4-1993

The Impact of Test Sophistication on Tenth Grade Students’ Michigan Educational Assessment Program (MEAP) Scores

Iris Williams
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Educational Assessment, Evaluation, and Research Commons, and the Educational Methods Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/1987

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
THE IMPACT OF TEST SOPHISTICATION ON TENTH GRADE
STUDENTS' MICHIGAN EDUCATIONAL ASSESSMENT
PROGRAM (MEAP) SCORES

by

Iris Williams

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Education
Department of Educational Leadership

Western Michigan University
Kalamazoo, Michigan
April 1993
THE IMPACT OF TEST SOPHISTICATION ON TENTH GRADE STUDENTS' MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM (MEAP) SCORES

Iris Williams, Ed.D.
Western Michigan University, 1993

The primary purpose of this study was to determine whether test sophistication training would result in more tenth grade students achieving performance levels on the Michigan Educational Assessment Program (MEAP) subtests in mathematics and reading which would entitle them to receive the state endorsed diploma.

Two groups of English II students (grade 10) from the same school participated in the study. One group received test sophistication instruction, and the other received no treatment. During the test sophistication training, students were instructed using the Ace That Test Preparation for Success Student Guide (Yarosz & Yarosz, 1991a) which was based on the areas of test wisdom that had been identified in the research as follows: motivation, time-on-task, learning by doing, reinforcement of learning, sharing of common problems and solutions by peers, and developing competence as a means of building confidence.

Secondary goals of the study were to determine if gender-related differences in achievement existed, and if there was an interaction effect between the type of group (instructed or not instructed) and gender, on achievement and
self-reports of how the students performed on the reading test.

A sample of 46 tenth grade students from a rural Michigan school district was selected for this study. Students were members of two tenth grade English II classes, which were determined to be equivalent on measures of grade point average, the Verbal Reasoning subtest of the Differential Aptitude Test (1982), and the pre-test of the Ace That Test test sophistication training program. Both groups were given the Ace That Test Form EY post-test, and also took the MEAP tests of mathematics, story reading and informational reading. Both also completed the self-reports of performance, effort and interest on the story and informational reading tests.

The data analysis showed that students who experienced the test sophistication training scored significantly higher ($p < .05$) on the post-test of test sophistication than those students who had not experienced the training. No other significant results were found.

The findings of this study suggest that the MEAP test may not be subject to the influence of test sophistication training.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700  800/521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
The impact of test sophistication on tenth-grade students' Michigan Educational Assessment Program (MEAP) scores

Williams, Iris, Ed.D.
Western Michigan University, 1993
ACKNOWLEDGEMENTS

I want to express my sincere appreciation to Dr. Eugene W. Thompson, chair of my doctoral committee, for the time, expertise, support and encouragement provided throughout the phases of this dissertation. I also want to thank the other members of my doctoral committee, Dr. Zoe Barley, for her committed guidance and encouragement, and Dr. Terre Davis, for her support and valuable ideas in the development of this investigation.

My gratitude goes to the members of the Martin Public Schools Board of Education for the proactive role they took in encouraging me, and in valuing what might be learned in the process. In addition, I am deeply grateful to Cyndee DeGraaf, Dave Pearson, June DeForest, Fred Bogdan and Carol Decker who helped in the implementation of this dissertation project.

I am also grateful to my son, Jim, for his loving encouragement and inspiration, and to my friends for caring.

Iris Williams
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................ ii  
LIST OF TABLES ............................................................................................................. vii  

CHAPTER  
I. INTRODUCTION ........................................................................................................... 1  
   The Problem ................................................................................................................ 1  
   General Statement ...................................................................................................... 4  
   Specific Questions ....................................................................................................... 5  
   Definition of Terms ..................................................................................................... 5  
   Hypotheses ................................................................................................................ 7  
   Delimitations ............................................................................................................. 9  
II. A REVIEW OF THE LITERATURE ............................................................................. 11  
   Achievement Tests and Test Sophistication ............................................................... 11  
   Time-on-Task ............................................................................................................. 15  
   Practice/Learning by Doing/Reinforcement ............................................................... 15  
   Confidence and Competence Building/Anxiety Reduction ........................................ 16  
   Motivation of Students ............................................................................................. 17  
   Test Sophistication and Criterion-Referenced Tests ................................................. 20  
   Achievement and Test Score Differences  
   Between Boys and Girls ............................................................................................. 21  
III. RESEARCH DESIGN AND METHODOLOGY ......................................................... 24
Table of Contents–Continued

CHAPTER

Research Design................................................................. 24
Sampling Procedures............................................................... 24
Threats to Validity........................................................................... 26
  Internal Threats to Validity...................................................... 27
  External Threats to Validity..................................................... 27
Effect Size......................................................................................... 30
The Changing Perception of Validity............................................. 32
Instruments...................................................................................... 33
  The Ace That Test Pre-Test, Instructional
  Program and Post-Test.............................................................. 33
  Reliability............................................................................. 38
  Ace That Test Instructional Program........................................ 39
  The Michigan Educational Assessment Program............... 40
Teacher............................................................................................. 44
Experimental Procedures.......................................................... 44
Analysis of Data............................................................................... 45

IV. FINDINGS.............................................................................................. 48
Equivalence of Groups.................................................................. 50
  Hypothesis 1: Type of Group and Test
  Sophistication Achievement...................................................... 51

iv

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table of Contents—Continued

CHAPTER

Hypothesis 2: Type of Group and Achievement on a Criterion-Referenced Test.......................................................... 53

Hypothesis 3: Type of Group and Perceptions About Testing...................................................................................... 55

Hypothesis 4: Type of Group and Proportion Qualifying for Diploma Endorsement.................................................. 57

Threats to Validity......................................................................................................................................................... 59

Results of Hypotheses Concerning Gender.................................................................................................................. 62

Hypothesis 5: Type of Group, Gender, and Post-Test.................................................................................................. 63

Hypothesis 6: Type of Group, Gender, and MEAP Math Test..................................................................................... 63

Hypothesis 7: Type of Group, Gender, and MEAP Story Reading Test....................................................................... 63

Hypothesis 8: Type of Group, Gender, and MEAP Information Reading Test.............................................................. 64

Hypothesis 9: Type of Group, Gender, and Self-Report on Story Reading................................................................. 65

Hypothesis 10: Type of Group, Gender, and MEAP Self-Reports on the Informational Reading Test..................... 65

Discussion of Findings Concerning Gender, Hypothesis 5 Through Hypothesis 10.................................................... 66

Effect Size................................................................................................................................................................. 67

Summary of Findings.................................................................................................................................................... 68

V. DISCUSSION AND RECOMMENDATIONS............................................................................................................. 70
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>70</td>
</tr>
<tr>
<td>Interpretation of the Results</td>
<td>71</td>
</tr>
<tr>
<td>Limitations of the Study</td>
<td>74</td>
</tr>
<tr>
<td>Implications of the Findings</td>
<td>75</td>
</tr>
<tr>
<td>Recommendations</td>
<td>77</td>
</tr>
<tr>
<td>General Recommendations</td>
<td>79</td>
</tr>
<tr>
<td>Future Study</td>
<td>79</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>APPENDICES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A. MEAP Administration Manual Information</td>
<td>81</td>
</tr>
<tr>
<td>B. Ace That Test Informational Materials</td>
<td>104</td>
</tr>
<tr>
<td>C. Parental Consent Form</td>
<td>113</td>
</tr>
<tr>
<td>D. Human Subjects Institutional Review</td>
<td></td>
</tr>
<tr>
<td>Board Letter of Approval</td>
<td>115</td>
</tr>
<tr>
<td>E. MEAP Copyright Permission Letter</td>
<td>117</td>
</tr>
<tr>
<td>F. Ace That Test Copyright Permission Letter</td>
<td>119</td>
</tr>
</tbody>
</table>

| BIBLIOGRAPHY                                           | 121  |
LIST OF TABLES

1. Comparability of Experimental and Control Groups on DAT-VR Scores, *Ace That Test* Pre-Test Scores, and GPA ................................................................. 25

2. Type of Group and Test Sophistication Achievement on *Ace That Test* Pre- and Post-Test Scores .................................................... 52

3. Type of Group and Scores on MEAF Math, Story Reading and Informational Reading Tests .................................................. 54

4. Type of Group and MEAP Self-Reports of Performance, Effort and Interest on Story and Informational Reading Tests ........................................... 56

5. Type of Group and MEAP Math, Reading: Diploma Endorsement .................................................. 58

6. ANOVA Results of Interaction Between Gender and Type of Group .................................................. 64
CHAPTER I

INTRODUCTION

The Problem

The continuing emphasis on improving student learning in America has resulted in both state and national movements to mandate student testing as a measure of the adequacy of instruction. As these efforts continue, testing procedures are a major issue in the controversy surrounding the significance and reliability of test scores. In the early seventies, the argument was documented as having two sides. Gregory R. Anrig, President of the Educational Testing Service in Princeton, N. J., described the debate (Anrig, 1992). On one side are those who believe that because tests are unfairly biased, they do not measure the true ability of an individual. On the other side are those who believe that tests can be considered valid measures of academic achievement and learning ability due to their standardization. In spite of this controversy, the use of standardized tests not only continues nationally, but is on the rise. A 1990 study released by the National Center on Education and the Economy at Rochester, N. Y., recommends that a national examination system be created. Under the proposed system, students would be expected to achieve a mastery certificate based on their ability to meet world class standards at age sixteen. Prototypes for such a system
are being field tested by The New Standards Project, underwritten by $2.5 million from the Pew Charitable Trusts, and the John D. and Catherine T. MacArthur Foundation (Wolk, 1992, p.3).

