R. L. Bangert-Drowns, J. A. Kulik, and C-L. C. Kulik, "Effects of Coaching Programs on Achievement Test Performance," Review of Educational Research 53(4) (1983):580-585.

estimates of student characteristics, coaching interventions, and study design. She also investigated whether or not coaching effects were different for the verbal and mathematics sections of the SAT. Becker documented the effect on SAT scores from the duration, the kind of the coaching and the study design. Becker concluded that if the comparison group studies can be taken as the most rigorous evaluations of the effects of coaching, then "we must expect only modest gains from <u>any</u> coaching intervention", on the average of about nine points on the SAT-V and nineteen points for the SAT-M.[66]

Review and Practice Tests

Whether or not "practice makes perfect" depends on several factors. Practice questions and sample tests are part of most every coaching situation. Research on learning over the last century shows that the most effective practice depends upon the time interval between repetitions, the frequency of the repetitions, and even the form of the repetition, that is, whether the practice is a review or a test. An understanding of the research findings may provide an insight as to how to structure review to obtain maximum benefits.

The research indicates that two or more opportunities to study the same material using the same amount of time are

[66]B. J. Becker, "Coaching for the Scholastic Aptitude Test: Further Synthesis and Appraisal," Review of Educational Research 60 (1990):405.

more effective than a single opportunity. In 1917 Edwards conducted a study with elementary school children. Edwards had one group of students study a history or an arithmetic lesson continuously for six and one-half minutes and another group study the same lesson for four minutes on one occasion and two and one-half minutes several days later. The group that had two opportunities to study the lesson performed about 30 percent better on the achievement measure.[67] Dempster found that the reviews that are spread out or distributed over lengthier periods of time produce results that are twice as effective as two massed presentations of the material and that advantage tends to increase as the frequency of the review increases.[68]

Spaced tests are more effective than massed tests, especially if the spaced tests are cumulative in nature. Frequently spaced testing results in higher levels of achievement than infrequent testing.[69] Spitzer did a study that involved 3605 sixth grade students from 91 elementary schools in Iowa. Students were given articles to read and were tested at different time intervals to determine how

---

[67]A. S. Edwards, "The Distribution of Time in Learning Small Amounts of Material," Studies in Psychology: Titchener Commemorative Volume (Worcester, Mass.: Wilson, 1917):209-213.

[68]F. N. Dempster, "The Spacing Effect: A Case Study in the Failure to Apply the Results of Psychological Research," American Psychologist 43 (1988):627-634.

[69]T. Landauer and R. Bjork, "Optimum Rehearsal Patterns and Name Learning," Practical Aspects of Memory (New York: Academic Press, 1978), pp.625-632.

much they had retained. If the test was given immediately after the reading, Spitzer found the students were able to retain material for much longer periods. He discovered that student who were not tested immediately forgot more in one day than students who had the benefit of immediate testing forgot in 63 days. When the material to be learned is first tested relatively soon after its introduction, Spitzer found that tests can actually be used to increase learning .[70]

Another benefit from spaced practice seems to be a deeper understanding of the concepts behind the learning. Research has not provided the answers to how or why this process occurs but students who are engaged in spaced practices or reviews have a richer, more elaborate understanding of the topic. It seems that spaced repetitions require students to engage in active, conscious processing, whereas a massed repetition or a single presentation tends to evoke a shallow, effortless processing.[71]

The Effects of Practice

Bangert-Drowns, Kulik, and Kulik investigated forty studies that looked at the concept of practice or retesting as an effect on learning. They found that the size of

---

[70]J. F. Spitzer, "Studies in Retention," Journal of Educational Psychology 31 (1939):646.

[71]M. A. McDaniel and M. E. J. Masson, "Altering Memory Representations Through Retrieval," Journal of Experimental Psychology: Learning, Memory, and Cognition 11 (1985):371-385.

practice effects turned out to be a function of three factors: similarity of practice and criterion tests, number of practice tests taken, and the innate ability of the student. The first of these, the similarity of the practice and criterion tests, indicated that when parallel forms of test were used the effect of one practice trail raised the criterion scores by .23 standard deviations. When practice and criterion tests were identical, the effects of one practice trial was to raise criterion scores by .42 standard deviations. In more familiar terms, the standard deviation on most IQ tests is 15 points and the standard deviation on most achievement measures is approximately ten months on a grade-equivalent scale. The gain from one practice trial on a parallel test is therefore, approximately three IQ points or two months in grade-level achievement; the gain from practice on an a test identical to the criterion test is approximately six IQ points or four months in grade-equivalent achievement.

The number of practice tests taken was the second factor to influence the gains. Effects increased directly--for both parallel and identical forms of the test with the greater amount of practice showing larger gains. For example, on parallel forms of an IQ test, the gain measured after one practice test would be approximately three IQ points, after four practice tests the gain was approximately seven IQ points, and a gain of approximately eleven IQ points after seven practice tests.

A third factor that affected the results of practice was the ability level of the students tested. Higher ability students gained more from practice than did other students. The relationship between ability and gains from practice was especially clear when the practice was on a test identical to the criterion, but it was also noticeable when practice was given on parallel forms of a test. Higher ability students are apparently able to grasp the lesson from a simple practice test more easily than are the lower ability students who may need more detailed explicit coaching.[72]

Test-Taking Skills

Teaching test-taking skills seem to have a positive effect at all grade levels. Studies have been done at the elementary and secondary levels with significant differences between the experimental and control groups. Fifth graders were given instruction in the Improving Test-Taking Skills (ITTS) program, and pre and post tested on the Iowa Test of Basic Skills (ITBS). There was a significant difference between the mean scores of the group that was taught the testing skills. Students in the experimental group scored higher on the Visual, Concepts, Problems and Total

---

[72]J. A. Kulik, C-L. C. Kulik, and R. L. Bangert-Drowns, "Effects of Practice on Aptitude and Achievement Test Scores," American Educational Research Journal 19:No. 3(Fall 1982):415-429.

Mathematics ITBS subtests.[73]

Third graders in Alabama were instructed using the Soaring High with Test-Taking Tactics-Mathematics Program and then tested on the Stanford Achievement Test. Significant differences were found when the group was considered heterogeneously and when the students were divided into three ability groups.[74] Students in Arizona in the third and sixth grades were given test-wiseness instruction using "Scoring High on the Cat", produced by Random House. Students were tested on the CAT, Form C, and a significant difference was found in mathematics but not reading. It was concluded that all students at all levels from kindergarten through college can learn from test-taking strategies.[75]

At the secondary level, many of the studies have involved the improvement of reading and verbal achievement scores. One study was conducted to see if secondary students whose reading abilities were below grade level could improve their

---

[73]Barbara Louise Benson-Pfiefle, "Effects on Achievement Test Scores Resulting From Teaching Test-Taking Skills in the Fifth Grade." Ed. D. dissertation, Loma Linda University, (1987).

[74]Martha Jean Hitt-Livingston, "The Effects of Testwiseness Instruction Using the Soaring High with Test-Taking Tactics--Mathematics Program on the Mathematics Scores of Third -Grade Students on the Stanford Achievement Test-Seventh Edition. Ed. D. The University of Alabama, (1987).

[75]Judith Anne Bishop, "The Effects of Instruction in TestWiseness on Score Improvement on the California Achievement Test Among Third and Sixth Grade Students." Ed. D. dissertation, Northern Arizona University, 1984.

reading comprehension and vocabulary test scores with the help of test-taking skills. Students were given twelve hours of test-taking skill instruction using published materials from World Book Company. The students were pre-tested with the Nelson-Denny Reading Test, Form E and post-tested with the Nelson-Denny Reading Test, Form F. The study showed a significant gain in mean scores of the experimental group. Test-taking strategies seem to have a positive effect on student achievement. Teachers should be aware that content knowledge alone does not produce a score that representative of student ability.[76]

Eleventh and twelfth graders were given 12-week Advanced Reading Course to determine if it would make a difference their the verbal scores on the SAT, the PSAT test, and the Nelson-Denny Reading Test. This study done in Georgia, included 50 students each in the experimental and control groups. The course consisted of vocabulary development, test-taking and critical thinking techniques, study skills, and outside reading. The tests showed a significant increase on all three tests. Students in the eleventh grade were able to raise their P-SAT scores by 61.6 and the twelfth grade students increased their SAT scores by 66 points. These were significant gains at the ($\rho \leq 0.01$)

---

[76]Diane Vida, "A Study of Test-Taking Skills and Achievement Scores Upon Secondary Students." Ed. D. dissertation, Drake University, (1985).

level of significance over the control group.[77]

At the college level, two sections of a biology class taught by the same instructor, at the Agricultural and Technical College, Morrisville, New York, participated in a study to determine if teaching test-taking strategies would have a measurable effect on test performance. During the first semester, twelve tests were administered to each class, which contained 20 students each, and the students' midterm grades were used to determine the equivalency of the two classes. The mean scores for the two classes were almost equal and the student t test of variance revealed no significant differences between the two classes. During the second semester, one class was taught a 15-minute lecture on a test-taking strategy at the beginning of each chapter. During this lecture, students were given suggestions about how to control test anxieties; how to list information which may be needed on the back of the exam before making an attempt to answer questions; how to answer all of the easy or known questions before tackling the harder ones; how to seek clues from the answers on the exam; how to read directions carefully and proofread the exam; and how to utilize all available time. At the end of the study, the experimental group had gained almost an eleven point

---

[77]Kathleen Brown Burke, "A Model Reading Course and Its Effects on the Verbal Scores of the Eleventh and Twelfth Grade Students on the Nelson-Denny Test, the Preliminary Scholastic Aptitude Test, and the Scholastic Aptitude Test (Coaching, Study Skills, Vocabulary)." Ph. D. dissertation, Georgia State University, 1986.

differential in the mean score. The resulting student t was significant at the .01 alpha level indicating that 99 percent of the difference in means could be attributed to the instruction in test-taking skills.[78]

Testing-taking strategies studies have shown improvement in the achievement levels of special education students in mainstreamed situations. Twenty-eight sixth-grade students who received instruction in Scorer, a test taking strategy, did significantly better than the control group on pre and post test reading scores.[79] Twelve middle school emotionally handicapped and learning disabled students were given test-taking strategies. Test scores in the students mainstream classes were used to determine growth. Results indicated that all but one student had increased their mean scores. Test-taking skills seem to provide students with an advantage to help them be successful.[80]

A study of instructional techniques that could potentially improve student scores of the ACT or SAT was conducted at a private computer camp in Illinois. The

---

[78]Joseph W. Culhane, "Should Test-Taking Strategies Be Taught?," The Clearing House, November 1982, pp.101-102.

[79]Shirley Ann Ritter, "Teaching Middle School Students To Use a Test Taking Strategy (Mainstreaming, Learning Strategies, Generalization)." Ph. D. dissertation, University of Illinois at Urbana-Champaign, 1985.

[80]Charles Allan Hughes, "A Test-Taking Strategy for Emotionally Handicapped and Learning Disabled Adolescents." Ph. D. dissertation, University of Florida, 1985.

control group was a semester course of regular college English taught to high school students that used ACT test discussion questions. Students in the experimental group were given a pretest and four weeks of treatment. Private interviews were held with students to provide motivation and to evaluate progress and attitudes toward testing. The researcher found no significant difference between the groups but the experimental group seemed to benefit from the intervention of computer coaching, teacher coaching, and cooperative learning by an increase in scores on sample ACT items.[81]

Preparation activities used to improve student performance on the ACT fall into three major categories: teaching test-taking strategies that are not directly related to the scope and content of the test; memorization or cram courses that rely on rote memory and teaching techniques of intelligent estimating; and academic classes which emphasize cognitive skills and involve a review of instruction of content knowledge and skills measured by the tests. Any or all of these categories can be incorporated into review classes.

Chapter Summary

At all levels, national, state, and local, there has been an outcry for measures of accountability of educational

---

[81]Claire Gunn Weaver, "A Study of Instructional Techniques to Prepare for the ACT Test." Ph. D. dissertation, Southern Illinois University at Carbondale, 1988.

performance. Since the service began, the College Board Testing Program has been named as one of those standards that can be easily and reliably used as an indicator of educational competence. College (SAT/ACT) testing scores were included in the first national "wall chart" for comparison a state's educational programs. Many states like California use the college testing scores in the state report card.

In the State of Michigan, all administrators must comply with Public Act 25. The authors of Public Act 25 sought to increase performance and the quality of schools through the legislation that requires administrators to communicate about the conditions of the schools through the Annual Report. ACT and SAT scores must be included in the Student Achievement section of the Annual Report. The ACT and the SAT scores have been consistently used as an indicator of quality education, as a standard for comparison among schools, and as a major yardstick for accountability.

In general, the analysis of over sixty different studies found varying results for the effects of coaching. Coaching produced the following average gains:

*On the SAT, a gain of about 15 points of .15 standard deviations on the scale of 200 to 800 points.

*On aptitude tests other than the SAT, the gain was equivalent to approximately six points or .43 standard deviations.

*On achievement tests, the gain was approximately two to three months on a grade-equivalent scale or .25 standard deviations.

The analysis of forty studies on the effects of practice found that simple practice did produce strong results. A single practice trial on a parallel (not identical) test produced a gain of three Intelligence Quotient (IQ) points or two to three months on a grade equivalent scale.

A survey of the studies showed that regardless of the type of testing, students made greater gains when:

*A pretest was given prior to coaching

*Practice was given on tests identical to the criterion tests, and

*Practice was given on a regular schedule over a longer period of time.

The overall gain from any coaching activity is better among high-ability students. The conjecture is that more-able students seem to be able to learn a concept directly from an example while the lower ability student may need more explicit instruction.

The frequent use of properly spaced reviews and tests in the classroom can dramatically improve classroom learning and retention. The use of cumulative questions on properly spaced testing is one of the keys to effective learning. In addition, research indicates that spaced repetitions can foster time-on-task and help students develop and maintain positive attitudes toward learning and school.

The fact that achievement tests and aptitude tests seem to be equally susceptible to the influence of special preparation programs may be important to those who have argued that school admissions decisions should be based on achievement tests rather than aptitude tests. Achievement tests can no longer be considered impervious to effects from special preparation programs.

CHAPTER III

DESIGN OF THE STUDY

Overview

Chapter Three will address the issues of research procedures, sample population, the materials used, the data collection process, and the limitations of the study.

This pretest/posttest study was designed to measure the effects of using practice mathematics items and test preparation activities on the ACT scores of students enrolled in a second-year algebra class. While the ACT test contains four sub-section tests, the researcher was concerned with and limited the study to the mathematics section only. The study was also designed to determine whether a student's gender or the number of years of mathematics study is related to performance on the mathematics section of a standardized test like the ACT.

Purpose of the Study

The primary purpose of this study was to determine whether student scores on the mathematics portion of the American College Testing Program (ACT) test could be improved by practicing sample test items and studying general test-taking strategies.

There were two secondary purposes included in this study. One was to determine whether a student's gender has any significance in ACT performance and the other was to determine whether the number of years of mathematics study affected achievement.

This study was designed to measure the effects of spaced practice of sample ACT mathematics questions and test-preparation suggestions on the mathematics section score of the ACT of students enrolled in a second-year algebra class.