Standardized testing is also increasing in some states. Michigan is a case in point. The Michigan Educational Assessment Program (MEAP), which was instituted in January of 1970, was intended to provide information on the status and progress of Michigan basic skills education to the State Board of Education. The MEAP test is a criterion-referenced test which assesses student performance on a basic set of objectives. Students are provided scaled scores which are rated as satisfactory or needing improvement, as well as raw scores achieved for each objective tested. The scores do not compare one student to another student; rather the scores assess the extent to which a given student demonstrates mastery of the objectives.

When the MEAP testing program was started in 1970, students in grades four, seven and ten were tested in math and reading. Significant changes to this original purpose had occurred by 1991. In addition to being used to provide information to the State Board of Education, the data are being publicly released, and the subjects tested expanded to include science at grades five, eight and eleven (Michigan State Board of Education, 1991). Tests for other subjects are under development.

Further emphasizing the importance of testing, Michigan's Legislature, in the 1991-92 state aid bill, mandated achievement of a score of 50 percent or
higher for all Michigan students commencing with the class of 1994, in order to receive a state-endorsed diploma.

The criteria used by the state to evaluate a student's qualification for a state endorsed diploma at the tenth grade level is a raw score of 50% or more items correct in math and reading, (at the eleventh grade level the students must achieve 50% of the items on the science test). All three tests must be passed to achieve the endorsed diploma, for the classes of '94, '95, and '96. Students not qualifying may receive a non-endorsed diploma. Beginning with the class of '97, however, only state-endorsed diplomas may be issued, and students must pass the required state test to qualify.

At the tenth grade level, students take the math and reading tests. The math test is composed of 120 items, therefore 60 items must be answered correctly to meet the required scaled score of 500. In reading there are two 20-question tests of constructing meaning from story and information selections. A scaled score of 300 or more on one of the two tests (16 items correct) is required to qualify for the state endorsed diploma. The science requirement comes into play for eleventh grade students.

Because of the thrust of not only continuing, but also increasing, emphasis on standardized testing, it seems sensible that educators would accept the fact of testing, and find ways to prepare students for test achievement that closely reflects their real level of performance. The achievement of higher scores would prepare students for greater academic achievement by reinforcing their success.
Throughout this paper, the terms test sophistication, test-wiseness and test-taking skills are used synonymously.

Some schools have already moved into the arena of teaching test sophistication, or test-wiseness skills to their students. If higher scores are gained by these students, an even bigger chasm in the measurement of student achievement may be created between those students whose schools provide little or inadequate test-taking skills instruction, and students who receive training.

Test sophistication, or test-wiseness has been considered by some to be a variable in achievement test scores, and research has validated its impact. There is evidence that test-taking skills can be learned and retained. A meta-analysis, a quantitative synthesis of findings of thirty studies on teaching test-taking skills to students was conducted by Bangert-Drowns, Kulik, and Kulik (1983). They found that test sophistication training had a positive effect, with a mean effect size of .25 on test performance in twenty five of the studies. Two years later, results of a second meta-analysis on twenty four studies of test-wiseness training programs indicated a mean effect size of .33, again suggesting that test-wiseness can be taught, and can increase achievement test scores (Samson, 1985).

General Statement

This study proposes to assess the influence of test-wiseness training on standardized test scores of tenth grade students of both sexes. Specifically, this study seeks to explore relationships between training in test sophistication and student
performance and student self-reports about how they did on the Michigan Educational Assessment Program (MEAP).

Specific Questions

1. Can test wiseness be learned, and test behavior be modified by group instruction in test-taking skills?

2. Do subjects who are informed that they must pass the MEAP tests in order to receive a state endorsed diploma, perform at a higher success rate when they are provided test-wiseness training than do students who are so informed, but who are not provided test-wiseness training?

3. Does classroom instruction in test sophistication training influence student attitudes and self-perceptions in the assessment of reading skills?

4. Are there differences between males and females, in their Self Reports of Performance, Effort and Interest on the MEAP Reading test, on the MEAP Reading and Math tests, and on Ace That Test pre- and post-test scores? (The Ace That Test materials are a newly available, commercial product intended for use at the secondary and post-secondary level, to bring about test sophistication. The pre- and post-tests measure a student’s knowledge of test-taking skills.)

Definition of Terms

Test sophistication, also referred to as test-wiseness, and test-taking skills, is defined as the individual’s capacity to maximize his/her utilization of the
characteristics and formats of the tests and/or test-taking situation to receive a higher score.

Student attitudes and self-perceptions is defined as the individual's capacity to identify and report their level of performance, effort or interest, when reading selections and answering questions with regard to performance on the MEAP reading test.

Effect size is defined as the magnitude of difference or relationship in the sample or population. It is the mean standardized difference between treatment and control groups, expressed in standard deviations. Cohen's (1965) definition of effect size is the degree to which a phenomenon exists, and he further categorizes effect sizes as small (.25), medium (.50) and large (1.0), while acknowledging that such classifications are ambiguous.

For this study, it is important to distinguish between the statistical significance of the findings and their practical significance which is indicated by effect size. The power of the test statistic with a sample of this small size may not produce statistical significance, but the effect size may indicate an important difference in groups. Effect size is most important in a high stakes testing environment, where, if even one more student qualifies for the state-endorsed diploma the outcome is highly meaningful to that single individual. The arbitrariness of cut scores, where the student's response to a single question can impact upon the his or her future opportunities becomes a poignant concern. Assuming a normal distribution with the cut score as the mean, an effect size of .25 would result in
10% more students passing.

Cook and Campbell (1979) and Borg and Gall (1983) support this line of reasoning with regard to effect size. Cook and Campbell (1979, p.46) address the problem of specifying magnitudes. They argue the importance of being conscious of the issue in the design phase of the research. What is needed is discussion of what level of effect is required, given the sample size of the research, to conclude that the treatment has made a significant practical difference.

Borg and Gall (1983, p. 380) define effect size as the magnitude of difference or relationship in a sample or population, and discuss two options for increasing the power of the chosen statistical significance test: increase sample size, or establish a higher probability level for rejecting the null hypothesis. Because it was not feasible to increase the sample size in this study, an alpha level of .05 is used. For purposes of this study, the concern is not statistical significance, but effect size. While the investigator considered the findings of the t test for statistical significance, effect size was also considered. If, for example, an effect size near .25 was found, the practical importance of that difference would be relevant to discussion. Conversely, if an insignificant effect size resulted, this finding would support the absence of statistical significance.

Hypotheses

The ten research hypotheses, stated as alternative hypotheses were as follows:
1. Students who experience test sophistication training score higher on a post-test of test-wisness skills than students not experiencing the training.

2. Students who experience test sophistication training score higher on the MEAP Math, Story Reading and Informational Reading subtests than students not experiencing the training.

3. The proportion of students who score high on the MEAP Self-report of Performance, Effort and Interest on story reading and informational reading is greater for students who experienced the test sophistication training than for students not experiencing the training.

4. The proportion of students who achieve the state endorsement in math and reading, is higher for students who experience the test sophistication training than for students who did not experience the training.

5. There is an interaction between student gender, and learning from the experience of test sophistication training, as measured by achievement on the Ace That Test post-test.

6. There is an interaction between student gender, and learning from the experience of test sophistication training, as measured by achievement on the MEAP mathematics test.

7. There is an interaction between student gender, and learning from the experience of test sophistication training, as measured by achievement on the MEAP story reading test.

8. There is an interaction between student gender, and learning from the
experience of test sophistication training, as measured by achievement on the MEAP informational reading test.

9. There is an interaction between student gender, and learning from the experience of test sophistication training, as measured by achievement on the MEAP Student's Self-report of performance, effort and interest on the story reading test.

10. There is an interaction between student gender, and learning from the experience of test sophistication training, as measured by achievement on the MEAP Students' Self-report of performance, effort and interest on the informational reading test.

Delimitations

The results of this study are limited in generalizability by the nature of its population. Subjects were persons who were in attendance at a rural Michigan public school, with a population which is 99 percent Caucasian. The setting is only typical of settings where large percentages of the population are Caucasian and rural in culture, as well as living conditions, and in their choice of public school attendance. It is unknown what effects attendance at such a rural, public, K-12, school (875 students) had on student attitudes and self-perceptions, and whether these would be different for the general population. The results are also limited to the Ace That Test (Yarosz & Yarosz, 1991b) test-wiseness training program.
In the decade of the sixties, researchers began the study of the process of test-taking, and the skills required to do it successfully. Their findings, and those of researchers who followed in subsequent decades, have been extensively reported in the professional literature. Test-wiseness principles and strategies have been identified, and differences between low and high achievers have been studied. Key findings are reported in the review of literature which follows in Chapter II.
CHAPTER II

A REVIEW OF THE LITERATURE

A number of variables, including motivation of pupils, time-on-task, learning by doing, reinforcement of learning, sharing of common problems and solutions by peers, and developing competence as a means of building confidence, are included in the test sophistication training program. The central concepts and related literature follow.

Achievement Tests and Test Sophistication

Considerable study has been done over the past three decades. Millman, Bishop, and Ebel (1965) surveyed 240 high school students using an unstructured questionnaire. They found that when the students were asked to make suggestions which would help a newly arrived U.S. student succeed on tests, they identified principles of test-wiseness: reading directions carefully, spending appropriate time on any one question, checking answers and guessing were considered most important. They also suggested planning their time, using a process of elimination, and using other questions for leads. These researchers studied a limited concept of test-wiseness, by excluding factors which emerged later in the literature: attitudes of anxiety or confidence, motivation, learning by doing (practice), reinforcement, time-on-task, etc., and restricted their analysis to the mechanics
of taking objective tests.