Research Questions

This study of coaching for a standardized achievement test like the American College Testing Program (ACT) using practice test items and test-taking strategies was designed to answer the following questions:

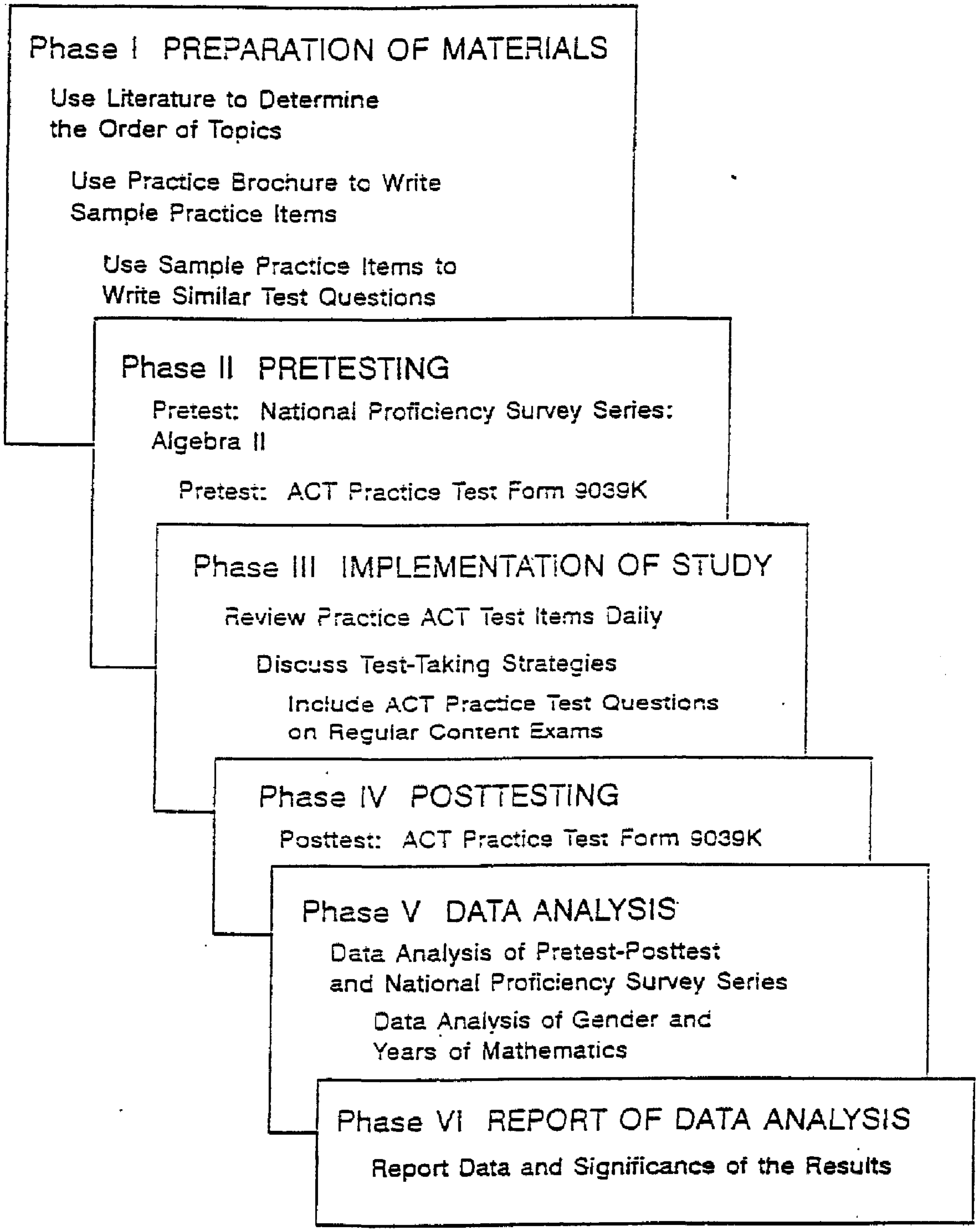
*Research Question One: Will the use of practice test items similar to the mathematics questions given on the American College Testing (ACT) exam and test-taking suggestions significantly improve the ACT mathematics score of second-year algebra students?

*Research Question Two: Does gender have any significance in the performance of second-year algebra students on the mathematics section of the ACT?

*Research Question Three: Does the number of years of prior mathematics study affect achievement on the mathematics section of the ACT?

*Research Question Four: Does the level of the student's prior achievement have any effect on the student's performance?

Figure 2. Research model used to study coaching effects on the mathematics section of American College Testing (ACT).



To accomplish these objectives, a comparison of ACT scores on a Practice Mathematics Exam prepared by American College Testing Program was made among eight sections of second-year algebra students. Students in four of the sections were taught test-taking strategies and review practice test items, while students in the four control groups were taught second-year algebra without the use of ACT practice materials.

In addition to the ACT pretest, all students took the National Proficiency Survey Series Algebra II exam, published by The Riverside Publishing Company, Chicago, Illinois, at the beginning of the semester. The duration of the study was 10 weeks. The textbook for each of the eight sections was Algebra Two and Trigonometry, by Alan Foster, James Rath, and Leslie Winters, Merrill Publishing Company, Columbus, Ohio, 1990. The test-taking strategies and practice ACT items came from materials suggested by the ACT or were written by the researcher. Each set of five ACT practice questions were reviewed during the week along with the regular Algebra II curriculum. The set of five correlating test questions were included in the next exam along with the regular Algebra test questions.

Sample items for practice and correlated test items are listed in the Appendix B and C. Test taking strategies are included in Appendix D. Specific mathematics testing strategies are included in Appendix E. Commonly used rules and formulas are included in Appendix F.

Students in all eight sections took a sample ACT test Form #9039K during the first week of the semester. Four classes were taught second-year algebra along with the test-taking skills and practice ACT items in mathematics. The four classes in the control group were taught second-year algebra only. At the end of the ten weeks, all students were tested again with the same sample ACT test Form #9030K and scores were compared.

The study was comprised of ten weeks of test-taking skills and practice ACT items. After the post-test, the control groups were given the same ACT practice information used with the experimental groups during the second ten weeks of the semester. Since the ACT has become such an important gatekeeper to all college-bound students, it would not be politically wise nor sound educational practice to deny half or our students equal opportunity to prepare for the ACT.

Sources of Data

In order to accomplish the objectives of this study, a comparison of the use of test-taking skills with content review and practice ACT questions was made among eight sections of second-year algebra students at a suburban high school in southeastern Michigan. This predominately blue-collar community had one high school, one middle school, and six elementary schools with a total enrollment of 4866 students. The ethnic/racial statistics show that 96.57% of the students are white and 3.43% are minorities. The high school enrollment is 1410 students. Second-year algebra students were chosen because that is usually the third year in the sequence of college preparatory classes. The ACT test is traditionally taken by most students at the end of their junior year so that scores will be available for them to apply for college admission at the beginning of their senior year.

<u>Sample</u>

The students involved in this study were tenth or eleventh grade students, who were between fifteen and seventeen years old. Students in the four experimental groups were taught test-taking skills and ACT practice test items along with the regular second-year algebra curriculum. The other four classes were taught second-year algebra without the use of the test-taking and item review materials. Each section of students had an approximate enrollment of thirty. The study consisted of 196 students, with 99 and 97 students respectively in the experimental and control groups. The students who selected the class were randomly assigned by the computer to the various sections. Second-year algebra is taught at two different levels, Honors Algebra and Intermediate/Advanced Algebra. There were two sections of Honors Algebra and six sections of Intermediate/Advanced Algebra. Within the six sections of Intermediate/Advanced Algebra, there was a group of tenth grade students with two years of mathematics that had taken Algebra I and Geometry in the eighth and ninth grades respectively. To balance the study, one section of Honors Algebra was in the experimental group and one in the control group. Each group of Honors Algebra had a different instructor. Three sections of Intermediate/Advanced Algebra were experimental and three were control groups. The control and experimental groups of Intermediate/Advanced Algebra each had a different instructor. See Table I and the accompanying description for precise breakdown of the sample population (Chapter IV, page 77).

Delmitations

There were four major limitations to this study. The first limitation was that while there are two major tests used for college admission purposes: The Scholastic Aptitude Test (SAT) and the American College Testing Program (ACT), this research study was limited to the ACT test. The reason the ACT was chosen instead of the SAT was that the colleges in the State of Michigan prefer to use the ACT over the SAT for admission purposes.

The second limitation of this study was that although the American College Testing Program (ACT) test covers the four areas of English, Mathematics, Reading, and Natural Sciences; this study was concerned only with the effect of teaching practice ACT items and test-taking strategies in the subject of mathematics.

The third limitation of this study was that the participants were limited to randomly-placed students in eight sections of second-year algebra in one high school.

The fourth limitation was that all four teachers did not teach all types of students. The constraints of the master schedule for the school did not allow the flexibility to arrange sections differently. Two teachers did not teach any tenth grade students or any eleventh grade students with one year of mathematics.

Chapter Summary

This chapter discussed the sources of data, the sample population, the research procedures, the materials used, and the data analysis employed in this study. The results of the data analysis are presented in Chapter IV.

CHAPTER IV

DATA ANALYSIS

Overview

The preceding chapter has described the design of the study, the testing instruments including practice ACT materials and the methods used to gather data. This chapter will discuss the research techniques used and analyze the significance of the data derived from the pretesting and posttesting.

Research Techniques

The experimental research used in this study was the randomized pretest-posttest design. The three assumptions concerning the subjects in both the experimental and control groups are: randomness, the subjects were assigned by chance by the computer scheduling program to each group; normality, the group contains a distribution of individuals from a general population on the characteristics of interest; and finally, homogeneity, the subjects of two or more samples have been drawn from populations of equal variances. Both groups were given a pretest and the National Proficiency Exam for Algebra II. The experimental group was given treatment and both groups were given a posttest.

The method by which data is statistically analyzed is determined by the nature of the methods used in research. If the study involved the use of a survey or ranked data using ordinal numbers, the researcher would use nonparametric techniques such as the Mann-Whitney U Test to determine significance. Parametric statistical techniques are used to analyze data that is generated by using subjects to test a hypothesis as was the case with this study. These parametric techniques do make the assumptions listed above concerning the population from which the samples involved in the study were drawn. The advantage of using parametric analysis is that they are generally more powerful than the nonparametric techniques and therefore are more likely to reveal a true difference or relationship if one exists.[82]

This study used the hypothesis testing method that involved the use of a null hypothesis and an alternate hypothesis. The null hypothesis is stated such that there is no difference between the population means of the two groups (difference of the two means is zero). Using appropriate testing methods, a value (probability of the occurrence happening by chance) is obtained. If the probability is small, the null hypothesis is rejected, thereby providing support for the research alternate hypothesis. In most educational research, it is customary to view an outcome as unlikely if the probability is less than or equal to five percent ($\rho \leq 0.05$). This is referred to as a 0.05 level of confidence or significance. When we reject the null hypotheses at the 0.05 level of significance, we are saying that the possibility of

---

[82]Jack R. Fraenkel and Norman E. Wallen, How to Design and Evaluate Research in Education (New York: Mc-Graw Hill, 1990), pp. 185-186.

obtaining such an outcome by chance is only five (or less) times in one hundred.[83]

There are three commonly used parametric tests for finding the difference between the means: the t-test, the analysis of variance (ANOVA), and the analysis of covariance (ANCOVA). The t-test is used to determine whether the difference between the means of two samples is significant. If the study involves more than two groups, the analysis of variance test (ANOVA) is used. The ANOVA is a more general form of the t-test that is used to analyze variation both within and between each of the groups. The null hypothesis is rejected when the probability of the occurrence is less than or equal to five percent ($\rho \leq 0.05$). The ANOVA is used in place of multiple t-tests to reduce the probability of making a Type 1 error. A Type 1 error is made when a true null hypothesis is rejected or considered to be false.[84]

This study used the analysis of covariance. The analysis of covariance (ANCOVA) is used when groups are given a pretest related in some way to the dependent variable and their mean scores on the pretest are found to differ. The ANCOVA is a combination of regression techniques (used for prediction) and the ANOVA techniques. The ANCOVA will allow the researcher to adjust the posttest

---

[83]Ibid., p. 181.

[84]Richard P. Runyon and Audrey Haber, Fundamentals of Behavioral Statistics (New York: Mc Graw Hill, 1991), pp. 306-307.

mean scores on the dependent variable for each group to compensate for the initial differences on the pretest. In some cases, researchers can use more than one covariate to make the groups balance. The pretest and the National Proficiency Exam for Algebra II are the independent variables (covariates) used for this research.[85]

There are two major purposes for using the covariate in any study. One purpose is to statistically equate groups that are different. Increasing the power of the statistical analysis is the second purpose for using the ANCOVA. When subjects are randomly assigned to different treatment groups, the inclusion of the covariate (quantitative variable) which is unrelated to the grouping condition (categorical variable) but related to the dependent variable may dramatically increase the power of the statistical testing. There is the expectation that there be a correlation that students with high scores on the pretest would also score high on the posttest. The square of the correlation coefficient (coefficient of determination) shows how much variance in the dependent variable is accounted for by the pretest. Some of the variance in the dependent variable can be accounted for by using the pretest results to predict the posttest results. Since the covariate (pretest & National Proficiency Exam for Algebra II) and the grouping variable (factor) have no relationship to each

---

[85]Robert L. Hale, Mystat Statistical Applications--Dos Version (Cambridge, Massachusetts, 1990), pp.126-127.

other, the variance in the dependent variable accounted for by the covariate will have no relationship to the categorical independent variable.[86]

Error variance is the variance not related to the treatment. Error variance can be decreased in any study if the variance in the dependent variable is accounted for by statistically removing the covariate. The F statistic contains the error variance in the denominator of the fraction, which means that as the denominator of the fraction decreases in error size the value of the overall F ratio increases. The inclusion of a covariate in the analysis will increase the value of the F ratio. However, one degree of freedom is lost for each covariate used so that the sum of the squares accounted for by the covariates must be worth the degrees of freedom lost to the error term. The mean-square value is obtained by dividing the sum-of-squares by the degrees of freedom. The F statistic grows larger as the mean-square error is made smaller and therefore the power of the statistical test is increased. The null hypothesis is rejected when the probability is less than or equal to five percent ($\rho \leq 0.05$) in an ANCOVA.[87]

Quantitative Results

The statistical program used to analyze the data in this study was MYSTAT. MYSTAT was published by Course

---

[86]Ibid., p. 127.

[87]Ibid., p. 129.

Technology, Inc. using SYSTAT, Inc. systems specifically adapted for use in statistical analysis in the educational community. The first four tables contain the descriptive statistics for the total group and each of the subgroups contained within the study.

Table I is a count table that shows the number of students taught by each of the four teachers. The students are categorized by their group, either control or experimental, by their class, either ten or eleven, the number of years of mathematics, either one or two, by the course level, either honors or intermediate, and by teacher. The total number of students in the study was 196.

Tables II, III, and IV give the mean and standard deviation on the on the pretest, posttest, and National Proficiency Exam for Algebra II. Table II shows the honors subgroup that was comprised of eleventh grade students. The experimental group contained 22 males and 10 females for a total of 32 students. The control group contained 20 males and 13 females for a total of 33 students. Table III show the intermediate group of eleventh grade students which contained 10 males and 27 females for a total of 37 students in the experimental group and 19 males and 19 females for a total of 38 students in the control group. Table IV shows the tenth grade students in the intermediate group were comprised of 13 males and 17 females for a total of 30 in the experimental group and 12 males and 14 females for a total of 26 students in the control group. The experimental subgroups contained 99 students in which the total number of males was 45 and the total number of females was 54. In the control subgroups of 97 students, the total number of males was 51 and the total number of females was 46. The total number of students in the study was 196 which was comprised of 96 males and 100 females.

This study used the data gathered from the pretest, the National Proficiency Exam, and the posttest to analyze the results based on six factors. The six factors to be considered for significance were: the group (control or experimental) that the student participated in, the gender (male or female) of the student, the number of years (one or two) of mathematics taken prior to Algebra II, the class (tenth or eleventh) of the student, the level (honors or intermediate) of student ability, and the teacher (1, 2, 3, 4).

Table I

Number of Students Taught
by
Each Faculty Member and Grade and Years of Mathematics

	11th Grade		10th Grade		
	1 Year of Mathematics	2 Years of Mathematics	1 Year of Mathematics	2 Years of Mathematics	
Teacher 1 Honors Control	0	33	0	0	33
Teacher 2 Honors Experimental	0	32	0	0	32
Teacher 3 Intermediate Control	13	25	8	18	64
Teacher 4 Intermediate Experimental	12	25	14	16	67
Totals	25	115	22	34	196

In Table I, the students are categorized by their group, either control or experimental, by their class, either ten or eleven, by the number of years of mathematics, either one or two, and by the course level, either honors or intermediate. The total number of students in the study was 196. The study was comprised of 140 eleventh graders, of which 115 had taken two years of mathematics and 25 who had only one year of mathematics prior to Algebra II. The tenth grade group of 56 total students was comprised of 34 students with two years and 22 students with one year of prior mathematics. Teacher 1 and Teacher 2 each taught one honors section of Algebra II and had 33 and 32 students, respectively. Teachers 1 and 2 taught eleventh grade only. Teacher 3 and Teacher 4 each taught three sections of Intermediate Algebra II and had 64 and 67 students, respectively.

Table II shows the descriptive statistical comparison for eleventh grade honors experimental and control groups with a break-down for gender. The National Proficiency Exam for Algebra II, as well as the pretest, scores were used to equalize the two groups. The mean on the National Proficiency Exam was 18.00 in the experimental group and 19.82 for the control group. In the experimental group, the mean of the students increased from 18.88 to 29.09 on the pretest/posttest (a gain of 54.08%). In the control group, the mean of the students increased from 21.94 to 29.55 on the pretest/posttest (a gain of 34.69%). The mean scores of the females in each group were lower than the males; however, the percentage of increase was greater for the female students in both groups. Female scores averaged an increase of 75.16% in the experimental group and 36.70% in the control group as compared to an increase of 46.47% for the males in the experimental group and 33.33% increase for the control group.

Table II

Descriptive Statistics for Teachers and Specialized Experimental and Control Groups

	Teacher 1 Honors 11th Grade Control	Pretest	Posttest	National		Teacher 2 Honors 11th Grade Experimental	Pretest	Posttest	National
n	Male	20	20	20		Male	22	22	22
Mean		22.20	29.60	20.80			20.14	29.50	18.64
SD		6.90	7.30	3.49			6.44	8.14	3.80
n	Female	13	13	13		Female	10	10	10
Mean		21.54	29.46	18.31			16.10	28.20	16.60
SD		3.37	5.77	3.27			4.30	5.29	2.29
n	Total	33	33	33		Total	32	32	32
Mean		21.94	29.55	19.82			18.88	29.09	18.00
SD		5.78	6.74	3.16			6.16	7.35	3.53

Table III

Descriptive Statistics
for
Teachers and Specialized Experimental and Control Groups

	Teacher 3 Intermediate 11th Grade Control	Pretest	Posttest	National		Teacher 4 Intermediate 11th Grade Experimental	Pretest	Posttest	National
n	Male	19	19	19		Male	10	10	10
Mean		16.79	18.58	11.42			19.20	22.80	15.20
SD		4.42	7.62	3.53			5.71	6.48	4.31
n	Female	19	19	19		Female	27	27	27
Mean		16.53	17.00	12.47			15.93	20.41	13.89
SD		3.28	4.83	4.11			4.26	5.95	2.91
n	Total	38	38	38		Total	37	37	37
Mean		16.66	17.79	11.95			16.81	21.05	14.24
SD		3.90	6.43	3.87			4.91	7.35	3.40

Table III shows the descriptive statistical comparison for the eleventh grade intermediate experimental and control groups with gender break-down. The mean on the National Proficiency Exam was 14.24 for the experimental group and 11.95 for the control group. In the experimental group, the mean of the students increased from 16.81 to 21.05 on the pretest/posttest (a gain of 25.22%). In the control group, the mean of the students increased from 16.66 to 17.79 on the pretest/posttest (a gain of 6.78%). The mean scores of the females in each group were lower than the males; however, the percentage of increase was greater for the females (28.12%) than the males (18.75%) in the experimental group. In the control group which seemed to be more evenly matched, the males increased 10.66% and the females increased 2.8%.

Table IV shows the descriptive statistical comparison for the tenth grade intermediate experimental and control groups with gender break-down. The mean on the National Proficiency Exam for Algebra II was 16.33 for the experimental group and 15.19 for the control group. In the experimental group, the mean of the students increased from 18.57 to 26.07 on the pretest/posttest (a gain of 40.39%). In the control group, the mean of the students increased from 21.42 to 26.08 (a gain of 21.76%). In the experimental group the females had a lower mean score than the males but the females scores increased 41.10% while the male increase was 39.60%. In the control group, the females had a higher mean score than the males but the percentage increase of 27.02% was greater for the males as compared the female increase of 18.54%.

Table IV

Descriptive Statistics
for
Teachers and Specialized Experimental and Control Groups

	Teacher 3 Intermediate 10th Grade Control	Pretest	Posttest	National		Teacher 4 Intermediate 10th Grade Experimental	Pretest	Posttest	National
n	Male	12	12	12		Male	13	13	13
Mean		17.58	22.33	12.08			20.00	27.92	17.62
SD		7.51	6.69	5.48			8.71	9.11	5.62
n	Female	14	14	14		Female	17	17	17
Mean		24.71	29.29	17.86			17.47	24.65	15.35
SD		5.91	7.95	4.41			4.83	6.85	3.97
n	Total	26	26	26		Total	30	30	30
Mean		21.42	26.08	15.19			18.57	26.07	16.33
SD		7.58	8.17	5.71			6.90	8.07	4.89

In all Tables V-XIX, the pretest score will show significance due to the fact that the covariate (pretest) must contribute to a significant level in variance to the posttest (dependent variable). Each table records the pretest at the significant confidence level of $\rho \leq 0.05$.

Since teacher's 1 and 2 taught no students with one year of mathematics or any tenth grade students, there are three ways to run a comparison with all cells being active. The first way that all teachers can be involved is to consider the eleventh grade students with two years of mathematics. (Tables V-X) The second analysis that can be made involved the best students that included the eleventh grade honors students and the tenth grade intermediate students with two years of mathematics. (Tables VII, XI-XV) The third analysis involved the intermediate level students in the eleventh grade with one or two years of mathematics. (Tables VIII, XVI-XIX)

Table V

Pretest/National Posttest Scores Analysis of Covariance for All Eleventh Grade Students With Two Years of Mathematics on the Variables: Level, Group and Gender

n = 115

Source	Sum of Squares	Degrees of Freedom	Mean Squared	F-Ratio	ρ **
Level	**535.883**	**1**	**535.883**	**18.703**	**.000⁺**
Group	**141.866**	**1**	**141.866**	**4.951**	**.028**
Gender	3.761	1	3.761	0.131	.718
Level * Group	.669	1	.669	0.023	.879
Level * Gender	34.592	1	34.592	1.207	.224
Group * Gender	32.480	1	32.480	1.134	.289
Level * Group * Gender	6.928	1	6.928	0.204	.624
Pretest	**1128.801**	**1**	**1128.801**	**39.396**	**.000⁺**
National	52.780	1	52.780	1.842	.178
Error	3008.546	105	28.653		

**** Null Hypothesis Rejected when BOLD** Ho: = Null Hypothesis Ha: = Alternate Hypothesis

Ho1: **There is no Level significant difference in posttest scores when comparing all Eleventh Grade Honors and Intermediate students who have taken two years of mathematics.**

Ha1: There is Level significant difference in posttest scores when comparing all Eleventh Grade Honors and Intermediate students who have taken two years of mathematics.

Ho2: **There is no Group significant difference in posttest scores when comparing all Eleventh Grade**

Control and Experimental students who have taken two years of mathematics.

Ha2: There is Group significant difference in posttest scores when comparing all Eleventh Grade Control and Experimental students who have taken two years of mathematics.

Ho3: There is no Gender significant difference in posttest scores when comparing all Eleventh Grade Male and Female students who have taken two years of mathematics.

Ha3: There is Gender significant difference in posttest scores when comparing all Eleventh Grade Male and Female students who have taken two years of mathematics.

Ho4: There is no Level and Group interactive significant difference in posttest scores when comparing all Eleventh Grade Honors/Intermediate and Control/Experimental students who have taken two years of mathematics.

Ha4: There is Level and Group interactive significant difference in posttest scores when comparing all Eleventh Grade Honors/Intermediate and Control/Experimental students who have taken two years of mathematics.

Ho5: There is no Level and Gender interactive significant difference in posttest scores when comparing all Eleventh Grade Honors/Intermediate and Male/Female students who have taken two years of mathematics.

Ha5: There is Level and Gender interactive significant difference in posttest scores when comparing all Eleventh Grade Honors/Intermediate and Male/Female students who have taken two years of mathematics.

Ho6: There is no Group and Gender interactive significant difference in posttest scores when comparing all Eleventh Grade Control/Experimental and Male/Female students who have taken two years of mathematics.

Ha6: There is Group and Gender interactive significant difference in posttest scores when comparing all

Eleventh Grade Control/Experimental and Male/Female students who have taken two years of mathematics.

Ho7: There is no Level, Group and Gender interactive significant difference in posttest scores when comparing all Eleventh Grade Honor/Intermediate, Control/Experimental and Male/Female students who have taken two years of mathematics.

Ha7: There is Level, Group and Gender interactive significant difference in posttest scores when comparing all Eleventh Grade Honors/Intermediate, Control/Experimental and Male/Female students who have taken two years of mathematics.

Table V shows the analysis of covariance for variables of level, group, and gender for the 115 eleventh grade students with two years of mathematics in the experimental and control groups. When the experimental groups was compared to the control group for all possible variations of the three factors, only two factors produced a statistically significant number at the alpha level $\rho \leq 0.05$. The level (honors or intermediate) showed a significant score of $\rho = 0.000$ and the group (experimental or control) showed a significant score of $\rho = 0.028$. Therefore, the null hypothesis that there is no level difference in posttest scores of all eleventh grade honors and intermediate students must be rejected. Students in the honors section did significantly better than the students in the intermediate sections. Also, the null hypothesis that there is no group significant difference in posttest scores of all eleventh grade honors and intermediate students who have taken two years of mathematics must be rejected. Gender was not significant. None of the interactive analysis showed significance. Eleventh grade students with two years of mathematics in the experimental group did significantly better than the eleventh grade students in the control group.

Table VI analyzed the same group of 115 students as Table V with respect of the variables of teacher and gender. At the alpha level $\rho \leq 0.05$, the teacher variable showed a significant score of $\rho = 0.000^{+}$, so that, the null hypothesis that there is no teacher significant difference in posttest scores when comparing all teachers of eleventh grade students who have taken two years of mathematics must be rejected. Neither gender nor any of the interactive analysis was significant.

The significance found in Tables V and VI indicate that the level, the group, and the teacher produced a significant score. Level (honors/intermediate) and group (experimental/control) are self explanatory in nature. The teacher significance was analyzed further to determine where the difference between the four teachers occurred (Tables VII-X).

Table VI

Pretest/National Posttest Scores Analysis of Covariance
for
All Eleventh Grade Students With Two Years of Mathematics
on the
Variables: Teacher and Gender
n = 115

Source	Sum of Squares	Degrees of Freedom	Mean Squared	F-Ratio	ρ **
Teacher	**689.948**	**3**	**229.983**	**8.027**	**.000⁺**
Gender	3.761	1	3.761	0.131	.718
Teacher * Gender	75.487	3	25.162	0.878	.455
Pretest	**1128.801**	**1**	**1128.801**	**39.396**	**.000⁺**
National	52.788	1	52.788	1.842	.178
Error	3008.546	105	28.653		

** **Null Hypothesis Rejected when BOLD** Ho: = Null Hypothesis Ha: = Alternate Hypothesis

Ho1: There is no Teacher significant difference in posttest scores when comparing all Teachers of Eleventh Grade students who have taken two years of mathematics.

Ha1: There is Teacher significant difference in posttest scores when comparing all Teachers of Eleventh Grade students who have taken two years of mathematics.

Ho2: There is no Gender significant difference in posttest scores when comparing all Eleventh Grade Male and Females students who have taken two years of mathematics.

Ha2: There is Gender significant difference in posttest scores when comparing all Eleventh Grade Male and Female students who have taken two years of mathematics.

Ho3: There is no Teacher and Gender interactive significant difference in posttest scores when comparing Teachers of all Eleventh Grade Male/Female students who have taken two years of mathematics.

Ha3: There is Teacher and Gender interactive significant difference in posttest scores when comparing Teachers of all Eleventh Grade Male/Female students who have taken two years of mathematics.

Table VII shows the analysis of covariance for all eleventh grade honor students with two years of mathematics with regard to the variables of teacher/group and gender. This comparison between teacher 1 (honors/control) and teacher 2 (honors/experimental) included 65 students. The analysis showed a significant ρ score of $\rho = 0.040$. At the alpha level of $\rho \leq 0.05$, the null hypothesis that there is no teacher significant difference in the posttest scores when comparing eleventh grade honors experimental group and the eleventh grade honors control group must be rejected. Neither gender nor teacher/group * gender was significant. Eleventh grade honors students with two years of mathematics in the experimental group did significantly better than their counterparts in the control group.

Table VII

**Pretest National Posttest Scores Analysis of Covariance
for
Eleventh Grade Honor Students With Two Years of Mathematics
on the
Variables: Teacher/Group and Gender**

n = 65

Source	Sum of Squares	Degrees of Freedom	Mean Squared	F-Ratio	ρ **
Teacher/Group	**108.508**	**1**	**108.508**	**4.416**	**.040**
Gender	49.970	1	49.970	2.034	.159
Teacher/Group * Gender	4.285	1	4.285	0.174	.678
Pretest	**702.004**	**1**	**702.004**	**28.573**	**.000+**
National	79.081	1	79.081	3.219	.078
Error	1449.580	59	24.569		

** **Null Hypothesis Rejected when BOLD** Ho: = Null Hypothesis Ha: = Alternate Hypothesis

Ho1: There is no Teacher/Group significant difference in posttest scores when comparing Teachers all Eleventh Grade Experimental and Control Honor students who have taken two years of mathematics.

Ha1: There is Teacher/Group significant difference in posttest scores when comparing all Eleventh Grade Experimental and Control Honor students who have taken two years of mathematics.

Ho2: There is no Gender significant difference in posttest scores when comparing all Eleventh Grade Male and Females Honor students who have taken two years of mathematics.

Ha2: There is Gender significant difference in posttest scores when comparing all Eleventh Grade Male and Female Honor students who have taken two years of mathematics.

Ho3: There is no Teacher/Group and Gender interactive significant difference in posttest scores when comparing Teachers of all Eleventh Grade Experimental/Control and Male/Female Honor students who have taken two years of mathematics.

Ha3: There is Teacher/Group and Gender interactive significant difference in posttest scores when comparing Teachers of all Eleventh Grade Experimental/Control and Male/Female Honor students who have taken two years of mathematics.

Table VIII shows the analysis of covariance of the eleventh grade intermediate students with two years of mathematics with regard to the factors of teacher/group and gender.

Table VIII

Pretest/National Posttest Scores Analysis of Covariance for All Eleventh Grade Intermediate Students With Two Years of Mathematics on the Variables: Teacher/Group and Gender
n = 50

Source	Sum of Squares	Degrees of Freedom	Mean Squared	F-Ratio	ρ **
Teacher/Group	72.693	1	72.693	2.105	.154
Gender	10.202	1	10.202	0.295	.590
Teacher/Group * Gender	15.083	1	15.083	0.437	.512
Pretest	**324.953**	**1**	**324.953**	**9.409**	**.004**
National	1.928	1	1.928	0.056	.814
Error	1519.678	44	1519.678		

** **Null Hypothesis Rejected when BOLD** Ho: = Null Hypothesis Ha: = Alternate Hypothesis

Ho1: There is no Teacher/Group significant difference in posttest scores when comparing Teachers of all Eleventh Grade Experimental and Control Intermediate students who have taken two years of mathematics.