Their view of test-wiseness as logically independent of the student's knowledge of the subject matter which the test items purport to measure, has been a consistent view throughout the development of the literature on the topic. Millman et al. (1965) raise questions for examination by future researchers: (a) what is the importance of test-wiseness, (b) to what degree can it be taught or measured, and (c) is it a substantial source of test variance?

College students cited test-wiseness as an important reason for success on examinations. Gaier (1962) asked 276 college students to assume that they achieved a grade of A or D on a test, and to identify what factors contributed to their A or D grade. The A group (136 students) indicated that understanding the test, the test characteristics, and comprehension and reasoning ability were factors contributing to their successful performance. The D group (140 students) identified not being able to understand and to reason, not understanding tests, and test characteristics as reasons for their unsuccessful performance.

Problem-solving styles of students were also studied with regard to test-wiseness, and possible impact upon test scores. Forty college students in two schools were interviewed as they took regular course examinations and were asked to explain why they chose the answers they did. As a result the problem solving styles of high and low test achievers were separated. Both Bloom and Broder (1950) and French (1965) identified that general problem solving techniques such as ability to reason logically, understanding test directions, and understanding the
nature of the test questions, were used by high performers.

The outline of test-wiseness principles cited by Millman et al. (1965) is divided into two areas. First, elements independent of test purpose include time-using strategy, error-avoidance strategy, guessing strategy, and deductive reasoning strategy. Elements which depend on the test purpose include intent consideration strategy and cue-using strategy.

In 1968, Wahlstrom worked with ninth graders, and hypothesized that instruction in test-taking strategy would enable higher scores than would be achieved by an equally matched examinee not instructed. He constructed two fifty-four item social studies tests of comparable difficulty, and used results of a DAT Verbal Reasoning Test to match groups of students by ability. Three groups were created, with the experimental group receiving 100 minutes of test-wiseness instruction, while the control group discussed occupations and the placebo group watched TV. The experimental group scored considerably higher in post-test exams while control and placebo groups showed little change.

Klutch, in 1976, studied test-wiseness instruction in concert with reduction of test anxiety and raised expectations of success. Her study was reflective of the movement by researchers to expand the study of test sophistication from the cognitive to the affective domains. Participants were eighty eighth-graders in an urban parochial school. Students were divided into three groups: twenty-nine received test sophistication training, twenty-six received unstructured counseling and twenty-five were not counseled. Assessment instruments were the DAT.
Abstract Reasoning Test, forms L and M, and a student questionnaire designed to elicit participants' self-expressed anxieties and inadequacies concerning test taking, and their evaluation and expectations of the results of tests as measures of their real ability. Trained counselors were used to conduct either test-sophistication or unstructured group counseling. Students were pre-tested, then randomly assigned to one of the three groups, which then met for five sessions, once each week, and post-tested on the DAT Abstract Reasoning Test form M, and post questionnaire.

Test sophistication participants followed a structured format of training in test-taking methods, attitudes and anxieties including practice exercises for cognitive materials, and relaxation exercises for anxiety reduction. Unstructured groups were counseled with a non-directive approach of group centered discussions. The noncounseled group received only the pre- and post-tests and questionnaires.

Klutch concluded on the basis of data analysis that test sophistication can be taught and learned in a group counseling setting, that test sophistication training leads to achieving higher scores on standardized tests, and that self-expressed anxiety is reduced with test sophistication training, and that sex of subject makes no significant difference in these areas. It is noted that the amount of time spent in these activities is not known (was it one hour a day for a total of five hours, or two hours daily for a total of ten?). Were it known, an even clearer picture of the findings could be gained.
Time-on-Task

In two separate meta-analyses which included fifty four individual studies, both Samson (1985) and Bangert-Drowns, Kulik, and Kulik (1983) found that the amount of time devoted to test taking skills affects test scores. They found that positive gains for time spent on training lie within the ten to twenty hour range, with an increase of fourteen percentile points at ten hours, and an increase of seventeen percentile points at twenty hours instruction. Thirty hours of instruction in test-taking skills achieved only two additional percentile points.

Practice/Learning by Doing/Reinforcement

Regardless of the descriptors used by various investigators, they reported repeatedly that students who practice test-taking skills by doing what is required on the tests, and who receive reinforcement for doing it, achieve positive results. Sarnacki (1979) cited a number of studies which demonstrate the positive effects of test-wiseness training, advocating specific attention be given to lack of practice opportunities. McLellan and Craig (1989) reviewed the literature and identified strategies which could be used by the classroom teacher, based upon their successful use prior. They include: (a) practice forums; (b) taking sample tests in the formal test setting; (c) spaced, as opposed to long periods of review; and (d) completion of a checklist of their existing test-taking habits to help them learn their individual test-taking profile. In addition a variety of mechanical techniques are
specifically identified: (a) familiarize students with directions, format and answer sheets in advance; (b) remind students to read carefully, pace themselves; (c) alert them to commonly used relationship patterns like same, type of, comes before, is the cause of, is more than; and (d) alert them to cues, to use deductive reasoning, and to make shrewd guesses, never skipping an answer.

Confidence and Competence Building/Anxiety Reduction

McLellan and Craig’s 1989 meta-analysis identified the importance of the teacher creating a comfortable and supportive atmosphere which reassures students, and the safeguarding of students’ emotional well-being. Sarnacki (1979) recommended that attention be given to test anxiety and to students’ lack of confidence in their own ability. Test-taking preparation was found advantageous on students’ level of anxiety, self-esteem and attitudes towards testing (Scruggs, White & Bennion, 1986). According to Wilson and Ritter (1986) approximately one in five students is subject to the debilitating effects of test anxiety, so test-taking preparation activities are beneficial psychologically to twenty percent of students.

Samson (1985) and Bangert-Drowns et al. (1983) both conclude that if an average student performs at the 50th percentile on a test, that same student would likely score at over the 60th percentile if exposed to test-taking skills. Similarly, they found low ability students to benefit in particular. When low, medium and high ability students have received the same training, low ability
students have made higher gains (Berliner, 1986). In another study, low socio-economic students received higher levels of treatment than other students, and benefitted more than twice as much according to Scruggs et al. (1986). This underscores the psychological advantage in the positive relationship between improvement in academic achievement and students’ self-concepts (Crittenden, 1984).

These studies have indicated test-taking skills have the following advantages to students: (a) the twenty percent who suffer psychologically may have some anxiety relieved, (b) low ability students make significant gains, and (c) students who make gains in achievement due to test preparedness may increase in self-esteem.

Motivation of Students

Researchers at the University of Michigan surveyed nearly 1,000 students in grades 2 to 11 using a 5-point scale where they agreed or disagreed with each statement (1 = strongly agree; 5 = strongly disagree) about standardized testing (Paris, Lawton, Turner & Roth, 1991). They found that students have a growing disillusionment about standardized testing, which increases over time and experience, and which is clearly apparent by high school. By this time, they report that students are skeptical that scores reflect intelligence, indicate good students, or indicate effective teachers. While younger students did not question the value of testing, older students question the purposes and uses of achievement tests.
A second trend reported by students is decreasing motivation to excel which the authors suggest may indicate an attempt by students who don't do well, to protect their own self-esteem by being able to claim that they didn't try.

A third trend observed was the assertion among older students that they felt less prepared to do well on tests. Despite years of repeated standardized testing, older students indicated they felt they had few appropriate strategies for testing. When compared to younger students, older students reported choosing negative testing behaviors like filling in the bubbles without reading the questions, or just guessing on confusing questions. They indicated a greater likelihood of cheating, becoming nervous, having difficulty concentrating, and looking for answers that matched the questions without reading the passage. Such strategies negate personal responsibility for outcomes, and avoid personal effort, thereby undermining motivation.

In a separate survey by these same researchers (Paris, Lawton, Turner & Roth, 1991) 250 Michigan students in Grades 4, 7, and 10 were surveyed. Shortly following taking the MEAP test, they were asked to agree or disagree with a series of statements about the test. Most students indicated they tried hard, and thought they did well. They indicated the test was not difficult, and there was little cheating. However, tenth grade students in particular reported using the negative strategies listed above, consistent with the results of the authors' earlier surveys. This group of 250 students was divided into a sample of high and low achievers, and examined to see if they reported differently. High achievers gave
positive responses. Low achievers reported filling in the bubbles without thinking, getting bored, and not finishing reading the passage. Thus positive self-perceptions, appropriate strategies and persistence separated high and low achievers.

This finding makes more imperative a recognition of differences that may have come to exist between schools whose students are trained in test sophistication, and those who are not. The fact that test sophistication programs are not standard or uniform, and are therefore elusive when researchers have attempted to define them confounds this understanding, and is underscored by Becker's exhaustive meta-analysis of studies reporting on coaching for the SAT.

Becker (Fall, 1990) found a number of variables which were inconsistently studied and reported in the literature on test sophistication with regard to the SAT. She examined such factors as definitions of coaching, content of coaching interventions, duration of coaching activities, quality of prior coaching studies, motivation of sample populations, selectivity of sample populations, gains due to retesting, and differential effects of coaching on SAT math and SAT verbal tests. Her conclusions are instructive to future investigation: (a) she concludes that the studies of coaching effectiveness have widely varying results; (b) that the variation in results arises from studies without comparison groups; (c) that there are two factors which relate to the magnitude of the found coaching effects: (1) study design and (2) study duration; and (d) that the set of coaching studies to date confound the features of their design and of the interventions used.
Because the studies reviewed in the various meta-analyses mentioned above were limited, results are applicable only to the particular achievement tests and intervention programs used. Before educators can use various intervention strategies with confidence, or decide not to use them at all, more research is needed on the extensively available, commercially prepared programs.

Deaton, Halpin and Alford studied the Scoring High program, which is presented as if it will have impact upon student achievement on the California Achievement Tests (Deaton, Halpin & Alford, 1987). A total of 925 students were divided into experimental and control groups in grades 1, 2, 4, and 5. The experimental groups were given the Scoring High program, while no instruction in test-taking skills was provided the control groups. Results of the comparisons between groups suggest that the program did not produce consistent increases in student scores.

The Scoring High program provides instruction and practice in areas tested by the CAT--reading, language, and mathematics--as well as test-taking skills and strategies. Effects differed with the particular subtests of the CAT analyzed. Reading comprehension and language expression scores were not affected by the coaching program, while mathematics and reading vocabulary were associated with differences attributable to the coaching program.

Test Sophistication and Criterion-Referenced Tests

A review of the literature related to the effects of test-wiseness instruction
on student performance on criterion-referenced tests resulted in little information. The literature which does exist focuses in areas other than the effects of test sophistication on student achievement. Articles abound which describe the criterion-referenced tests used by various states; others report on such questions as the effect on performance of the student's prior knowledge of content, the ability of teachers to predict student performance, and the effects of allowing students more time to take tests. It appears that while many programs of criterion referenced testing have been developed, they have not been evaluated with regard to test-taking skills. Many "home remedies" for increasing student performance are described.

Achievement and Test Score Differences Between Boys and Girls

What is the impact of gender on the achievement of students? The American Association of University Women conducted a study in which school-age boys and girls were surveyed. Titled "Shortchanging Girls, Shortchanging America," Avasthi (1991) reports the study indicated that girls enter high school with less self-confidence, fewer dreams for the future, and less faith in their abilities to succeed than do boys. While concluding that both boys and girls suffer from low self-esteem levels in adolescence, girls emerge from this period with markedly lower self-worth.

According to the report, girls begin to feel less competent in school, particularly with regard to math and science, with a drop in self-esteem becoming
noticeable between the sexes starting around age ten. Girls' ratings of the extent to which they are good at things drop from 45% in elementary school, to 29% in middle school, and further to 23% in high school. The AAUW study found that the decline for boys goes from 55% in elementary school to 48% in middle school, and to 42% in high school. The report further indicates that as girls slowly lose self-confidence with age, they worry about appearance, spend more time on hair and makeup than homework, and expect to be homemakers, not scientists or mathematicians.

Is there reason to believe that there is a biological difference between boys and girls that causes one gender or the other to be superior in any particular academic area such as mathematics or language? Not according to a review of the literature reported by Good and Brophy (1986). They report that research in intellectual functioning does not reveal consistent differences, thus, there is no reason to think that boys and girls cannot succeed equally well in various fields of study or employment. They found that what differences can be indicated are 'trivial when measured, and that performance differences are likely due to differences in sex-role expectations. They support this conclusion by citing various studies on boys' reading performance, which differed greatly from one culture to another, with similar findings across studies of girls and performance in mathematics. Good and Brophy (1986) assert that in spite of girls' early advantage over boys in elementary grades, the fact that over time, their achievement falls progressively farther behind that of boys, indicates the gradual change in the relationship
between gender role and student role.

The topic of test sophistication has been studied for the past three decades, and continues to be an issue for educators and students alike. Because of continued pressure on schools to increase student performance on state-mandated tests, it is likely that the study of ways to improve student test scores will continue as well. The present study is intended to contribute to the discussion as it continues to evolve.
CHAPTER III

RESEARCH DESIGN AND METHODOLOGY

Research Design

This study uses a quasi-experimental pre-test, post-test, control group design, where the sample consists of two existing classes of tenth grade students.

Sampling Procedures

The students in the study are enrolled in English II (10th grade). They are divided into two sections of twenty-three students each, which are offered during the fourth and fifth hours of the school day. Has selection bias operated in the creation of the two class sections? A review of course assignments shows that all 10th grade band students (eight) are enrolled in the fifth hour class. Six are females, and two are males. These eight students have an accumulated mean grade point average (GPA) of 3.33. The accumulated mean GPA of the fourth hour class is 2.43, while the accumulated GPA of the fifth hour class with the band members' values extracted is 1.84. Overall the accumulated GPA of the fifth hour class is 2.33.

While band students must be assigned to the fifth hour class, and while the distribution of boys and girls differs between classes (4th hour M=12, F=11; 5th
hour M=7, F=16), there seem to be few other differences. No statistically significant differences in mean scores were found between the groups on both the pre-test of test-wiseness skills, the DAT Verbal Reasoning test, and grade point average.

T-tests were used to examine possible differences between groups. As indicated in Table 1, although differences in mean scores were found on each of the three measures tested, the differences were not statistically significant at the .05 alpha level. The experimental and control groups are considered to be equivalent.

A weakness of the study design is that random assignment of subjects to

Table 1

Comparability of Experimental and Control Groups on DAT-VR Scores, Ace That Test Pre-Test Scores, and GPA

<table>
<thead>
<tr>
<th>Type of Group</th>
<th>Type of Measure</th>
<th>m</th>
<th>SD</th>
<th>SE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. Control</td>
<td>DAT-VR</td>
<td>52.8824</td>
<td>30.1515</td>
<td>7.3128</td>
<td>.6506</td>
</tr>
<tr>
<td></td>
<td>DAT-VR</td>
<td>57.1905</td>
<td>27.8830</td>
<td>6.0846</td>
<td></td>
</tr>
<tr>
<td>Exp. Control</td>
<td>pre-test</td>
<td>16.1364</td>
<td>3.091</td>
<td>.659</td>
<td>.398</td>
</tr>
<tr>
<td></td>
<td>pre-test</td>
<td>17.0476</td>
<td>3.840</td>
<td>.838</td>
<td></td>
</tr>
<tr>
<td>Exp. Control</td>
<td>GPA</td>
<td>2.5241</td>
<td>.792</td>
<td>.169</td>
<td>.743</td>
</tr>
<tr>
<td></td>
<td>GPA</td>
<td>2.4200</td>
<td>1.248</td>
<td>.266</td>
<td></td>
</tr>
</tbody>
</table>

p = .05

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
experimental and control groups was not possible. While using existing classroom
groups as subjects allowed the researcher to conduct the project in an applied set-
ting, the results cannot be as readily generalizable as would those from a more
rigorous randomized pre-test post-test control group design. This type of sample
heightens the researcher's concern for internal and external validity within the
design of the study.

Also, in a school setting, particularly in a small school, students have both
the motive and the opportunity to discuss their daily lives with each other. It is
possible that students in the experimental group could share information about
the instructional program with those students who are in the control group. The
control group's access to the instructional program was limited by having the
teacher retain possession of all materials, both before and after class. Exceptions
occurred in the case of several student absences, which did require individual
access to the materials, in order to make up work missed in class at home. An
additional strategy was to time the MEAP testing to start on the day immediately
following the conclusion of the post-test at the end of the instructional program
(to limit the span of time in which information might be shared among students
from different groups).

Threats to Validity

There are a number of potential internal and external threats to the validity
of the study. Internal threats include selection and testing in particular, while
external threats include the Hawthorne, John Henry and Pygmalion effects, and
the novelty and experimenter threats.

**Internal Threats to Validity**

The first internal threat to validity is selection. To address the potential risk
that the effect may be due to initial differences between groups as opposed to the
treatment, the researcher studied the composition of the experimental and control
groups. Three measures were analyzed. No statistically significant differences
between means were found on measures of grade point average, *Ace That Test*
(Yarosz & Yarosz, 1991b) pre-test scores of test-taking skills, or the DAT Verbal
Reasoning Test.

A second threat to internal validity is repeated testing. Because the subjects
were both pre-tested and post-tested on test-taking skills, it is possible that famil-
liarity with a test could enhance performance by students remembering items and
responses. This threat was addressed by using alternate forms of the pre- and
post-test.

**External Threats to Validity**

Five external threats to validity are considered in the study: the Hawthorne,
John Henry and Pygmalion-like effects, and the novelty and experimenter threats.
The first three of these, Hawthorne, John Henry and Pygmalion-like effects can
result from the attempts of subjects, who know they are in a study, to look good.
Findings from studies with these effects are not the result of the experimental treatment; rather, they result from the subjects' knowledge that they are subjects in a research study.

The subjects in this study were aware that they were part of a research effort. The researcher met with the two groups of tenth graders and provided the following explanation:

We are trying something we think could be helpful to our students in taking tests, but we don't know if it is of value. We want to know if it is good, so we're going to give it to a group of you before you take the Michigan Educational Assessment Program tests. All other students will receive the training after the MEAP has been taken. We are asking for your help with this so we can figure out if this is something that will help our kids do better on their tests. As you may know, in order to attain a state-endorsed diploma, you must achieve a satisfactory score on the MEAP tests. This requirement started with last year's tenth graders. Fifty two of them took the MEAP tests last year, and only nineteen of them qualified for the diploma. We are looking for a way to help you do well on the tests. All students taking English II will take a pre-test which will show you the skills you already have in test-taking. Then one class will receive 16 hours of training, and all students will then be post-tested to see if the training made a difference. Following the post-test, all students will take the MEAP tests. Then those students who did not receive the instruction prior to MEAP will receive it. When the MEAP scores are reported back to us, we will look to see if there is a difference between groups in the proportion of students receiving satisfactory scores. Your parents will be asked to sign a release which gives us permission to use your test score information (see Appendix C). There is no penalty to you if you choose not to allow your test scores to be used. It is not going to affect your grades.