Ha1: There is Teacher/Group significant difference in posttest scores when comparing Teachers of all Eleventh Grade Experimental and Control Intermediate students who have taken two years of mathematics.

Ho2: There is no Gender significant difference in posttest scores when comparing all Eleventh Grade Male and Females Experimental students who have taken two years of mathematics.

Ha2: There is Gender significant difference in posttest scores when comparing all Eleventh Grade Male and Female Intermediate students who have taken two years of mathematics.

Ho3: There is no Teacher/Group and Gender interactive significant difference in posttest scores when comparing Teachers of all Eleventh Grade Experimental/Control and Male/Female Intermediate students who have taken two years of mathematics.

Ha3: There is Teacher/Group and Gender interactive significant difference in posttest scores when comparing Teachers of all Eleventh Grade Experimental/Control and Male/Female Intermediate students who have taken two years of mathematics.

This comparison between teacher 3 (intermediate control) and teacher 4 (intermediate experimental) included 50 students. At the alpha level, $\rho \leq 0.05$, there were no significant scores for any factors.

Table IX

Pretest/National Posttest Scores Analysis of Covariance for All Eleventh Grade Control Students With Two Years of Mathematics on the Variables: Teacher/Level and Gender

n = 58

Source	Sum of Squares	Degrees of Freedo m	Mean Squared	F-Ratio	ρ **
Teacher/Level	**268.023**	**1**	**268.023**	**8.316**	**.006**
Gender	29.166	1	29.166	0.284	.596
Teacher/Level * Gender	30.071	1	30.071	0.933	.339
Pretest	**657.847**	**1**	**657.847**	**20.412**	**.000⁺**
National	10.976	1	10.976	0.341	.562
Error	1675.885	52	1675.885		

** **Null Hypothesis Rejected when BOLD** Ho: = Null Hypothesis Ha: = Alternate Hypothesis

Ho1: There is no Teacher/Level significant difference in posttest scores when comparing Teachers of all Eleventh Grade Control students who have taken two years of mathematics.

Ha1: There is Teacher/Level significant difference in posttest scores when comparing Teachers of all Eleventh Grade Control students who have taken two years of mathematics.

Ho2: There is no Gender significant difference in posttest scores when comparing all Eleventh Grade Male and Female Control students who have taken two years of mathematics.

Ha2: There is Gender significant difference in posttest scores when comparing all Eleventh Grade Male and Female Control students who have taken two years of mathematics.

Ho3: There is no Teacher/Level and Gender interactive significant difference in posttest scores when comparing Teachers of all Eleventh Grade Honor/Intermediate and Male/Female Control students who have taken two years of mathematics.

Ha3: There is Teacher/Level and Gender interactive significant difference in posttest scores when comparing Teachers of all Eleventh Grade Honor/Intermediate and Male/Female Control students who have taken two years of mathematics.

Table IX shows the analysis of covariance the eleventh grade control students with two years of mathematics with regard to the variables of teacher/level and gender. This comparison between teacher 1 (honors control) and teacher 3 (intermediate control) included 58 students. At the alpha level, $\rho \leq 0.05$, the teacher/level showed a significant score of $\rho = 0.006$ and the null hypothesis that there is no teacher/level significant difference in posttest scores when comparing teachers of eleventh grade control students with two years of mathematics must be rejected. Neither gender nor teacher/level * gender was significant.

Table X shows the analysis of covariance for all eleventh grade experimental students with two years of mathematics with regard to the variables of teacher/level and gender. This comparison between teacher 2 (honors experimental) and teacher 4 (intermediate experimental) included 57 students. At the alpha level, $\rho \leq 0.05$, the teacher/level showed a significant score of $\rho = .003$ and the null hypothesis that there is no teacher/level significant difference in posttest scores when comparing teachers of all eleventh grade experimental students who have taken two years of mathematics must be rejected.

Table X

Pretest/National Posttest Scores Analysis of Covariance for All Eleventh Grade Experimental Students With Two Years of Mathematics on the Variables: Teacher/Level and Gender

n = 57

Source	Sum of Squares	Degrees of Freedom	Mean Squared	F-Ratio	ρ **
Teacher/Level	**258.167**	**1**	**258.167**	**10.006**	**.003**
Gender	31.266	1	31.266	1.212	.276
Teacher/Level * Gender	3.523	1	3.523	0.137	.713
Pretest	**416.861**	**1**	**416.861**	**16.157**	**.000+**
National	56.344	1	56.344	2.184	.146
Error	1315.843	51	1315.843	25.801	

** **Null Hypothesis Rejected when BOLD** Ho: = Null Hypothesis Ha: = Alternate Hypothesis

Ho1: There is no Teacher/Level significant difference in posttest scores when comparing Teachers of all Eleventh Grade Experimental students who have taken two years of mathematics.

Ha1: There is Teacher/Level significant difference in posttest scores when comparing Teachers of all Eleventh Grade Experimental students who have taken two years of mathematics.

Ho2: There is no Gender significant difference in posttest scores when comparing all Eleventh Grade Male and Female Experimental students who have taken two years of mathematics.

Ha2: There is Gender significant difference in posttest scores when comparing all Eleventh Grade Male and

Female Experimental students who have taken two years of mathematics.

Ho3: There is no Teacher/Level and Gender interactive significant difference in posttest scores when comparing Teachers of all Eleventh Grade Honor/Intermediate and Male/Female Experimental students who have taken two years of mathematics.

Ha3: There is Teacher/Level and Gender interactive significant difference in posttest scores when comparing Teachers of all Eleventh Grade Honor/Intermediate and Male/Female Experimental students who have taken two years of mathematics.

Neither gender nor teacher/Level * gender was significant. A significant score was found between each pairing of teachers except between teacher 3 (intermediate control) and teacher 4 (intermediate experimental).

The second way to analyze the data was to compare all tenth grade intermediate students with two years of mathematics to the eleventh grade honors students with two years of mathematics. The tenth grade intermediate students had Algebra I in the eighth grade and were the same level as the eleventh grade honors students who did not have the opportunity to take Algebra I until the ninth grade. (Table VII, XI-XV).

Table XI show the analysis of covariance for the 99 tenth grade intermediate and eleventh grade honors students with two years of mathematics with regard to the variables of class, group, and gender. At the alpha level of $\rho \leq 0.05$, the group showed a significant score of $\rho = 0.010$, so that, the null hypothesis that there is no group significant difference in posttest scores when comparing all eleventh grade honor and tenth grade intermediate, control and experimental students who have taken two years of mathematics must be rejected. Class, gender, and all interactive variations showed no significance.

Table XII analyzed the same group of 99 students as Table XI with regard to teacher and gender. At the alpha level of $\rho \leq 0.05$, the teacher showed a significant score of $\rho = 0.039$, therefore, the null hypothesis that there is no teacher significant difference in posttest scores when comparing teachers of all eleventh grade honors and tenth grade intermediate students who have taken two years of mathematics must be rejected. Neither gender nor teacher * gender was significant.

Table XI

**Pretest/National Posttest Scores Analysis of Covariance
for
Tenth Grade Intermediate and Eleventh Grade Honor Students
With Two Years of Mathematics
on the
Variables: Class, Group and Gender
n = 99**

Source	Sum of Squares	Degrees of Freedom	Mean Squared	F-Ratio	ρ **
Class	37.547	1	37.547	1.757	.188
Group	**148.725**	**1**	**148.725**	**6.961**	**.010**
Gender	11.144	1	11.144	0.522	.472
Class * Group	.143	1	.143	0.007	.935
Class * Gender	15.513	1	15.513	0.726	.396
Group * Gender	.950	1	.950	0.044	.833
Class * Group * Gender	4.252	1	4.252	0.199	.657
Pretest	**1510.185**	**1**	**1510.185**	**70.686**	**.000+**
National	49.895	1	49.895	2.335	.130
Error	1901.470	89	21.365		

** **Null Hypothesis Rejected when BOLD** Ho: = Null Hypothesis Ha: = Alternate Hypothesis

Ho1: There is no Class significant difference in posttest scores when comparing all Eleventh Grade Honors and Tenth Grade Intermediate students who have taken two years of mathematics.

Ha1: There is Class significant difference in posttest scores when comparing all Eleventh Grade Honors and Tenth Grade Intermediate students who have taken two years of mathematics.

Ho2: There is no group significant difference in posttest scores when comparing all Eleventh Grade Honor and Tenth Grade Intermediate, Control and Experimental students who have taken two years of mathematics.

Ha2: There is group significant difference in posttest scores when comparing all Eleventh Grade Honor and Tenth Grade Intermediate, Control and Experimental students who have taken two years of mathematics.

Ho3: There is no Gender significant difference in posttest scores when comparing all Eleventh Grade Honors and Tenth Grade Intermediate Male and Female students who have taken two years of mathematics.

Ha3: There is Gender significant difference in posttest scores when comparing all Eleventh Grade Honors and Tenth Grade Intermediate Male and Female students who have taken two years of mathematics.

Ho4: There is no Class and Group interactive significant difference in posttest scores when comparing all Eleventh Grade Honors and Tenth Grade Intermediate Control/Experimental students who have taken two years of mathematics.

Ha4: There is Class and Group interactive significant difference in posttest scores when comparing all Eleventh Grade Honors and Tenth Grade Intermediate Control/Experimental students who have taken two years of mathematics.

Ho5: There is no Class and Gender interactive significant difference in posttest scores when comparing all Eleventh Grade Honors and Tenth Grade Intermediate and Male/Female students who have taken two years of mathematics.

Ha5: There is Class and Gender interactive significant difference in posttest scores when comparing all Eleventh Grade Honors and Tenth Grade Intermediate and Male/Female students who have taken two years of mathematics.

Ho6: There is no Group and Gender interactive significant difference in posttest scores when comparing all Eleventh Grade Honors and Tenth Grade Intermediate, Control/Experimental and Male/Female students who have taken two years of mathematics.

Ha6: There is Group and Gender interactive significant difference in posttest scores when comparing all Eleventh Grade Honors and Tenth Grade Intermediate, Control/Experimental and Male/Female students who have taken two years of mathematics.

Ho7: There is no Class, Group and Gender interactive significant difference in posttest scores when comparing all Eleventh Grade Honor and Tenth Grade Intermediate, Control/Experimental and Male/Female students who have taken two years of mathematics.

Ha7: There is Class, Group and Gender interactive significant difference in posttest scores when comparing all Eleventh Grade Honor and Tenth Grade Intermediate, Control/Experimental and Male/Female students who have taken two years of mathematics.

Table XII

Pretest/National Posttest Scores Analysis of Covariance
for
Tenth Grade Intermediate and Eleventh Grade Honor Students
With Two Years of Mathematics
on the
Variables: Teacher and Gender

n = 99

Source	Sum of Squares	Degrees of Freedom	Mean Squared	F-Ratio	ρ **
Teacher	**185.738**	**3**	**61.913**	**2.898**	**.039**
Gender	11.144	1	11.144	0.522	.472
Teacher * Gender	27.755	3	7.585	0.355	.786
Pretest	**1510.185**	**1**	**1510.185**	**70.686**	**.000⁺**
National	49.895	1	49.895	2.330	.130
Error	1901.470	89	21.365		

** **Null Hypothesis Rejected when BOLD** Ho: = Null Hypothesis Ha: = Alternate Hypothesis

Ho1: There is no Teacher significant difference in posttest scores when comparing Teachers of all Eleventh Grade Honors and Tenth Grade Intermediate students who have taken two years of mathematics.

Ha1: There is Teacher significant difference in posttest scores when comparing Teachers of all Eleventh Grade Honors and Tenth Grade Intermediate students who have taken two years of mathematics.

Ho2: There is no Gender significant difference in posttest scores when comparing all Eleventh Grade Honors and Tenth Grade Intermediate Male and Female students who have taken two years of mathematics.

Ha2: There is Gender significant difference in posttest scores when comparing all Eleventh Grade Honors and Tenth Grade Intermediate Male and Female students who have taken two years of mathematics.

Ho3: There is no Teacher and Gender interactive significant difference in posttest scores when comparing teachers of all Eleventh Grade Honors and Tenth Grade Intermediate and Male/Female students who have taken two years of mathematics.

Ha3: There is Teacher and Gender interactive significant difference in posttest scores when comparing Teacher of all Eleventh Grade Honors and Tenth Grade Intermediate and Male/Female students who have taken two years of mathematics.

The significance found in Tables XI and XII indicate that the group, and the teacher produced a significant ρ score. Group (experimental/control) is self-explanatory in nature. The teacher significance can be analyzed further to determine where the teacher significance occurs between the four teachers, (Tables VII, XIII-XV).

Table VII shows the analysis of covariance for all eleventh grade honor students with two years of mathematics with regard to the variables of group/teacher and gender. This comparison between Teacher 1 (honors control) and teacher 2 (honors experimental) included 65 students. The analysis showed a significant score of $\rho = 0.40$ for group/teacher. At the alpha level of $\rho \leq 0.05$, the null hypothesis that there is no teacher significant difference in the posttest scores when comparing eleventh grade honors experimental group and the eleventh grade honors control group must be rejected. Neither gender nor teacher/group * gender were significant. Eleventh grade honors students with two years of mathematics in the experimental group did significantly better than their counterparts in the control group.

Table XIII shows the analysis of covariance of the tenth grade intermediate students with two years of mathematics with regard to the factors of teacher/group and gender. This comparison between teacher 3 (intermediate control) and teacher 4 (intermediate experimental) included 34 students. At the alpha level, $\rho \leq 0.05$, the teacher/group showed a significant score of 0.033, so that, the null hypothesis that there is no teacher/group significant difference in posttest scores when comparing teachers of all tenth grade intermediate control/experimental students who have taken two years of mathematics must be rejected. Neither gender nor teacher/group * gender was significant.

Table XIII

Pretest/National Posttest Scores Analysis of Covariance
for
Tenth Grade Intermediate Students With Two Years of Mathematics
on the
Variables: Teacher/Group and Gender

n = 34

Source	Sum of Squares	Degrees of Freedo m	Mean Squared	F-Ratio	ρ **
Teacher/Group	**75.348**	**1**	**75.348**	**5.023**	**.033**
Gender	.049	1	.049	0.003	.955
Teacher/Group * Gender	3.785	1	3.785	0.252	.619
Pretest	**714.159**	**1**	**714.159**	**47.606**	**.000⁺**
National	2.112	1	2.112	0.141	.710
Error	420.037	28	15.001		

** **Null Hypothesis Rejected when BOLD** Ho: = Null Hypothesis Ha: = Alternate Hypothesis

Ho1: There is no Teacher/Group significant difference in posttest scores when comparing Teachers of all Tenth Grade Intermediate Control/Experimental students who have taken two years of mathematics.

Ha1: There is Teacher/Group significant difference in posttest scores when comparing Teachers of all Tenth Grade Intermediate Control/Experimental students who have taken two years of mathematics.

Ho2: There is no Gender significant difference in posttest scores when comparing all Eleventh Grade Honors and Tenth Grade Intermediate Male and Female students who have taken two years of mathematics.

Ha2: There is Gender significant difference in posttest scores when comparing all Tenth Grade Intermediate Male and Female students who have taken two years of mathematics.

Ho3: There is no Teacher and Gender interactive significant difference in posttest scores when comparing teachers of all Tenth Grade Intermediate and Male/Female students who have taken two years of mathematics.

Ha3: There is Teacher and Gender interactive significant difference in posttest scores when comparing Teacher of all Tenth Grade Intermediate and Male/Female students who have taken two years of mathematics.

Table XIV shows the analysis of covariance for eleventh grade honors students and tenth grade intermediate students in the control group with two years of mathematics with regard to the variables of teacher/class and gender.