A Hawthorne-like effect occurs when the subjects in the experimental group have improved performance which is not the result of the experimental treatment; rather the improvement in performance occurs because the subjects are responding to the experience of being in the study (Borg & Gall, 1983).
The possibility of a John Henry-like effect exists when the control group students know they are in a study, and the control group performs above its usual level, not because of the treatment provided to the experimental group or the control group; rather the higher performance of the control group is the result of competition with the experimental group (Borg & Gall, 1983).

In the present study, the students knew they were part of a study and this could have motivated them to compete with the experimental group and to outperform them on the MEAP test. A strategy for countering a John Henry-like effect was using the Ace That Test post-test for both the experimental and control groups, prior to the MEAP, to identify any differences between the time of pre- and post-testing which may have occurred during the treatment phase of the study. If the control group had performed better than the experimental group on the post-test of test taking skills, the possibility of John Henry-like effect could be noted.

A Pygmalion-like effect can occur when what is expected by the experimenter causes changes in the behavior of the subjects, rather than changes occurring as a result of the treatment (Borg and Gall, 1983). In this study, this was countered by the explanation to the subjects, in which the experimenter explained that whether or not this program would be of value was unknown. The expectation established could be described as "help us find out if this makes a difference", rather than "this will work".

The novelty threat, where findings can result from the enthusiasm generated
by trying something new, was countered by two factors in this study. First, the treatment phase of the experiment lasted eighteen instructional days, almost an entire month of the school year, and was presented in the student’s regular instructional setting. It seems likely that any initial enthusiasm may have diminished by the completion of the program. Second, tenth grade students are simply not enthusiastic about taking the MEAP test (see page 18); rather they tend to place little value upon it.

The experimenter threat, where the behavior of subjects may be unintentionally affected by characteristics or actions of the experimenter was addressed. The strategy used was to have the same teacher teach all students in both groups. The teacher was instructed to treat the two classes identically except for the training in the Ace that Test material.

Additionally, all students were tested on the MEAP together in the same setting, so that the test conditions were identical. Their pre- and post-test experiences with Ace That Test materials were in their regular English II classroom to avoid drawing undue attention to the project which might influence student expectations.

Directions for activities were followed in accordance with the prescribed procedures of each test (see Appendix A for details).

Effect Size

Scientific methodology includes a discussion of the value and place of
significance testing, with some participants arguing that effect size is more important in the interpretation of findings. Davidson and Giroir (1989) define effect size as how much of the dependent variable is accounted for by the independent variable. Problems they identify in the use of significance testing include errors of interpretation, whereby the researcher underinterprets an outcome with a large effect size which is statistically nonsignificant, or overinterprets a statistically significant outcome which involves a small effect size. They argue that the inclusion of effect size in result interpretation can prevent such errors.

Hays (1981, p. 293) is emphatic in asserting that "virtually any study can be made to show significant results if one uses enough subjects." Thus, consideration must be given to the size of the sample as the researcher considers how to put it in context. If, for example, a nonsignificant result would be significant with a few more subjects, such information may helpful. Similarly, where a few less subjects would change a significant result to a nonsignificant result, the researcher may have a responsibility to so indicate. According to Thompson (1988, p. 147) when results are significant and sample size is small, or when sample size is large and results are not significant, the results would be far more worthy of note. Because effect size is not influenced by sample size, results are more accurately portrayed when effect sizes are reported and interpreted.

Accordingly, results of the present study will include tests of statistical significance and calculations of effect size where appropriate. With a sample population of twenty-three subjects in the experimental group, and twenty-three
subjects in the control group, the small sample size is considered a factor to evaluate in grasping the magnitude of the effect of the treatment. The use of effect size, which is a standardized measure, allows comparison of outcomes, and serves as a backup to the test statistic. Should an effect size near .25 occur, the practical importance of this difference will be discussed.

The Changing Perception of Validity

Relations among the three approaches to validation, in which content, construct, and criterion were considered separate but equal types of validity, have changed (Geisinger, 1992). Content and criterion validities are now only facets of overall validity. The meaning first attached to construct validation has been increasingly taken on as test validation. Thus, an important aspect of test construction consists of carefully developed definitions of the construct to be measured. Also important is expert review indicating that the test items are consistent with the construct definitions and theory. Construct validation subsumes all other approaches to test validation.

With the publication of the 1974 standards for educational and psychological tests and manuals (American Psychological Association, 1974) note was made that expert judgment of test content was no longer considered adequate for content validation, rather validation concerned inferences made regarding the test scores. The 1985 Standards (American Educational Research Association, American Psychological Association & National Council for Measurement in Education,
1985) clarify that the larger change is that attention is focused on evidence which supports test uses, rather than test characteristics.

When can an inference from a test score be considered sufficiently valid? According to Cole and Moss (1989) sufficient validity occurs when a variety of types of evidence support the instrument's plausibility and eliminate primary counterinferences. Discussion of the Ace That Test Pre-test, Instructional Program and Post-test below points out the strengths and weaknesses of the validity as reported.

Instruments

The instruments for this study are (1) the Ace That Test Commercial Pre-test, Instructional Program and Post-test; (2) the MEAP, Michigan Educational Assessment Program tests in mathematics and reading for grade ten, 1992; and (3) the MEAP Student Self-report of Performance, Interest and Effort test items for each reading selection, grade 10, 1992.

The Ace That Test Pre-Test, Instructional Program and Post-Test

The Ace That Test materials are a newly available commercial product intended for use at the secondary and post-secondary level, to bring about test sophistication (Yarosz & Yarosz, 1991b). The test scores of students trained in the test sophistication program would more accurately reflect students' achievement levels by reducing any influence on those scores which is brought about by
the students' lack of knowledge about how to take tests. Developed over a twenty-five year period of research by Dr. Edward Yarosz, Ed.D, Associate Professor of Education at Hunter College, the program is based upon findings in the literature on test sophistication and test-wiseness. The pre-test and post-test components have the primary purpose of providing diagnostic information, and estimates of confidence building in test-taking. Two equivalent forms of the test of test-taking skills (Form DY and EY) were developed so that any post-test gains can be attributed to the program, and not to possible test-retest practice effect.

Just what validity do the Ace That Test Pre-test and Post-test have? According to Yarosz and Yarosz (1991b), the tests are supported with content and construct validity. Content validity is subjective, and has been constructed by the traditional process of expert jury who attempt to establish through opinion the extent to which the content of the tests matches what is taught in the Ace That Test program, and the extent to which what is taught in the Ace That Test program matches what is known about the content of test-wiseness. Yarosz and Yarosz (1991b) assert that the content and construct validity for the Ace That Test materials are demonstrated by a variety of measures of supporting evidence as follows.

Throughout its 25 year history of development, according to Yarosz and Yarosz, the content of the Ace That Test program (lessons) was compared with findings on test sophistication and test-wiseness as they were reported in the literature. They claim that nearly all principles and techniques identified in the
literature are included in the lessons. For example, Millman, Bishop and Ebel's (1965, p.711) outline of test-wiseness principles is recognized and cited by almost all writers, and their principles are covered in the Ace That Test program.

Behavioral scientists have come to support the concept of validation as construct validation. In order to assess the construct validity of any test, and to answer the question at hand in the present study, the researcher attempts to assess the extent to which the test may be said to measure a theoretical construct or trait. According to Anastasi (1988) this is accomplished by establishing inter-relationships among behavioral measures, and gradually accumulating information from various sources. Such sources may consist of empirical studies employing test scores, or normative data in which test score differences may be studied across variables such as age, gender, ethnicity, and setting. Evidence for this validation can come from any data that throws light on the nature of the trait being considered. It is the accumulation of research results which allows for judgments of construct validity. The investigator establishes an hypothesis or a series of hypotheses about the characteristics of persons scoring high on the test, as opposed to those scoring low. Through the process of hypothesizing, a theory about the nature of the construct the test is purported to measure is formed. Cronbach (1988) comments that the process of theorizing is never finished, because as new concepts are generated, the interpretation of existing tests must be reconsidered. A variety of types of evidence supporting the test's plausibility are needed, in which content considerations are apparent, while emphasis is
placed on appropriate uses of the test, and inferences which can be drawn from it.

The case to support the existence of construct validity in the Ace That Test pre- and post-tests is built by Yarosz and Yarosz in part by a number of empirical studies. First, Rutan (1974) in his doctoral dissertation at Rutgers University, compared twenty-five experimental and twenty-three control groups of suburban high school students on pre- and post-test measures of test sophistication and on the Math and Verbal Reasoning subtests of the Differential Aptitude Test. He used the original 30 item version of the test of test sophistication. The experimental group subjects were taught test sophistication using the first versions of the Ace That Test Preparation for Success Student Guide and the Ace That Test Preparation for Success Teacher Instruction Manual. Statistically significant differences, at the .05 level of significance using a two-tailed t-test, were found between experimental and control groups, which indicated higher scores for the experimental group. Yarosz and Yarosz (1991b) assert that these findings favor a judgment of the existence of a measure of construct validity within the test of test sophistication.

Osborne (1985) conducted another study at Hunter College. A 42-item version of the test of test sophistication was used. Thirty-one college freshmen were divided into experimental and control groups. Using Rutan's design of pre-test, treatment and post-test she found the same level of statistically significant difference, with the experimental group scores higher. Yarosz and Yarosz cite this as
another piece of supporting evidence toward construct validity. In 1988, Donald Yarosz studied "The Effect of a Test Sophistication Training Program on Male College Athletes," using a 56-item version of the test of test sophistication. He used a pseudo-experimental design (with pre- and post- tests and no control group). The experimental group of 22 football players had statistically significant higher post- than-pre-test results, as measured by the $t$ test of significance at the .05 level, and $r$ was equal to .54 between scores on the post-test of test sophistication and the participants' semester GPA'S. Yarosz and Yarosz (1991b) indicate that these findings can be interpreted as favoring construct validity.