Table XIV

Pretest/National Posttest Scores Analysis of Covariance for Eleventh Grade Honors Control and Tenth Grade Intermediate Control Students With Two Years of Mathematics on the Variables: Teacher/Class and Gender

n = 51

Source	Sum of Squares	Degrees of Freedom	Mean Squared	F-Ratio	ρ **
Teacher/Class	7.205	1	7.205	0.394	.533
Gender	.939	1	.939	0.051	.822
Teacher/Class * Gender	8.871	1	8.871	0.485	.490
Pretest	**797.646**	**1**	**797.646**	**43.651**	**.000⁺**
National	**43.292**	**1**	**43.292**	**5.105**	**.029**
Error	827.304	45	827.304		

** **Null Hypothesis Rejected when BOLD** Ho: = Null Hypothesis Ha: = Alternate Hypothesis

Ho1: There is no Teacher/Class significant difference in posttest scores when comparing Teachers of all Eleventh

Grade Honor Control and Tenth Grade Intermediate Control students who have taken two years of mathematics.

Ha1: There is Teacher/Class significant difference in posttest scores when comparing Teachers of all Eleventh Grade Honor Control and Tenth Grade Intermediate Control students who have taken two years of mathematics.

Ho2: There is no Gender significant difference in posttest scores when comparing all Eleventh Grade Honor Control and Tenth Grade Intermediate Control Male and Female students who have taken two years of mathematics.

Ho2: There is Gender significant difference in posttest scores when comparing all Eleventh Grade Honor Control and Tenth Grade Intermediate Control Male and Female students who have taken two years of mathematics.

Ho3: There is no Teacher/Class and Gender interactive significant difference in posttest scores when comparing teachers of all Eleventh Grade Honor Control and Tenth Grade Intermediate Control and Male/Female students who have taken two years of mathematics.

Ha3: There is Teacher/Class and Gender interactive significant difference in posttest scores when comparing teachers of all Eleventh Grade Honor Control and Tenth Grade Intermediate Control and Male/Female students who have taken two years of mathematics.

The comparison between teacher 1 (honors control) and teacher 3 (intermediate control) included 51 students. At the alpha level, $\rho \leq 0.05$, none of the factors showed a significance. Teacher/class, gender, and teacher/class * gender were not significant. The null hypotheses must be accepted as true.

Table XV shows the analysis of covariance for all eleventh grade honors experimental students and tenth grade intermediate experimental students with two years of mathematics with regard to the variables of teacher/class and gender. This comparison between teacher 2 (honors experimental) and teacher 4 (intermediate experimental) included 48 students. At the alpha level, $\rho \leq 0.05$, none of the factors were significant. Teacher/class, gender, and teacher/class * gender were not significant. The null hypotheses must be accepted as true. There is no significant difference between the two control teachers or between the two experimental teachers. It appears that the teaching between the two control teachers and the two experimental teachers was equivalent. A plausible explanation for the fact that there is significance between the control and experimental sections of teacher 1 and 2 and teacher 3 and 4 was that the experimental treatment made the difference.

Table XV

Pretest/National Posttest Scores Analysis of Covariance for Eleventh Grade Honors Experimental and Tenth Grade Intermediate Experimental Students With Two Years of Mathematics on the Variables: Teacher/Class and Gender

n = 48

Source	Sum of Squares	Degrees of Freedom	Mean Squared	F-Ratio	ρ **
Teacher/Class	26.641	1	26.641	1.000	.323
Gender	4.122	1	4.122	0.167	.685
Teacher/Class * Gender	21.000	1	21.000	0.853	.361
Pretest	**750.872**	**1**	**750.872**	**30.486**	**.000**[+]
National	1.155	1	1.155	0.047	.830
Error	1034.465	42	24.630		

** **Null Hypothesis Rejected when BOLD** Ho: = Null Hypothesis Ha: = Alternate Hypothesis

Ho1: There is no Teacher/Class significant difference in posttest scores when comparing Teachers of all Eleventh Grade Honor Experimental and Tenth Grade Intermediate Experimental students who have taken two years of mathematics.

Ha1: There is Teacher/Class significant difference in posttest scores when comparing Teachers of all Eleventh Grade Honor Experimental and Tenth Grade Intermediate Experimental students who have taken two years of mathematics.

Ho2: There is no Gender significant difference in posttest scores when comparing all Eleventh Grade Honor Experimental and Tenth Grade Intermediate Experimental Male and Female students who have taken two years of mathematics.

Ho2: There is Gender significant difference in posttest scores when comparing all Eleventh Grade Honor Experimental and Tenth Grade Intermediate Experimental Male and Female students who have taken two years of mathematics.

Ho3: There is no Teacher/Class and Gender interactive significant difference in posttest scores when comparing teachers of all Eleventh Grade Honor Experimental and Tenth Grade Intermediate Experimental and Male/Female students who have taken two years of mathematics.

Ha3: There is Teacher/Class and Gender interactive significant difference in posttest scores when comparing teachers of all Eleventh Grade Honor Experimental and Tenth Grade Intermediate Experimental and Male/Female students who have taken two years of mathematics.

The third possible analysis was to compare the eleventh grade students with either one or two years of mathematics taught by teacher 3 and teacher 4. (Tables VIII, XVI-XIX)

Table XVI shows the analysis of covariance for all 75 eleventh grade intermediate students with regard to the variables of years of mathematics, teacher/group and gender.

Table XVI

Pretest/National Posttest Scores Analysis of Covariance for All Eleventh Grade Intermediate Students on the Variables: Years of Mathematics, Teacher/Group and Gender

n = 75

Source	Sum of Squares	Degrees of Freedo m	Mean Squared	F-Ratio	ρ **
Years of Mathematics	90.321	1	90.321	3.456	.068
Teacher/Group	**124.341**	**1**	**124.341**	**4.757**	**.033**
Gender	2.835	1	2.835	0.108	.743
Years of Math * Teacher/Group	1.860	1	1.860	0.071	.790
Years of Math * Gender	5.684	1	5.684	0.217	.643
Teacher/Group * Gender	.062	1	.062	0.002	.961
Years * Teacher/Group * Gender	26.642	1	26.642	1.019	.316
Pretest	**412.349**	**1**	**412.349**	**15.777**	**.000+**
National	8.579	1	8.579	0.328	.569
Error	1698.881	65	26.137		

** **Null Hypothesis Rejected when BOLD** Ho: = Null Hypothesis Ha: = Alternate Hypothesis

Ho1: There is no Years of Mathematics significant difference in posttest scores when comparing all Eleventh Grade Intermediate students who have taken one or two years of mathematics.

Ha1: There is Years of Mathematics significant difference in posttest scores when comparing all Eleventh Grade Intermediate students who have taken one or two years of mathematics.

Ho2: There is no Teacher/Group significant difference in posttest scores when comparing Teachers of all Eleventh Grade Intermediate Experimental and Control students who have taken one or two years of mathematics.

Ha2: There is Teacher/Group significant difference in posttest scores when comparing Teachers of all Eleventh Grade Intermediate Experimental and Control students who have taken one or two years of mathematics.

Ho3: There is no Gender significant difference in posttest scores when comparing all Eleventh Grade Intermediate Male and Female students who have taken one or two years of mathematics.

Ha3: There is Gender significant difference in posttest scores when comparing all Eleventh Grade Intermediate Male and Female students who have taken one or two years of mathematics.

Ho4: There is no Years of Mathematics and Teacher/Group significant difference in posttest scores when comparing Teachers of all Eleventh Grade Intermediate Control/Experimental students who have taken one or two years of mathematics.

Ha4: There is Years of Mathematics and Teacher/Group significant difference in posttest scores when comparing Teachers of all Eleventh Grade Intermediate Control/Experimental students who have taken one or two years of mathematics.

Ho5: There is no Years of Mathematics and Gender significant difference in posttest scores when comparing all Eleventh Grade Intermediate students who have taken one or two years of mathematics.

Ha5: There is Years of Mathematics and Gender significant difference in posttest scores when comparing all Eleventh Grade Intermediate students who have taken one or two years of mathematics.

Ho6: There is no Teacher/Group and Gender significant difference in posttest scores when comparing Teachers of all Eleventh Grade Intermediate Male and Female students.

Ha6: There is Teacher/Group and Gender significant difference in posttest scores when comparing Teachers of all Eleventh Grade Intermediate Male and Female students.

Ho7: There is no Years os Mathematics and Teacher/Group and Gender significant difference in posttest scores when comparing Teachers of all Eleventh Grade Intermediate Male and Female students who have taken one or more years of mathematics.

Ha7: There is Years os Mathematics and Teacher/Group and Gender significant difference in posttest scores when comparing Teachers of all Eleventh Grade Intermediate Male and Female students who have taken one or more years of mathematics.

At the alpha level of $\rho \leq 0.05$, a significant score of $\rho = 0.033$ was found for teacher/group, so that, the null hypothesis that there is no teacher/group significant difference in posttest scores when comparing teachers of all eleventh grade intermediate experimental and control students who have taken one or two years of mathematics must be rejected. Years of mathematics, gender, and any interactive combination thereof was not significant.

Table XVII shows the analysis of covariance for all 25 eleventh grade intermediate students with one year of mathematics with regard to the variables of teacher/group and gender.

Table XVII

Pretest/National Posttest Scores Analysis of Covariance
for
All Eleventh Grade Intermediate Students With One Year of Mathematics
on the
Variables: Teacher/Group and Gender

n = 25

Source	Sum of Squares	Degrees of Freedom	Mean Squared	F-Ratio	ρ **
Teacher/Group	**49.983**	**1**	**49.983**	**4.679**	**.043**
Gender	.017	1	.017	0.002	.966
Teacher/Group * Gender	11.989	1	11.989	1.305	.268
Pretest	**83.631**	**1**	**83.631**	**9.103**	**.007**
National	6.854	1	6.854	0.746	.399
Error	174.559	19	9.187		

** **Null Hypothesis Rejected when BOLD** Ho: = Null Hypothesis Ha: = Alternate Hypothesis

Ho1: There is no Teacher/Group significant difference in posttest scores when comparing Teachers of all Eleventh Grade Intermediate Experimental and Control students who have taken one year of mathematics.

Ha1: There is Teacher/Group significant difference in posttest scores when comparing Teachers of all Eleventh Grade Intermediate Experimental and Control students who have taken one year of mathematics.

Ho2: There is no Gender significant difference in posttest scores when comparing all Eleventh Grade Intermediate Male and Female students who have taken one year of mathematics.

Ha2: There is Gender significant difference in posttest scores when comparing all Eleventh Grade Intermediate Male and Female students who have taken one year of mathematics.

Ho3: There is no Teacher/Group and Gender significant difference in posttest scores when comparing Teachers of all Eleventh Grade Intermediate Male and Female students who have taken one year of Mathematics.

Ha3: There is Teacher/Group and Gender significant difference in posttest scores when comparing Teachers of all Eleventh Grade Intermediate Male and Female students who have taken one year of Mathematics.

At the alpha level of $\rho \le 0.05$, a significant score of $\rho = 0.043$ was found for teacher/group, so that, the null hypothesis that there is no group/teacher significant difference in posttest scores when comparing teachers of all eleventh grade intermediate experimental and control students who have taken one year of mathematics must be rejected. Neither gender nor teacher/group * gender was significant. Table VIII previously displayed on page 91?, shows the analysis of covariance of the eleventh grade intermediate students with two years of mathematics with regard to the factors of teacher/group and gender. This comparison between teacher 3 (intermediate control) and teacher 4 (intermediate experimental) included 50 students. At the alpha level, $\rho \le 0.05$, there was no significant score for any factors.

Table XVIII shows the analysis of covariance for all 38 eleventh grade intermediate control students with one or two years of mathematics with regard to the variables of years of mathematics and gender. At the alpha level of $\rho \le 0.05$, no significant score was found for years of mathematics, gender, or any combination thereof. All null hypotheses in Table XVIII must be accepted as true.

Table XIX shows the analysis of covariance for all 37 eleventh grade intermediate experimental students who have taken one or two years of mathematics with regard to the variables of years of mathematics and gender.

Table XVIII

Pretest/National Posttest Scores Analysis of Covariance for All Eleventh Grade Intermediate Control Students on the Variables: Years of Mathematics and Gender

n = 38

Source	Sum of Squares	Degrees of Freedom	Mean Squared	F-Ratio	ρ **
Years of Mathematics	110.071	1	110.071	3.523	.070
Gender	.106	1	.106	0.003	.954
Years of Mathematics * Gender	34.079	1	34.079	1.091	.304
Pretest	**259.160**	**1**	**259.160**	**8.294**	**.007**
National	9.126	1	9.126	0.792	.593
Error	999.865	32	31.246		

** **Null Hypothesis Rejected when BOLD** Ho: = Null Hypothesis Ha: = Alternate Hypothesis

Ho1: There is no Years of Mathematics significant difference in posttest scores when comparing all Eleventh Grade Intermediate Control students who have taken one year or two years of mathematics.

Ha1: There is Years of Mathematics significant difference in posttest scores when comparing all Eleventh Grade

Intermediate Control students who have taken one year or two years of mathematics.

Ho2: There is no Gender significant difference in posttest scores when comparing all Eleventh Grade Intermediate Control students who have taken one or two years of mathematics.

Ha2: There is Gender significant difference in posttest scores when comparing all Eleventh Grade Intermediate Control students who have taken one or two year of mathematics.

Ho3: There is no Years of Mathematics and Gender interactive significant difference in posttest scores when comparing all Eleventh Grade Intermediate Control Male and Female students who have taken one or two years of Mathematics.

Ho3: There is Years of Mathematics and Gender interactive significant difference in posttest scores when comparing all Eleventh Grade Intermediate Control Male and Female students who have taken one or two years of Mathematics.

At the alpha level $\rho \leq 0.05$, no significant score was found for years of mathematics, gender, or any combination thereof. All null hypotheses in Table XIX must be accepted as true. The overall comparison between the control and the experimental groups showed a significant difference for teacher/group. The only other significant score was between the eleventh grade control and experimental students with one year of mathematics. The fact that the students with one year of mathematics would show a gain when the students with two years of mathematics did not is unusual?

Table XIX

**Pretest/National Posttest Scores Analysis of Covariance
for
All Eleventh Grade Intermediate Experimental Students
on the
Variables: Years of Mathematics and Gender**

n = 37

Source	Sum of Squares	Degrees of Freedom	Mean Squared	F-Ratio	ρ **
Years of Mathematics	10.008	1	10.008	0.488	.490
Gender	1.168	1	1.168	0.057	.813
Years of Mathematics * Gender	17.113	1	17.113	0.519	.448
Pretest	**159.859**	**1**	**159.859**	**7.800**	**.009**
National	63.104	1	63.104	3.079	.089
Error	635.345	31	70.495		

** **Null Hypothesis Rejected when BOLD** Ho: = Null Hypothesis Ha: = Alternate Hypothesis

Ho1: There is no Years of Mathematics significant difference in posttest scores when comparing all Eleventh Grade Intermediate Experimental students who have taken one year or two years of mathematics.

Ha1: There is Years of Mathematics significant difference in posttest scores when comparing all Eleventh Grade Intermediate Experimental students who have taken one year or two years of mathematics.

Ho2: There is no Gender significant difference in posttest scores when comparing all Eleventh Grade Intermediate Experimental students who have taken one or two years of mathematics.

Ha2: There is Gender significant difference in posttest scores when comparing all Eleventh Grade Intermediate Experimental students who have taken one or two year of mathematics.

Ho3: There is no Years of Mathematics and Gender interactive significant difference in posttest scores when comparing all Eleventh Grade Intermediate Experimental Male and Female students who have taken one or two years of Mathematics.

Ho3: There is Years of Mathematics and Gender interactive significant difference in posttest scores when comparing all Eleventh Grade Intermediate Experimental Male and Female students who have taken one or two years of Mathematics.

Chapter Summary

This chapter has discussed the statistical analysis process in general and in reference to this study. The analyses of the statistical data derived from the experimental and control program as it applied to the aims of this study are explained in detail. Conclusions that can be drawn from these results, their implications for teaching and learning, and recommendations for further study will be discussed in the next chapter.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Overview

Testing has been influential in shaping education in several ways. Testing has been used to assess the effectiveness of teaching and learning and as tool for comparison on a local, state, national, or international level. Tests have been used to determine what curricula are taught and to determine whether a student graduates from high school. Testing has been used to determine college admission. It is within the last context that this study was done. The use of coaching, practice, and test-taking skills instruction have been a controversial issue for the last fifteen to twenty years. The primary purpose of this study was to determine whether student scores on the mathematics portion of the American College Testing Program (ACT) test could be improved by practicing sample test items and studying general test-taking strategies. Review of prior research in this area showed that in general practice and test-taking skills instruction did seem to have a positive effect in the area of mathematics. One key factor related to the effectiveness of practice in improving scores was the design of the study. Prior research found that the pretest/posttest design yielded larger gains. Teaching test-taking skills produced a positive effect at all grade levels. Studies done at the elementary, secondary, and college level have produced significant differences between the experimental and control groups. The best gains in scores were found when the practice was given on a regular basis over a longer period of time, the practice items were similar to the actual test questions, and a pretest was given prior to coaching. The overall gain from coaching activities is better among high-ability students. This study used a pretest, had practice items similar to the actual test, used practice spaced over a period of time, and held students responsible by including practice test items on regular content testing. The overall results of this study reiterated the findings of prior research.

This pretest/posttest study was designed to measure the effect of using practice mathematics items and test preparation activities on the American College Testing Program (ACT) scores of students enrolled in a second-year algebra class. Students in the experimental group were given spaced practice of sample ACT mathematics questions and test-preparation suggestions for the mathematics section of the ACT. The primary purpose of this study was to determine whether student scores on the mathematics portion of the ACT test could be improved by practicing sample test items and studying general test-taking strategies. There were three secondary purposes included in this study. One was to determine whether a student's gender had any significance in ACT performance, the second was to determine whether the number of years of mathematics study affected achievement, and the third purpose was to examine the role of the student's prior achievement level.

This study of coaching for a standardized achievement test like the American College Testing Program (ACT) using practice test items and test-taking strategies was designed to answer the following research questions:

*Research Question One: Will the use of practice test items similar to the mathematics questions given on the American College Testing (ACT) exam and test-taking suggestions significantly improve the ACT mathematics score of second-year algebra students?

*Research Question Two: Does gender have any significance in the performance of second-year algebra students on the mathematics section of the ACT?

*Research Question Three: Does the number of years of prior mathematics study affect achievement on the mathematics section of the ACT?

*Research Question Four: Does the level of the student's prior achievement have any effect on the student's performance?

The researcher conducted the study in six phases. In Phase I materials were prepared for use with the students. ACT practice materials were used by the researcher as a model to write sample practice questions and test items (Appendix B and C). The items were sequenced to match the order of topics traditionally taught in second-year algebra. The pretesting of the eight sections of second-year algebra students was done in Phase II using the ACT Practice Test Form 9039K and the National Proficiency Survey Series: Algebra II. Phase III was the implementation of the ten-week study. The experimental groups received treatment with daily practice ACT test items and similar practice items on content exams. The posttesting was done in Phase IV using the ACT Practice Test Form 9039K. Phase V was the statistical analysis of the data generated by the study. Phase VI was the reporting of statistical significance and the implications of the results.

The students in the study were compared on the dependent variables of years of mathematics (one or two), class (tenth or eleventh), level (honors or intermediate), teacher (1,2,3, or 4), gender (male or female), and group (control or experimental).

Summary Analysis of Findings

In a response to the four research questions, the results of the study along with a repetition of the questions are presented for inspection. In all cases, an alpha level of $\rho \leq 0.05$ is considered significant.

Research Question One

The first question was:

> Will the use of practice test items similar to the mathematics questions given on the American College Testing (ACT) exam and test-taking suggestions significantly improve the ACT mathematics score of second-year algebra students?

Findings: This study was analyzed in three different subsets: Subset A: all eleventh grade students with two years of mathematics; Subset B: the eleventh grade honors students and the tenth grade intermediate students, each with two years of mathematics; and Subset C: the eleventh grade intermediate students with one or two years of mathematics. Significance was found for the 115 eleventh grade/two years of mathematics students in Subset A for three areas. Level was found to be significant at $\rho = 0.028$, Group was found to be significant at $\rho = 0.000^{\cdot}$, and Teacher was found to be significant at $\rho = 0.000^{\cdot}$. The analysis of the 99 eleventh grade honors students and the tenth grade intermediate students with two years of mathematics in Subset B showed significance for two areas. Group was found to be significant at $\rho = 0.010$ and Teacher was found to be significant at $\rho = 0.039$. The analysis of the 75 eleventh grade intermediate students with one or two years mathematics in Subset C showed significance for one area. Teacher/Group was found to be significant at $\rho = 0.033$.

Discussion: In each subset, teacher/group showed significance. Students in the experimental group who received treatment outscored their counterparts in the control group. The null hypothesis that there is no group significant difference when comparing students from the experimental to the control group is false and therefore, must be rejected. As an aside, in Subset A where some students were in the honors level and others in the intermediate level, Level was found to be significant. This result seems to support the prior research that students with more prior achievement seem to benefit more from coaching practice situations.

Research Question Two

The second question was:

> Does gender have any significance in the performance of second-year algebra students on the mathematics section of the ACT?

Findings: All three subsets of the study contained male and females. In Subset A gender produced a score of $\rho = 0.718$, in Subset B gender produced a score of $\rho = 0.472$, and in Subset C gender produced a score of $\rho = 0.743$.

Discussion: In all three subsets of the study, gender did not come close to a significant score of $\rho < 0.05$. It appears that gender does not have any bearing on a student's ability to learn or benefit from coaching practice or test-taking suggestions. Male and Female students seem to have an equal opportunity to learn skills and/or process information equally well to be successful on the ACT. We must accept the null hypothesis that there is no gender significant difference in posttest scores when comparing all students in the study who have taken one or two years of mathematics.

Research Question Three

The third question was:

> Does the number of years of prior mathematics study affect achievement on the mathematics section of the ACT?

Findings: The only subset of the study where the number of years of mathematics (one or two) was not the same was in Subset C. Subset C analyzed the 75 eleventh grade intermediate students taught by Teacher 3 and Teacher 4. Years of mathematics was not found to be significant with a $\rho = 0.068$ (Table XVI).

Discussion: Since the years of mathematics produced a score of $\rho = 0.068$, the null hypothesis that there is no years of mathematics significant difference in posttest scores when comparing all eleventh grade intermediate students who have taken one or two years of mathematics must be accepted as true. There were 25 students with one year of mathematics and 50 students with two years of mathematics for a total of 75 students in the study. Is it possible that the imbalance of numbers gave a skewed result?

Research Question Four

The fourth question was:

> Does the level of student's prior achievement have any effect on the student's performance?

Findings: In the analysis of the 115 eleventh grade students with two years of mathematics (Table V), the level of that student showed a significant score of $\rho = 0.000^{+}$.

Students in the honors sections showed more improvement than their intermediate counterparts.

Discussion: The students who had the better prior achievement did seem to gain more from the practice than the average college preparatory students. This result could be seen in the analysis of all the eleventh students with two years of mathematics. The results were confirmed with the analysis of eleventh grade control students in Table IX and in the analysis of the eleventh grade experimental students in Table X.

Conclusions

Subset A: (Eleventh Grade Honors and Intermediate Students with Two Years of Mathematics)

In the analysis of the eleventh grade honors and intermediate students with two years of mathematics, the researcher found that there was significance in the level (honors/intermediate) that can be explained by the fact that the honors students were selected for the honor class according to their superior prior achievement. The expectation would be that the honor students would have scored better than their intermediate counterparts. Table V and VI show that there was also significance in group and teacher. There was significance between the honors control and honors experimental group shown in Table VII. Significance was not found in the comparison between the intermediate control and intermediate experimental groups (Table VIII). Tables IX and X showed that there was teacher/level significance between the honors and intermediate control and the honors and intermediate experimental students. The researcher found overall significance between control and experimental groups in level which would be expected . Significance was also found with teacher and group. The difficulty is that since each distinct group also had a distinct teacher, the researcher cannot be sure which factor group or teacher produced the results. The combination of teacher 1 and teacher 3 (control, Table IX) compared to teacher 2 and teacher 4 (experimental, Table X) produced a significant result. Teacher 1 and teacher 2 (honors control and experimental, Table VII) produced a significant result. Teacher 3 and teacher 4 (intermediate control and experimental, Table VIII) did not show a significance. Significance was also found between teacher 1 and teacher 3 (honors and intermediate control, Table IX) and teacher 2 and teacher 4 (honors and intermediate experimental, Table X). It appears that the honors students gained more from the teaching or the treatment than did their intermediate counterparts. However, a significant difference was found between the honors students that had teacher 1 and teacher 2. It would be expected that there would be no significant difference between these students were at the same prior achievement level yet the experimental group did significantly better than the control group. In the intermediate comparison between teacher 3 and teacher 4, the same difference was not found. The researcher would like to consider the analysis of Subset B before drawing any more conclusions. Appendix I contains a summary of subset A tables with significance and a figure that shows relationships between the control and experimental groups.

Subset B: (Eleventh Grade Honors and Tenth Grade Intermediate Students with Two Years of Mathematics)

These students were at the same prior achievement level. The difference between them was that the tenth grade intermediate students had the opportunity to take Algebra I in the eighth grade, while the eleventh grade students were not taught Algebra I until ninth grade. In this analysis, the researcher had the opportunity to make a comparison between all four teachers. In the overall analysis of control students (teachers 1 and 3) to experimental students (teachers 2 and 4), the researcher found significance for the variables of group (Table XI) and teacher (Table XII). It is not possible to distinguish between these two variables since each distinct group had a distinct teacher. Either the teacher or the treatment or both had an effect on the students. As the researcher continued to examine the relationship between each teacher, an interesting result appears. Significance is found between the eleventh grade control and experimental groups (teachers 1 and 2, Table VII) and in the tenth grade control and experimental groups (teachers 3 and 4, Table XIII). However, significance is not found between the tenth and eleventh grade control students (teachers 1 and 3, Table XIV); nor is it found between the tenth and eleventh grade experimental students (teachers 2 and 4, Table XV). Therefore, the researcher might conclude that since there is no significant difference between the teachers in the control and experimental groups, that the treatment administered to the experimental groups had an effect. If the researcher were to consider the information from both Subset A and Subset B, it would seem that the better students benefitted more from the teaching or the treatment. This is evidenced by the fact that both tenth and eleventh grade students at the higher level showed a significant difference in the study. The eleventh grade intermediate students with two years of mathematics (teacher 3 and 4, Table VIII) showed no difference. This information also seemed to indicate that since the teaching between the control teachers (1 and 3) and the experimental teachers (2 and 4) with the better tenth and eleventh grade students showed no significant difference and that there was no difference between the eleventh intermediate control and experimental teachers (3 and 4) that possibly the teaching was not the key factor. The conclusion might be that the treatment more than the teaching had a positive effect on the students. Appendix J contains a summary of subset B tables with significance and a figure that shows relationships between control and experimental groups.

Subset C: (Eleventh Grade Intermediate Students with One or Two Years of Mathematics)

In the comparison of the eleventh grade intermediate students taught by teachers 3 and 4, significance was found between the control and experimental groups for the variable of teacher/group. In the detailed analysis of the two teachers, the data showed that there was no significance between the control students with one or two years of mathematics (Table XVIII) or the experimental students with one or two years of mathematics (Table XIX). There was no significant difference between the eleventh grade control and experimental groups (Table VIII). The only significance in this subset was between the eleventh grade control and experimental students with one year of mathematics (Table XVII). It appears that students who had only one year of mathematics received benefit from the treatment while two year students did not. A plausible explanation of this phenomenon was that the students with one year of mathematics had taken Algebra I the preceding year. Their knowledge base of Algebra I might have been fresher than their counterparts with two years of mathematics who had taken Algebra I and Geometry prior to Algebra II. It appears that the treatment works for the better student as indicated by the prior research in this area. It also appears that for the normal college prep student the treatment works better if the students takes the mathematical sequence of Algebra I, Algebra II, and Geometry instead of the traditional sequence of Algebra I, Geometry, and Algebra II. Appendix K contains a summary of subset C tables with significance and a figure that shows relationships between the control and experimental groups.

Recommendations for Further Research

The first recommendation is to continue research about coaching for standardized tests like the ACT. The study could be replicated with a different population (e.g., heterogenous classes) instead of stratified ones. The study design could be changed to use one instructor for both the control and the experimental groups. The study could be done using a different standardized test like the SAT.

The second recommendation is that a two year pretest/posttest study could be implemented using actual Pre ACT scores from the P-ACT test given to high school sophomores in October each year as the pretest score. Treatment could be given in the Geometry in the sophomore year and Second-year Algebra classes in the junior year. Most students take the actual ACT test in the spring of the junior year. The posttest score for the study would be the student's actual ACT test score from the first attempt at the ACT.

The third recommendation is that a study could be designed to compare the results from a free public school test preparation course to one that is commercially presented (e.g. Stanley Kaplan).

The fourth recommendation is that the researcher could design a study to study the effects of preparation on students who must take graduate level exam like the Graduate Record Exam (GRE).

The fifth recommendation for further study could do a follow-up study of the students who went on to college after a coaching or test preparation course. Did these students perform up the predicted indicator of the test or did the test preparation inflate their score to a point that it invalidated the value of the admission screening process?

Summary

Students who received coaching help did significantly outscore their counterparts in the control group. The use of practice test items and test-taking suggestions has a positive effect on the student's ability to achieve on a standardized test like the ACT. The gender of the student has no bearing on the ability to process and to improve achievement on a standardized test like the ACT. The prior research and the results of this study indicate that the teaching of mathematical content may not be all that is needed for optimizing performance on standardized tests like the ACT. There are several implications for educators concerning the effects of practice. Educators should be using spaced practice of materials over a long period of time. The performance gains on the test are directly related to how closely practice items are matched to the actual test. The use of a pretest/posttest design with students seems to produce better results in that it gives students a picture of what they are going to learn. This preview of coming attractions (the pretest) seems to produce better results. The old adage, "Practice makes perfect" has different implications for educators today.

Preparing for the ACT Assessment

Contents

ACT endorses the *Code of Fair Testing Practices in Education*, a statement of the obligations to test takers of those who develop, administer or use educational tests and data. The *Code* sets forth criteria for fairness in four areas: developing and selecting appropriate tests, interpreting test scores, striving for fairness, and informing test takers. ACT is committed to ensuring that each of its testing programs upholds the *Code*'s standards as they apply to test developers.

A copy of the full *Code* may be obtained free of charge from ACT Publications, P.O. Box 168, Iowa City, Iowa 52243, 319/337-1429

A Message to Students

The best indication of how well you will do in college is a measure of how well you can perform the skills necessary for college coursework. The ACT Assessment—chances are, you and your classmates call it simply "the ACT"—measures these skills in four major curriculum areas: English, mathematics, reading, and science reasoning. These areas are tested because they include the major areas of instruction in most high school and college programs.

This booklet, which is provided free of charge, is intended to help you do your best on the ACT. It summarizes general test-taking strategies, describes the content of each of the tests, provides specific tips for each, and lets you know what you can expect on the test day. Included in this booklet are a practice test—a "retired" form of the ACT Assessment that was administered to students on a national test date—and a sample answer sheet and scoring instructions.