In 1989 and 1990 two groups of graduate students (consisting of thirty-three and thirty-two students each) who were in the counselor training program at Hunter College, were provided a teacher-directed review of all principles covered in the Ace That Test manual by Dr. Edward Yarosz using the tests of test sophistication. In 1989 he used the 32-item test of test sophistication, and in 1990 the 34-item test. He found that there was significant improvement for both groups on the test sophistication tests on the basis of pre- and post-test comparisons, both at the .05 level of significance using a $t$ test. Yarosz and Yarosz (1991b) claim further substantiation of the presence of construct validity on the basis that, since 1976, more than twenty-five graduate students at Hunter College have conducted similar studies using all of the versions of the test of test sophistication, obtaining similar results with a wide variety of clients and settings.

While Yarosz and Yarosz (1991b) cite a variety of supporting evidence to
build the case of verification of construct validity, they may have neglected to address two methods of verification which are based upon comparison. One method would be to compare the results of the Ace That Test program with another measure of the same construct.

A second method of substantiating the Ace That Test program's construct validity would be to compare its results to the results of a measure of a similar but different construct, with a finding of a low level of correlation. It is not clear that such findings have been reported. It is suggested that the authors expand their validation efforts accordingly, or report such findings more fully in the Ace That Test materials if such comparisons have been made. In addition, it would be helpful if Yarosz and Yarosz would include a discussion of effect size to further illuminate the findings of the studies they cite.

Reliability

Since 1985, two tests of the reliability of the Ace That Test test of test sophistication found the Pearson product moment coefficient of correlation to be only .75. Similarly, Mainzer (1988) found his $r$ was equal to .70. These are low scores of reliability. Both $r$'s were low for two reasons: (1) some of the subjects were motivated between the start and end of the testing to learn more about test sophistication, and did so; and (2) the spread of scores was limited in both groups (few high scores, especially on pre-tests). In his study Von Zerneck (1989) showed an $r$ of .88, and it was replicated by Dr. Edward Yarosz in 1990 with an
$r$ of .91. These study results better support the reliability of the present test forms as pre- or post-tests. Because alternate forms of the pre-test and post-test were used, as opposed to test-retest procedures, a lower correlation can be received more favorably.

**Ace That Test Instructional Program**

Before anything about test-taking skills was taught, the pre-test was administered, and students were provided a profile of their individual strengths and weaknesses, which is written in such a way as to provide positive reinforcement (see Appendix B). This feedback suggested the amount of focus the student should give to each area of skill development when taking the instructional program.

The principles and techniques reported in the literature are addressed in the program's sixteen hours of lessons, and comprise the pre- and post-tests to ensure content validity. The areas covered in the instructional program are divided into nine lessons, using both direct and indirect questioning: (1) using directions well, (2) tactics for multiple-choice questions, (3) multiple choice key words, (4) tactics for paragraph-reading questions, (5) using time wisely, (5) tactics for essay exams, (6) unusual types of questions, (7) guessing and changing answers, (8) attacking teacher-made tests, and (9) preparing the night and hours before. The lessons include forty-nine brief checks about how to do well on tests.

Learning principles that are used in the instructional program, and which
are built into the teacher's guide and student guide activities include: (a) the teacher elicits pupil motivation initially and at regular intervals thereafter, in any program in which learning is to happen; (b) learning by doing is a more effective skill development technique than lecturing; (c) learning requires frequent reinforcement; (d) sharing of common concerns and problems by peers builds confidence and learning; and (e) developing competence is one way to reduce anxiety and build confidence. By completing the exercises, students become more familiar with different types of tests, get practice in responding to tests, and have an opportunity to share correct answers with other students and the teacher, who provides the group with answers to missed questions, and facilitates group discussion.

The Michigan Educational Assessment Program

Three parts of the MEAP at the tenth grade level were used as measures of mean differences between the experimental and control groups, following the test sophistication treatment. The MEAP program is a statewide testing program funded by the legislature. It is revised annually, and provides information about how a student's performance compares to the State Board of Education's expectation of satisfactory performance of essential skills for students at grade level.

First, the Basic Skills Mathematics Test performance can be interpreted in two ways: (1) in terms of the content or topics included, and (2) in terms of the process or types of thinking required. The test is developed in thirteen strands,
eight of which are content-focused: (1) whole numbers and numeration; (2) fractions, decimals, ratio and percent; (3) measurement; (4) geometry; (5) statistics and probability; (6) algebraic ideas; (7) problem solving and logical reasoning; and (8) calculator literacy and use. Five additional strands are process-focused and include: (1) conceptualization, (2) mental arithmetic, (3) estimation, (4) computation, and (5) applications and problem solving.

The criteria used by the state to evaluate qualification for a state-endorsement on the diploma requires that students pass both of the tenth grade tests, which are in mathematics and reading, and the eleventh grade test in science. If a student fails to achieve the 50% level on any of these measures as described below, a non-endorsed diploma may be issued for graduates in 1994, 1995, and 1996. Potential graduates failing to achieve the 50% level in 1997 in any of the three tests will not receive a diploma.

The criteria for success in mathematics is a scaled score of 500, which is raw score of 50% or more of the 120 test items in math. In reading, the MEAP test requires students to read story and informational selections, and to respond to four categories of questions. They are: (1) constructing meaning (how well students understood what was read), (2) topic familiarity (a measure of some of the previous information the student already knew about the topics in the reading selections), (3) knowledge about reading (the student's knowledge about and understanding of reading strategies and skills needed to understand each reading selection), and (4) student self-report of performance, effort and interest (how the
student viewed his performance effort and interest while reading the selections and answering the test questions). The criteria used by the state to evaluate qualification for a reading endorsement on the diploma is a scaled score of 300 (or sixteen items correct) on twenty items constructing meaning from a story selection or on twenty items constructing meaning from an informational selection. The eleventh grade science test is not a focus of this study and is not discussed further.

The Michigan Department of Education sets the parameters for administration of the MEAP Tests. In 1992, the testing activities were scheduled to be conducted between September 21 and October 23. Five weeks were provided to allow for adequate time to schedule around the many school activities taking place in the Fall. Students need approximately 2.5 hours to complete testing for reading and 2.5 hours for mathematics. It is recommended that the testing sessions be no longer, nor more frequent, than one hour per school day. The process can become a lengthy one, when consideration is given to all of the various make-up sessions which are required due to individual student absences.

Testing procedures used are described precisely and in detail in Appendix A. Students were assembled for the MEAP testing experience first hour of the day on October 12, 13, 14, 15, 16, and 20 in the school cafeteria. This was done to provide a uniform testing experience for all tenth grade students. Group testing sessions were scheduled for test make-ups necessitated by student absences occurring on the scheduled testing date on October 21, 22, or 23. These sessions
were held in the library or in a smaller conference room, depending on the size of the group being tested. Seventh, eighth and eleventh grade students needing make ups were also part of these sessions. One student from the control group (there were no students absent from the experimental group) was given one make-up test in the conference room on October 22.

Two students, one from the experimental and one from the control group, were provided the audiotape version of the MEAP test which is available to special education students with reading deficiencies (these students were enrolled in both Special Education Reading and English II classes). These students were tested during their regular instruction time in the special education classroom. Because of delays in shipping the audiotapes to local schools, this testing took place during the week of October 26. Results were not available for inclusion in the study. In addition one control group student did not take the MEAP reading tests.

Reading and mathematics test materials consist of separate test booklets and answer sheets for reading and math, metric rulers and acetate grids and protractors for each student. Number two pencils are used to bubble the answer sheets and complete identification information. Optional research reports were requested which will provide test score information disaggregated by experimental and control group, and by boys and girls within each group. Student answer sheets were coded accordingly.
Teacher

The instructional program was taught by the students' regularly assigned English II teacher, who instructs the tenth grade students. Instruction was presented traditionally, as a unit in the regularly scheduled school day for sixteen consecutive days. Training provided to the teacher included an explanation of the program and its conceptual background (on the Ace That Test informational video); an individualized inservice by a professional knowledgeable about the Ace That Test product; and a detailed instructional guide for the lessons.

Experimental Procedures

Tenth grade students were divided into experimental and control groups, based upon their assignment to scheduled sections of English II. The fourth hour class had twenty-three students, and a mean grade point average of 2.43. The fifth hour class had twenty-three students, and a mean GPA of 2.33.

T-tests were used to examine possible differences between groups. Three measures were studied: (1) grade point average, (2) Ace That Test pre-test scores, and (3) Differential Aptitude Test Verbal Reasoning scores. As indicated on Table 1, although differences in mean scores were found on each of the three measures tested, the differences were not statistically significant at the .05 alpha level. The experimental and control groups are considered to be equivalent.

The fifth hour class was the control group, with the fourth hour class the
experimental group. The control group took the pre-test, the post-test and the MEAP tests, while the experimental group took the pre-test, received the 16 hour instructional program, took the post-test and the MEAP tests.

Analysis of Data

Experimental and control group student mean scores obtained on the following dependent variables were used as the unit for analysis: (a) the Ace That Test post-test; (b) the MEAP mathematics subtest; (c) the MEAP Story Reading subtest; (d) the MEAP Informational Reading subtest; and (e) the MEAP Self-report of Performance, Effort and Interest on Story and Informational Reading subtests. Differences in the proportions of students in the experimental and control groups who qualified for state endorsement on their diplomas through satisfactory MEAP scores were also examined.