Read this booklet carefully and take the practice test well before the test day so you will be familiar with the ACT, what it measures, and the strategies you can use to do your best on it.

Table I

Number of Students Taught
by
Each Faculty Member and Grade and Years of Mathematics

	11th Grade		10th Grade		
	1 Year of Mathematics	2 Years of Mathematics	1 Year of Mathematics	2 Years of Mathematics	
Teacher 1 Honors Control	(1) 0	(2) 33	(3) 0	(4) 0	33
Teacher 2 Honors Experimental	(5) 0	(6) 32	(7) 0	(8) 0	32
Teacher 3 Intermediate Control	(9) 13	(10) 25	(11) 8	(12) 18	64
Teacher 4 Intermediate Experimental	(13) 12	(14) 25	(15) 14	(16) 16	67
Totals	25	115	22	34	196

(#) indicates the cell number used in the following charts

APPENDIX I

Subset A: 11th Grade Honors and Intermediate Students with 2 Years

V

(115)
Level Group

2,10 vs 6,14

VI

(115)
Teacher

2,10 vs 6,14

VII

(65)
Teacher/Group

2 vs 6
Honors

VIII

(50)

10 vs 14
Inter.

IX

(58)
Teacher/Level

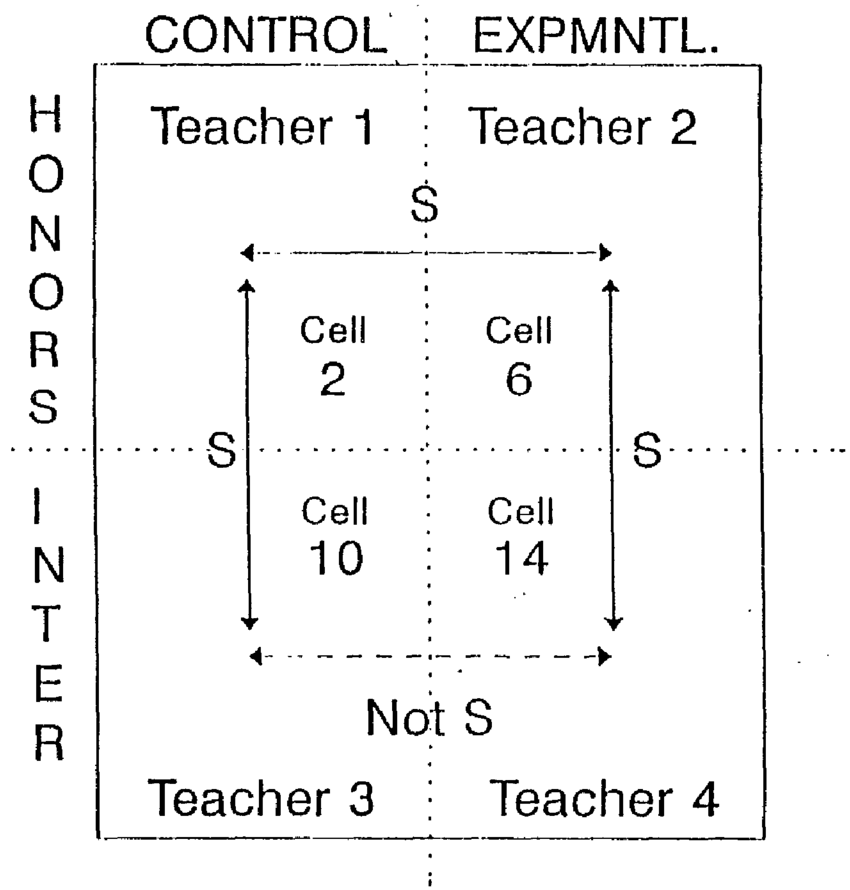
2 vs 10
Control

X

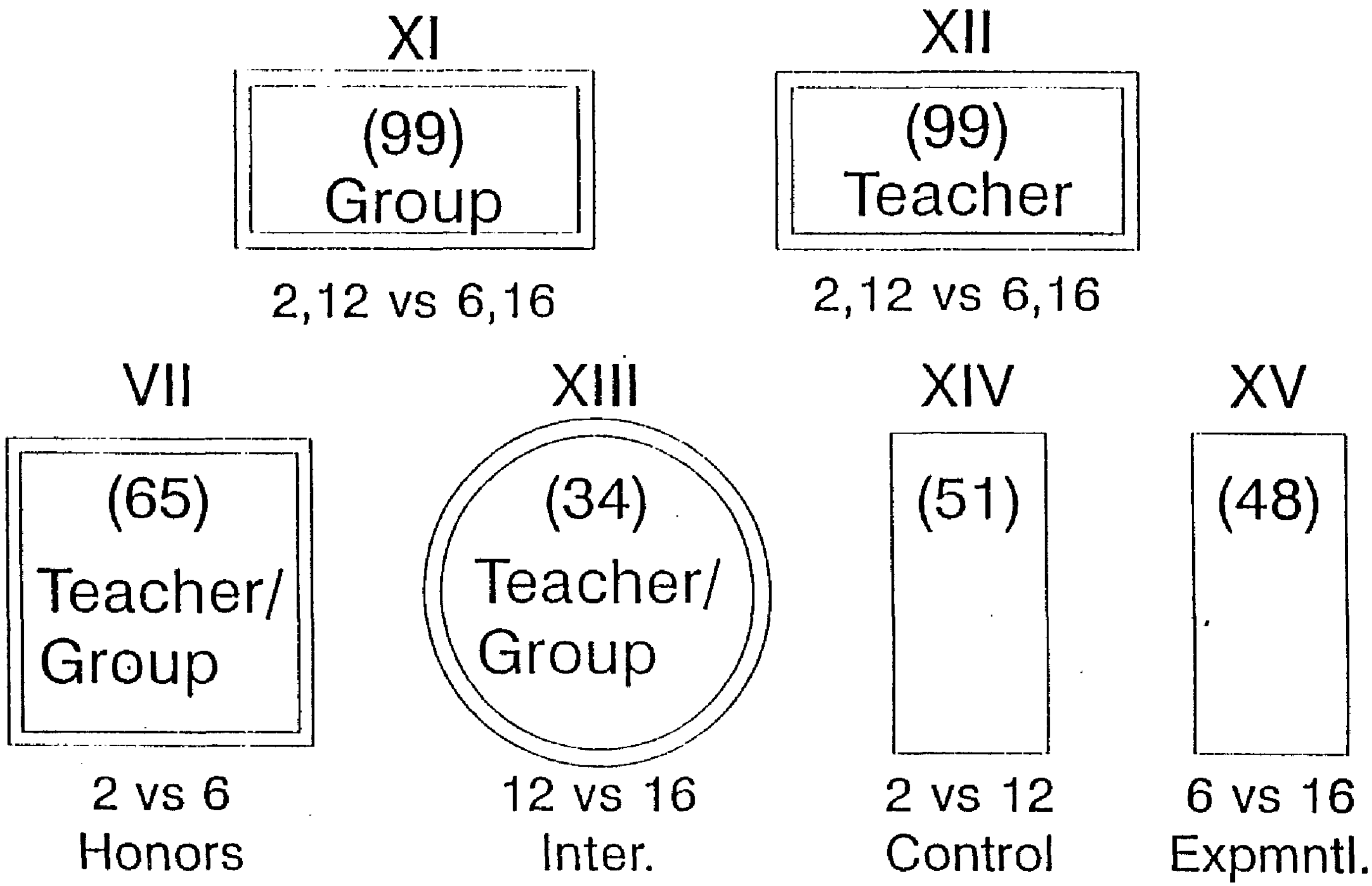
(57)
Teacher/Level

6 vs 14
Expmntl.

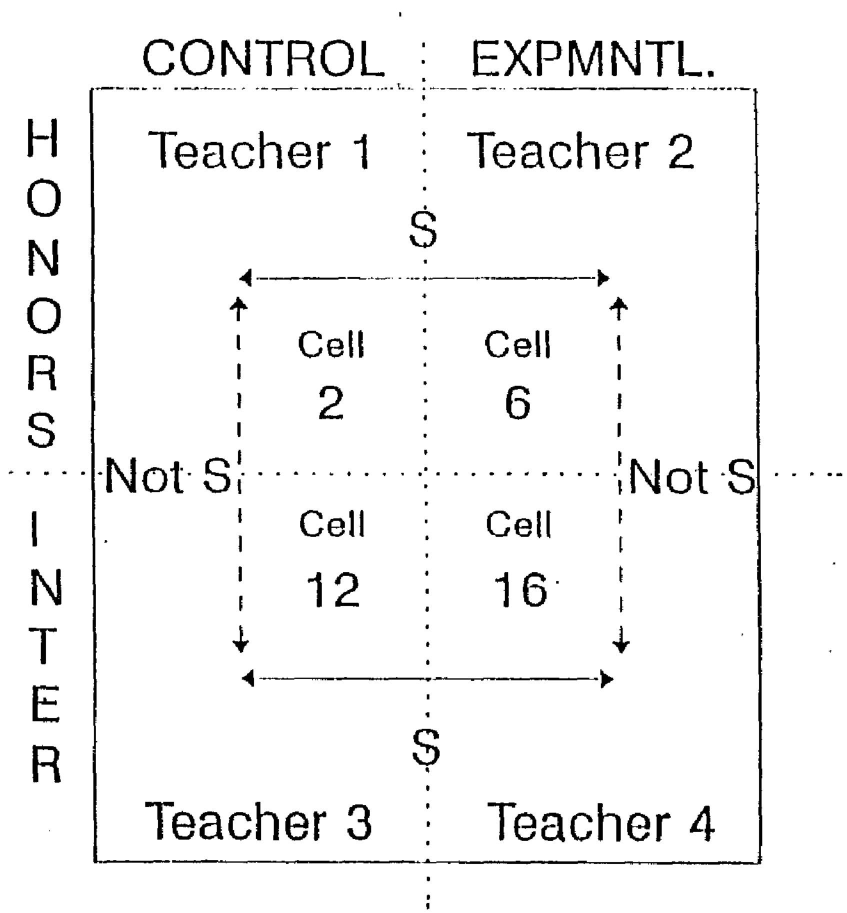
Subset A: 11th Grade Honors and Intermediate Students with 2 Years



Subset B: 11th Grade Honors and 10th Grade Intermediate Students with 2 years of Mathematics



Subset B: 11th Grade Honors and 10th Grade Intermediate students with 2 Years of Mathematics



Subset C: 11th Grade Intermediate Students with 1 or 2 Years

XVI

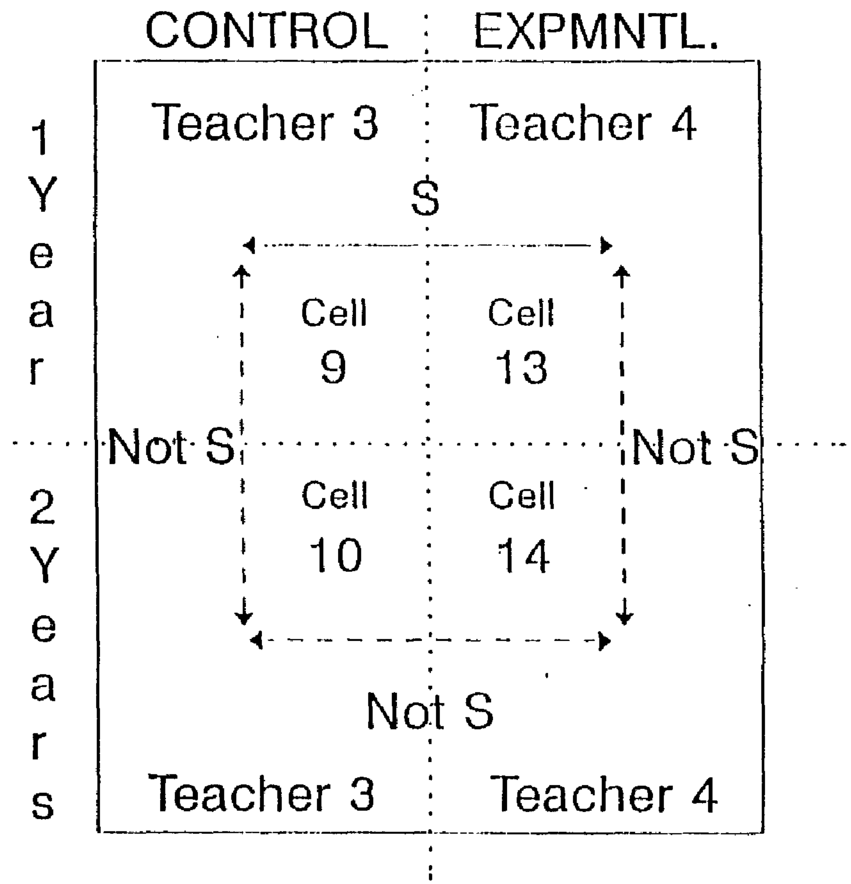
(75)

Teacher Group

9,10 vs 13,14

XVII	VIII	XVIII	XIX
(25) Teacher Group	(50)	(38)	(37)
9 vs 13	10 vs 14	9 vs 10	13 vs 14
1 Year	2 Years	Control	Expmntl.

Subset C: 11th Grade Intermediate Students with 1 or 2 Years



SELECTED BIBLIOGRAPHY

Alderman, D. L. and Powers, D. E. "The Effects of Special Preparation on SAT Verbal Scores." American Educational Research Journal 17 (1980):239-253.

American College Testing Program. Preparing for the ACT Assessment. Iowa City: Iowa, 1991.

Antes, Richard. Preparing Students for Taking Tests. Bloomington, Indiana: Phi Delta Kappa Educational Foundation, 1989.

Archbald, Douglas and Newmann, Fred M. Beyond Standardized Testing: Assessing Authentic Academic Achievement in the Secondary School Reston, VA: National Association of Secondary School Principals, 1988.

Ashworth, Kenneth H. "Standardized Testing: A Defense." The College Board Review (Winter 1989-90):23-24.

Bangert-Drowns, R. L.; Kulik, J. A.; and Kulik, C-L. C. "Effects of Coaching Programs on Achievement Test Performance." Review of Educational Research 53(4) (1983):571-585.

Becker, B. J. "Coaching for the Scholastic Aptitude Test: Further Synthesis and Appraisal." Review of Educational Research 60 (1990):373-417.

Bell, Terrel H. "Reflections One Decade After A Nation At Risk." Phi Delta Kappan 74 (April 1993):592-597.

Benson-Pfiefle, Barbara Louise. "Effects on Achievement Test Scores Resulting From Teaching Test-Taking Skills in the Fifth Grade." Ed. D. dissertation, Loma Linda University, (1987).

Bishop, Judith Anne. "The Effects of Instruction in TestWiseness on Score Improvement on the California Achievement Test Among Third and Sixth Grade Students." Ed. D. dissertation, Northern Arizona University, 1984.

Bracey, Gerald W. "Testing: Some Cautionary Tales. Phi Delta Kappan 72 (November 1991):225.

Braswell, James S. "Changes in the SAT in 1994." The Mathematics Teacher 85(1) (January 1992):16-21.

Boyer, Ernest L. High School: A Report on Secondary Education in America. New York: Harper & Row, 1983.

Burns, P. C. "Intensive Review as a Procedure in Teaching Arithmetic." Elementary School Journal 60 (1970):205-211.

Brown-Burke, Kathleen. "A Model Reading Course and Its Effects on the Verbal Scores of the Eleventh and Twelfth Grade Students on the Nelson-Denny Test, the Preliminary Scholastic Aptitude Test, and the Scholastic Aptitude Test (Coaching, Study Skills, Vocabulary)." Ph. D. dissertation, Georgia State University, 1986.

Coffin, Gregory C. "Computers Can Help Students Improve SAT Scores". NASSP Bulletin 72 (October 1988):78-83.

Cole, Beverly. "College Admissions and Coaching." Negro Educational Review 38 (April-July 1987):125-135.

College Board. APT Guide 1991-92 for High Schools and Colleges. New York: College Board Publications, 1991.

College Board. Taking the SAT, 1991-92. New York: College Entrance Examination Board, 1991.

College Entrance Examination Board, Taking the SAT: A Guide to the Scholastic Aptitude Test and the Test of Standard Written English. New York: ERIC Document Reproduction Service No. ED 203, 1978.

College Entrance Examination Board. Effects of Coaching on Scholastic Aptitude Test Sores. New York: ERIC Document Reproduction Service No. ED 169, 1968.

Comras, Jay. Improving College Admission Test Scores--ACT Mathematics Student Workbook. Reston, Virginia: National Association of Secondary School Principals, 1985.

Crosby, Emeral A. "The 'At Risk' Decade." Phi Delta Kappa 74(8) (April 1993):598-604.

Crouse, James and Trusheim, Dale. "Much Ado about the SAT." Phi Delta Kappan 72 (October 1991):254.

Culhane, Joseph W. "Should Test-Taking Strategies Be Taught?" The Clearing House (November 1982):101-102.

Darling-Hammond, Linda. "Achieving Our Goals: Superficial or Structural Reforms." Phi Delta Kappan 71 (December 1990):286-95.

Darling-Hammond, Linda. "The Implications of Testing Policy for Quality and Equality." Phi Delta Kappan 72

(November 1991):220-225.

Dempster, F. N. "The Spacing Effect: A Case Study in the Failure to Apply the Results of Psychological Research." American Psychologist 43 (1988):627-634.

Edwards, A. S. "The Distribution of Time in Learning Small Amounts of Material." Studies in Psychology: Titchener Commemorative Volume. Worcester, Mass.: Wilson, 1917.

Elmes, D. G.; Dye, C. J.; and Heredelin, N. J. "What is the Role of Affect in the Spacing Effect?" Memory and Cognition 11 (1983):144-151.

Evans, F. R. and Pike L. W. "The Effects of Instruction for Three Mathematics Item Formats." Journal Of Educational Measurement 10 (1973):257-272.

Federal Trade Commission, Bureau of Consumer Protection. Effects of Coaching on Standardized Admission Examinations: Revised Statistical Analyses of Data Gathered By Boston Regional Office of the Federal Trade Commission. Washington, D. C.: Federal Trade Commission, Bureau of Consumer Protection, NTIS No. PB-296, 1979.

Foster, Alan, Rath, James, & Winters, Leslie. Algebra Two and Trigonometry (Columbus: Merrill Publishing Company, 1990).

Fraenkel, Jack R. and Wallen, Norman E. How to Design and Evaluate Research in Education. New York: McGraw-Hill, 1990.

Gallagher, Richard P. "How to Study." Phi Kappa Phi Journal. (Spring 1988):40-41.

Gay, L. R. "Temporal Position of Reviews and Its Effect on the Retention of Mathematical Rules." Journal of Educational Psychology 64 (1973):171-182.

Glass, G. V. "Primary, Secondary, and Meta-analysis of Research." Educational Researcher 5 (1976):3-8.

Good, T. L. and Grouws, D. A. "The Missouri Mathematics Effectiveness Project: An Experimental Study in Fourth-Grade Classroom." Journal of Educational Psychology 71 (1979):355-362.

Goodlad, John I. A Place Called School: Prospects for the Future. New York: McGraw-Hill, 1984.

Gronlund, Norman E. and Linn, Robert L. Measurement and

Evaluation in Teaching 6th ed. New York: Macmillian, 1990.

Hale, Robert L. Mystat Statistical Applications--Dos Version. Cambridge, Massachusetts: Course Technology, Inc., 1990.

Harris, Major L. and Rohfeld, Rae. "SAT/ACT Preparation Program: A Team Approach." National Council on Community Services and Continuing Education (1983).

Hitt-Livingston, Martha Jean. "The Effects of Testwiseness Instruction Using the Soaring High with Test-Taking Tactics--Mathematics Program on the Mathematics Scores of Third-Grade Students on the Stanford Achievement Test-Seventh Edition. Ed. D. The University of Alabama, (1987).

Hughes, Charles Allan. "A Test-Taking Strategy for Emotionally Handicapped and Learning Disabled Adolescents." Ph. D. dissertation, University of Florida, 1985.

Hymel, Glenn M. and Guedry-Hymel, Linda. "Promoting Study Skills and Test-Taking Techniques". NASSP Bulletin 71 (October 1987):97-100.

Jencks, C. and Crouse, J. "Should We Relabel the SAT...Or Replace It?" New Directions for Testing and Measurement: Measurement, Guidance, and Program Improvement, No. 13. San Francisco: Jossey-Bass, 1982.

Jordan, Samuel Jr. Assessment of Standardized Tests Scores and The Black College Environment. Paper presented at Southern Conference on Afro-American Studies. (Jackson, Mississippi: March 27-28, 1987).

Kaagan, S. S. and Coley, R. J. "State Education Indicators: Measured Strides Missing Steps." Princeton, N. J.: Center for Policy Research in Education, 1989.

Kaplan, Stanley H. "Stanley H. Kaplan Tells How to Improve Scores on Standardized Tests." Bottom Line/Personal, 15 February 1990, pp. 11-12.

Kearney, Philip C. "Uses and Abuses of Assessment and Evaluation Data by Policymakers." Educational Measurement: Issues and Practice 23 (Fall 1983):9-12.

Kulik, J. A.; Bangert-Drowns, R. L.; and Kulik, C-L. C. "Effectiveness of Coaching for Aptitude Tests." Psychological Bulletin 95(2) (1984):179-188.

Kulik, J. A.; Kulik, C-L. C.; and Bangert-Drowns, R. L. "Effects of Practice on Aptitude and Achievement Test Scores." American Educational Research Journal 19:(3) (Fall 1982):415-429.

Kulik, J. A.; Kulik, C-L. C.; and Cohen, P. A. "Effectness of Computer-based College Teaching: A Meta-Analysis of Findings." Review of Educational Research 50 (1980):134-135.

Landauer, T. and Bjork, R. "Optimum Rehearsal Patterns and Name Learning." Practical Aspects of Memory. New York: Academic Press, 1978.

Ligon, G. D. and Jones, P. "Preparing Students for Standardized Testing: One District's Perspective." Paper presented at the annual meeting of the American Research Association, April 1, 1982.

Matter, M. K. "Legitimate ways to Prepare Students for Testing: Being Up Front to Protect Your Behind." in Hall, J. and Wolmut, R., eds., National Association of Test Directors 1986 Symposia. Oklahoma City: Oklahoma City Public Schools, 1986.

McCormick, Kathleen. "Cramming for College." Changing Times, September 1987, p. 61, 64.

McDaniel, M. A. and Masson, M. E. J. "Altering Memory Representations Through Retrieval." Journal of Experimental Psychology: Learning, Memory, and Cognition 11 (1985):371-385.

Mehrens, William A. "National Tests and Local Curriculum: Match or Mismatch?" Educational Measurement: Issues and Practice 3 (March 1984):10.

Mehrens, William A. and Kaminski, J. "Methods for Improving Standardized Test Scores: Fruitful, Fruitless, or Fraudulent?" Educational Measurement: Issues and Practices 8 (January 1989):14-22.

Messick, S. "Issues of Effectiveness and Equity in the Coaching Controversy: Implications for Educational and Testing Practice." Educational Psychologist 17 (1982):67-91.

Michigan Department of Education. "Annual Report Carries $ Incentives and Noncompliance Penalty." Michigan Education Report. Lansing, MI: Department of Education, September 1990.

Michigan Department of Education. "Quality Education Package

Will Drive Reform." Michigan Education Report. Lansing, MI: Department of Education, September 1990.

National Commission on Excellence in Education. A Nation At Risk. Washington, D.C.: Government Printing Office, 1983.

National Council of Teachers of Mathematics, Commission on Standards for School Mathematics. Curriculum and Evaluation Standards for School Mathematics. Reston, VA: The Council, 1989.

National School Public Relations Association. School Public Relations: The Complete Book. Arlington, VA: National School Public Relations Association, 1986.

Noble, Candace A. and Mullen, Kenneth B, "The ACT Assessment in the 1990's." The Mathematics Teacher 85(1) (January 1992):22-25.

Nungester, R. J. and Duchastel, P. C. "Testing Versus Review: Effects on Retention." Journal of Educational Psychology 74 (1982):18-22.

Oakes, J. Educational Indicators: A Guide of Policymakers. New Brunswick: Center for Policy Research in Education, 1986.

Popham, James W. "The Merits of Measurement-Driven Curriculum." Phi Delta Kappan 68 (May 1987):679-682.

Porter, Andrew. "External Standards and Good Teaching: The Pros and Cons of Telling Teachers What to Do." Educational Evaluation and Policy Analysis 11 (1989):354.

Rea, C. P. and Modigliani, V. "The Effect of Expanded Versus Massed Practice on the Retention of Multiplication Facts and Spelling Lists." Human Learning 4 (1985):11-18.

Resnick, Daniel and Lauren. "Standards, Curriculum and Performance: A Historical and Comparative Perspective," Educational Researcher (1987):5-20.

Resnick, Lauren B. Education and Learning to Think. Washington, D.C.: National Academy Press, 1987.

Reynolds, J. H. and Glaser, R. "Effects of Repetition and Spaced Review upon Retention of a Complex Learning Task." Journal of Educational Psychology 55 (1964):297-308.

Ritter, Shirley Ann. "Teaching Middle School Students To Use a Test Taking Strategy (Mainstreaming, Learning Strategies, Generalization)." Ph. D. dissertation, University of Illinois at Urbana-Champaign, 1985.

Runyan, Richard P. and Haber, Audrey. Fundamentals of Behavioral Statistics. New York: Mc-Graw Hill, Inc, 1991.

Saxon, John. "Incremental Development: A Breakthrough in Mathematics." Phi Delta Kappan 63 (July 1982):482-484.

Scannell, Dale P., Consulting Editor. National Proficiency Survey Series: Algebra II. Chicago, Illinois: The Riverside Publishing Company, 1989.

Selden, R. W. "Missing Data: A Progress Report from the States." Phi Delta Kappan 69 (July 1988):492-494.

Sesnowitz, Michael; Bernhardt, Kenneth L.; and Knain, D. Matthew. "An Analysis of the Impact of Commercial Test Preparation Courses on SAT Scores." American Education Research Journal 19(3) (Fall 1982):429-441.

Shepard, L. A. and Kreitzer, A. E. "The Texas Teacher Test." Educational Researcher 16 (June 1987):22-31.

Slavin, R. E. Educational Psychology: Theory Into Practice 2nd ed. Englewood, Cliffs, N.J.: Prentice Hall, 1988.

Smith, Frederick L. "Commercial Coaching and SAT Scores: The Effects on College Preparatory Students in Private Schools." Journal of College Admissions (Spring 1989):2-9.

Snedecor, Peter J. "Coaching : Does It Pay-Revisited." Journal of College Admissions (Fall 1989):15-18.

Spitzer, J. F. "Studies in Retention." Journal of Educational Psychology 31 (1939):665-676.

Slack, W. V. and Porter, D. "The Scholastic Aptitude Test: A Critical Appraisal." Harvard Educational Review 50 (1980):154-155.

Spitzer, J. F. "Studies in Retention." Journal of Educational Psychology 30 (1939): 641-656.

The American College Testing Program. The Act Assessment Form 9039K. Iowa City, Iowa: ACT Publications, 1990.

Tyler, Ralph W. "The Impact of External Testing Programs." in Findlay, Warren G., ed. The Impact and Improvement of School Testing Programs. Chicago: University of Chicago Press, 1963.

Urman, Howard Neal. "Ethnic Differences and the Effects of Testwiseness Training on Verbal and Math Achievement." Ph. D. dissertation, University of Southern California, 1982.

U.S. Department of Education. Measuring Up: Questions and Answers about State Roles in Educational Accountability. Report of the OERI State Accountability Study Group. Washington, D.C.: Government Printing Offices, 1988b.

U.S. Department of Education. America 2000: An Educational Strategy. Washington, D.C.: Government Printing Office, 1991.

Vernon, P. E. "Practice and Coaching Effects in Intelligence Tests." Educational Forum 18 (1954):259-280.

Vida, Diane. "A Study of Test-Taking Skills and Achievement Scores Upon Secondary Students." Ed. D. dissertation, Drake University, 1985.

Weaver, Claire Gunn. "A Study of Instructional Techniques to Prepare for the ACT Test." Ph. D. dissertation, Southern Illinois University at Carbondale, 1988.

Wigdor, Alexandra K. and Garner, Wendell R., eds. Ability Testing: Uses, Consequences, and Controversies. Washington, D.C.: National Academy Press, 1987.

ABSTRACT

THE EFFECTS OF TEACHING PRACTICE REVIEW ITEMS
AND TEST-TAKING STRATEGIES
ON
THE ACT MATHEMATICS SCORES OF
SECOND-YEAR ALGEBRA STUDENTS

by

PATRICIA KOVACH MC MANN

May, 1994

Advisor: Dr. Edward Simpkins

Major: Administration and Supervision-General

Degree: Doctor of Education

This pretest/posttest study was designed to measure the effect of using practice mathematics items and test preparation activities on the American College Testing Program (ACT) scores of students enrolled in a second-year algebra class. The total number of students in the study was 196 (97 control group, 99 experimental). Students in the experimental group were given spaced practice of sample ACT mathematics questions and test-preparation suggestions for the mathematics section of the ACT. The primary purpose of this study was to determine whether student scores on the mathematics portion of the ACT test could be improved by practicing sample test items and studying general test-taking strategies. There were three secondary purposes included in this study. One was to determine whether a student's gender had any significance in ACT performance, the second was to determine whether the number of years of mathematics study affected achievement, and the third

purpose was to examine the role of the student's prior achievement level.

The students in the study were compared on the dependent variables of years of mathematics (one or two), class (tenth or eleventh), level (honors or intermediate), teacher (1,2,3, or 4), gender (male or female), and group (control or experimental).

The study found that students in the experimental group who received treatment outscored their counterparts in the control group. Gender did not show significance in the students ability to learn or benefit from coaching practice or test-taking suggestions. The prior number of years of mathematics taken showed no significance. Students in the honor section showed more improvement than their average college preparatory counterparts. The number of years of prior mathematics did not show significance.

The detailed analysis of the study can be found in the text.

AUTOBIOGRAPHICAL STATEMENT

NAME Patricia Kovach Mc Mann

EDUCATION

Ed. D.	Wayne State University, Detroit, Michigan-1994 Major: General Administration and Supervision
Ed. Sp.	Wayne State University, Detroit, Michigan-1990 Major: General Administration and Supervision
M. Ed.	Wayne State University, Detroit, Michigan-1969 Major: Mathematics Education
B. S.	California University of Pennsylvania-1967 California, Pennsylvania Major: Secondary Education-Mathematics

PROFESSIONAL EXPERIENCE

1975-1994	Department Chairperson, Mathematics Roosevelt High School Wyandotte, Michigan
1971-1994	High School Mathematics Teacher Roosevelt High School Wyandotte, Michigan
1967-1970	Junior High School Teacher Lincoln Junior High School Wyandotte, Michigan
1969-1979	Adjunct Instructor Wayne County Community College Taylor, Michigan

PROFESSIONAL ORGANIZATIONS

Association for Supervision and Curriculum Development
Detroit Area Council Teachers of Mathematics - Rec. Sec.
Michigan Association of Mathematics Supervisors
Michigan Council of Teachers of Mathematics
Michigan Association of Computer Users in Learning
National Association of Secondary School Principals
National Council of Teachers of Mathematics
National Council of Supervisors of Mathematics
Phi Delta Kappan, Wayne State University Chapter