Ten research hypotheses were postulated for this study. The statistical or null forms of these hypotheses are as follows:

1. The difference between the post-test mean scores of students that experienced the Ace That Test instructional program and students who did not experience the program is zero.

2. The difference between the mean scores of students who experienced the Ace That Test program on the MEAP Math and Reading subtests, and students who did not experience the program is zero.

3. The difference in the proportion of students who experienced the Ace
That Test program and who reported high, moderate or low on the MEAP Self-report of Performance, Effort and Interest on Story Reading and Informational Reading, and students who did not experience the program and who reported high, moderate or low is zero.

4. The difference in the proportion of students who experienced the Ace That Test training and who qualified for state endorsement in Math and Reading on their diplomas and the proportion of students who did not experience the training and who qualified for state endorsement in Math and Reading on their diplomas is zero.

5. There is no interaction between gender and type of group (instructed or not instructed) for Ace That Test post-test scores.

6. There is no interaction between gender and type of group (instructed or not instructed) for MEAP math scores.

7. There is no interaction between gender and type of group (instructed or not instructed) for MEAP story reading scores.

8. There is no interaction between gender and type of group (instructed or not instructed) for MEAP informational reading scores.

9. There is no interaction between gender and type of group (instructed or not instructed) for MEAP Students' Self-reports of performance, effort and interest on the story reading test.

10. There is no interaction between gender and type of group (instructed or not instructed) for MEAP Students' Self-reports of performance, effort and
interest on the informational reading test.

T-tests were used to test differences in mean scores of the two groups. Chi-square tests were used to test the differences in the proportions of students from the experimental and control groups who reported high, moderate or low levels of performance, effort and interest on MEAP, and who achieved the MEAP cut scores on math and reading. Interaction between gender and type of group was tested using two-factor Anova.

All calculations for statistical significance were computed using the statistical package SPSS-X, Norusis (1988). Computer Center facilities at Western Michigan University were used for this purpose. Effect sizes were calculated manually by the researcher. Since large samples may produce statistically significant results with practically insignificant differences (and vice versa), effect sizes are useful in pointing up possible strengths or weaknesses of the statistical findings. The following chapter reports the findings which resulted from the data analysis.
CHAPTER IV

FINDINGS

The purpose of this study was to address the following problem: what is the effect of test sophistication training on tenth grade student scores on the subtests of the Michigan Educational Assessment Program?

The design of the study involved the participation of tenth grade students in two class sections of English II, one of which received the commercially prepared Ace That Test training in test sophistication. Both groups were given the Ace That Test pre-test to assess pre-existing knowledge of test-wiseness; the experimental group received the 16 hour instructional program, and both groups received the Ace That Test post-test, MEAP Math, Reading and Self-report sub-tests.

In this study, the independent variables were the Ace That Test instructional program and gender. The dependent variables were learning, and perceptions about testing, as measured by the Ace That Test pre- and post-tests, the MEAP Tests of Math, Story Reading and Informational Reading and Self-reports of Performance, Effort and Interest on the MEAP reading tests.

When the two independent variables are considered ten different hypotheses can be tested. The null hypotheses are stated in Chapter III. The ten research hypotheses, stated as alternative hypotheses were as follows:
1. Students who experience test sophistication training score higher on a post-test of test-wiseness skills than students not experiencing the training.

2. Students who experience test sophistication training score higher on the MEAP Math, Story Reading and Informational Reading subtests than students not experiencing the training.

3. The proportion of students who score high on the Meap Self-report of Performance, Effort and Interest on story reading and informational reading is greater for students who experienced the test sophistication training than for students not experiencing the training.

4. The proportion of students who achieve the state endorsement in math and reading, is higher for students who experience the test sophistication training than for students who did not experience the training.

5. There is an interaction between student gender, and learning from the experience of test sophistication training, as measured by achievement on the Ace That Test post-test.

6. There is an interaction between student gender, and learning from the experience of test sophistication training, as measured by achievement on the MEAP mathematics test.

7. There is an interaction between student gender, and learning from the experience of test sophistication training, as measured by achievement on the MEAP story reading test.

8. There is an interaction between student gender, and learning from the
experience of test sophistication training, as measured by achievement on the
MEAP informational reading test.

9. There is an interaction between student gender, and learning from the
experience of test sophistication training, as measured by achievement on the
MEAP Student's Self-report of performance, effort and interest on the story reading test.

10. There is an interaction between student gender, and learning from the
experience of test sophistication training, as measured by achievement on the
MEAP Students' Self-report of performance, effort and interest on the informational reading test.

The discussion of the findings for each of the research hypotheses follows
the discussion of the comparability of the experimental and control groups.

Equivalence of Groups

In this study design, where one group of subjects received the treatment
(instruction in test sophistication), and the other group did not, it is important to
assess the comparability of the groups to determine if they are equivalent. Three
measures were examined to answer this concern. These included the students' scores on the Verbal Reasoning subtest of the Differential Aptitude Test, the students' grade point averages, and the students' scores on the Ace That Test pre-test of test-wiseness skills. Table 1 reports the findings of these three measures.
T-tests were used to examine possible differences between groups. As indicated in Table 1, although differences in mean scores were found on each of the three measures tested, the differences were not statistically significant at the .05 alpha level. The experimental and control groups are considered to be equivalent.

**Hypothesis 1: Type of Group and Test Sophistication Achievement**

The null form of Hypothesis 1 tested in this study is: the difference between the post-test mean scores of students that experienced the Ace That Test instructional program and students who did not experience the program is zero.

In order to test this hypothesis two intact classroom groups of tenth grade English II students who had been determined to be equivalent on measures of grade point average, Differential Aptitude Test Verbal Reasoning scores and Ace That Test pre-test (form DY, a 34 item test on which results are reported as the number of items correct) scores of test-wiseness were selected as experimental and control groups.

The experimental group received the sixteen hours of instruction known as the Ace That Test program. Both groups took the 34 item form EY post-test. Post-test scores were reported on an interval scale from 0-34, as the number of items answered correctly. The Ace That Test computer scoring center was used to grade the tests.

The distribution of the test scores of all students in the study ranged from
11 to 25 points. The maximum possible score was 34 points. Test scores of the group receiving the program ranged from 15 to 25, while scores of untrained students in the control group were from 11 to 23 points. The results of the t-test of these data are presented in Table 2.

T-tests were used to determine whether either the experimental or the control group gained in performance on the post-test scores. As noted in the column totals of Table 2 the post-test mean score obtained for the instructed group (19.4545) is significantly higher than the pre-test mean score of that group (p < .05) while for the noninstructed group there is no difference (p < .05). A test of the practical significance of this finding was the calculation of effect size by the researcher using the formula $Es = (\overline{X_e} - \overline{X_c})/SD_c$. Effect size was calculated to be 1.49, indicating that this is statistically significant result that is not

<table>
<thead>
<tr>
<th>Type of Group</th>
<th>Type of Measure</th>
<th>m</th>
<th>SD</th>
<th>SE</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp.</td>
<td>pre-test</td>
<td>16.1364</td>
<td>3.091</td>
<td>.659</td>
<td>.000*</td>
</tr>
<tr>
<td></td>
<td>post-test</td>
<td>19.4545</td>
<td>2.483</td>
<td>.529</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>pre-test</td>
<td>17.0476</td>
<td>3.840</td>
<td>.838</td>
<td>.120</td>
</tr>
<tr>
<td></td>
<td>post-test</td>
<td>15.7727</td>
<td>2.476</td>
<td>.540</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05
over-interpreted. (Cohen has indicated that an effect size of .25 is small, .50 is medium, and 1.0 is large.) Therefore, the null hypothesis of no differences in test-wiseness achievement between the group receiving the training and the group not trained was rejected.

**Hypothesis 2: Type of Group and Achievement on a Criterion-Referenced Test**

The null form of Hypothesis 2 tested in this study is: The difference between the mean scores of students who experienced the Ace That Test program on the MEAP Math and Reading subtests, and students who did not experience the program is zero.

In order to test this hypothesis, MEAP math, story reading and informational reading test scores from students in the experimental and control groups were examined. The State of Michigan established a cut score of 500 in mathematics for students qualifying for the endorsed diploma. The distribution of the test scores for both groups ranged from 461 to 587. In the experimental group the distribution ranged from 485 to 545, while in the control group it ranged from 461 to 587.

In reading, the State of Michigan established a cut score of 300 on either the story reading, or the informational reading test for students qualifying for the endorsed diploma. In story reading the distribution of the test scores for both groups ranged from 286 to 347; the experimental group range was from 291 to 347, while the control group range was from 286 to 347. In informational reading
the distribution of the test scores for both groups ranged from 274 to 335 with the experimental range from 277 to 316 and the control group range from 274 to 335.

The results of the t-tests of these data are presented in Table 3.

The observed difference between mean test scores of experimental and control groups was not statistically significant on math, story reading or informational reading ($p < .05$). A review of the mean scores shows that although not statistically significant, on each MEAP subtest, the group that received the instruction scored slightly higher than the control group. In math, the difference between group mean scores was 2.8637 (on a range of 461 to 587); in story reading the difference between group mean scores was 1.8766 (on a range of 287 to 347); and

Table 3

<table>
<thead>
<tr>
<th>Type of Group</th>
<th>MEAP Test</th>
<th>m</th>
<th>SD</th>
<th>SE</th>
<th>P</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. Math</td>
<td>505.2273</td>
<td>16.338</td>
<td>3.483</td>
<td>.677</td>
<td>.1041</td>
<td></td>
</tr>
<tr>
<td>Control Math</td>
<td>502.3636</td>
<td>27.507</td>
<td>5.864</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp. St.R.</td>
<td>308.5909</td>
<td>22.863</td>
<td>4.874</td>
<td>.758</td>
<td>.1139</td>
<td></td>
</tr>
<tr>
<td>Control St.R.</td>
<td>306.7143</td>
<td>16.463</td>
<td>3.592</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp. In.R.</td>
<td>303.2273</td>
<td>15.669</td>
<td>3.341</td>
<td>.907</td>
<td>.0336</td>
<td></td>
</tr>
<tr>
<td>Control In.R.</td>
<td>302.6190</td>
<td>18.062</td>
<td>3.942</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$p = .05$
in informational reading the difference between mean scores was .6083 (range of 274 to 335). These are small differences. A greater understanding of these scores was achieved by calculating the effect sizes.

The practical significance of each finding was assessed by the calculation of effect size. They are .1041 in math, .1139 in story reading, and .0336 in informational reading; all three scores are below Cohen’s .25 level of a low effect size. Practically speaking, then, it may be said that negligible differences were found. As a consequence, the null hypothesis of no differences in achievement between students experiencing the test sophistication training, and students not experiencing the training, was retained.

There was not enough evidence to conclude that a significant difference in achievement on the MEAP math and reading subtests existed between students who experienced the test sophistication training, and those who did not.

Hypothesis 3: Type of Group and Perceptions About Testing

The null form of Hypothesis 3 tested in this study is: the difference in the proportion of students who experienced the Ace That Test program and who reported high ratings on the MEAP Self-report of Performance, Effort and Interest on story reading and informational reading, and students who did not experience the program and who reported high ratings is zero.

In order to test this hypothesis, chi-square tests were used to study the proportion of students reporting low, moderate or high levels of performance, effort
and interest, by type of group (experimental or control). Observed significance levels ($p \geq .05$), ranging from .10209 to .68126 were found, indicating no statistically significant differences. See Table 4.

This table also indicates the number of students rating themselves at each level in each category, by group. The null hypothesis is retained.

Table 4

Type of Group and MEAP Self-Reports of Performance, Effort and Interest on Story and Informational Reading Tests

<table>
<thead>
<tr>
<th></th>
<th>Story Reading</th>
<th></th>
<th>Informational Reading</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exp.</td>
<td>Con.</td>
<td>$x^2$</td>
<td>Sig.</td>
</tr>
<tr>
<td>Performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>3</td>
<td>1</td>
<td>1.24726</td>
<td>.53600</td>
</tr>
<tr>
<td>Mod</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>High</td>
<td>8</td>
<td>10</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Effort</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>3</td>
<td>2</td>
<td>4.56383</td>
<td>.10209</td>
</tr>
<tr>
<td>Mod</td>
<td>16</td>
<td>10</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>High</td>
<td>3</td>
<td>9</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Interest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>12</td>
<td>7</td>
<td>3.07197</td>
<td>.21524</td>
</tr>
<tr>
<td>Mod</td>
<td>5</td>
<td>10</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>High</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Observed significance values were not < .05.

Note: With 23 students in each group, results were unavailable for 2 special education students (1 per group), and 1 student in the control group refused to take the reading tests.
It was expected that more students in the group receiving the training would assess their performance effort and interest higher than the untrained students. This did not occur. These students did not report greater numbers of high performance on the tests, or of high effort on the test, or of high interest in the test. Had their confidence or motivation been strengthened by the training program, an indication could have been made here. On the other hand, larger numbers of students from the control group reported high levels of effort on the tests. This may be an indication that the students trained in test-wiseness did not perceive the tests as being as hard as untrained students did.

These findings help to counter possible assertions that the interference of motivation or a Hawthorne-like effect contaminated the results. There were no statistically significant differences favoring the experimental group in how well the students thought they performed, how hard they said they tried, or how interested they said they were.

**Hypothesis 4: Type of Group and Proportion Qualifying for Diploma Endorsement**

The null form of Hypothesis 4 tested in this study is: The difference in the proportion of students who experienced the *Ace That Test* training and who qualified for state endorsement in Math and Reading on their diplomas, and the proportion of students who did not experience the training and who qualified for state endorsement in Math and Reading is zero. Chi-square tests were used to
test the differences in the proportions of students from the experimental and control groups who qualified for state endorsement. Table 5 shows the findings.

While not statistically significant, the findings show that on the three measures represented, math, reading and both math and reading combined, more students receiving the training passed the tests (50% of the students in the experimental group, as compared to 36.3% in the control group). Observed significance levels did not indicate statistical significance at the .05 level. Without findings

| Type of Group and MEAP Math, Reading: Diploma Endorsement |
|---------------------------------|---------------------|---------------------|
| Exp.  | Con.  | Type of Group   | Gender |
|       |       | Exp.  | Con.  |       |       | Significance                |     |
| Exp.  | Con.  |      |       |       |       | Male | Female |       |       |       |       |     |     |
| Both Math & Reading              |                     |       |       |       |       |       |       |       |       |       |     |     |
| Pass     | 11  | 8   | .61743 | .43200 | 8   | 11 | .34925 | .55454 |     |     |     |
| Fail     | 11  | 13  |       |       | 8   | 16  |       |       |     |     |     |
| Math     |       |     |       |       |     |     |     |     |     |     |     |     |     |
| Pass     | 12  | 8   | 1.46667 | .22587 | 8   | 12 | .02876 | .86534 |     |     |     |
| Fail     | 10  | 14  |       |       | 9   | 15  |       |       |     |     |     |
| Reading  |       |     |       |       |     |     |     |     |     |     |     |     |     |
| Pass     | 18  | 17  | .00532 | .94186 | 13  | 22 | .98496 | .98496 |     |     |     |
| Fail     | 4   | 4   |       |       | 3   | 5   |       |       |     |     |     |

Observed significance values were not < .05.

Note: With 23 students in each group, results were unavailable for 2 special education students (1 per group), and 1 student in the control group refused to take the reading tests.
of statistical significance, it must be assumed that the differences in proportions of pupils passing is due to chance. Further investigation would be needed to draw any other conclusion. The null hypothesis is retained.

Threats to Validity

It may be questioned whether the results from these four hypotheses could be attributed to other factors. Could normal maturation, the changes that occur naturally over time, have been a determinant on Hypothesis 1, where the experimental group learned the Ace that Test program? Would those students have done as well on the post-test eighteen school days later, due to the fact that they are eighteen days older? Because a control group was used and was subject to the same period of maturation, and achieved lower means scores on the post-test, normal maturation can be ruled out as a rival explanation. The group whose mean score increased was, as expected, the group whose members were trained.

Another possible concern that may be raised regarding these four achievement related hypotheses is the effect that being in a study in itself may have on either group. For example, a Hawthorne-like effect occurs when the experimental group excels simply because the subjects are in a study. This can be ruled out on the basis of the difference between the mean scores of the groups. While the experimental group did achieve statistically significant higher scores than the control group on the Ace That Test post-test this was expected because they had received the training. No similar findings occurred on any other measures, ruling
out a possible Hawthorne-like effect. This was a concern because the experimental group subjects knew they were in a study.

Concerns about the possible influence of a John Henry-like effect, where the control group students, knowing they are in a study but are not receiving the treatment, are motivated to outperform the trained students, are also alleviated. It was possible that the control group may have been motivated to compete with the other group and outperform them on the post-test, or the MEAP math and reading tests. This did not occur. Control group students did not show a consistent pattern of higher levels of achievement on MEAP math or reading tests, or on self-reports of performance, of effort, or of interest than students from the experimental group. They did show a statistically significant difference in effort on the story reading subtest, which may be an indication that the story reading test was more difficult for these students than for the trained students. However, this finding of only one statistically significant difference between experimental and control groups, (and no difference between male and female outcomes which is reported on the following pages), also supports the conclusion that the threat of selection has been controlled.

A Pygmalion-like effect can occur when what is expected by the experimenter causes changes in the behavior of the subjects, rather than changes occurring as a result of the treatment. In this study, this was countered by an explanation to the subjects that whether or not this program would be of value is unknown. The expectation established can be described as "help us find out if
this makes a difference," rather than "this will work." Findings do not support the presence of a Pygmalion-like effect. On the contrary, the expected results did not occur. Trained students did not achieve higher MEAP scores than students not trained in test sophistication.

The novelty, or disruption threat, where the experimental results may be partially due to the enthusiasm or disruption generated by the newness of the treatment was considered. It is countered by the length of the treatment, eighteen instructional days including the pre- and post-tests, which would require a concerted effort at sustained enthusiasm. It is also countered by the routinely changing nature of the material in the English II class. In this situation, units of curriculum are taught, with the content and direction of the course changing as mastery of curriculum is achieved. Because it is not unusual for work to be completed on a topic, and for a new topic of study to be undertaken, the insertion of a unit on test-taking skills can be viewed by students as a normal part of instruction, rather than as a novelty. As reported in Chapter II (page 18) there is a general lack of tenth grade student enthusiasm for taking the MEAP test. This absence of enthusiasm is a countering element to the threat of novelty or disruption which might contaminate the findings of a study because of the subjects excitement about it. Results of the study indicate that any possible enthusiasm did not carry over into performance or self-reports on the MEAP tests.

The experimenter threat, where the behavior of subjects may be unintentionally affected by characteristics or actions of the experimenter was addressed. The
strategy used was to have the same teacher teach all students in both groups. The teacher was instructed to treat the two classes identically except for the training in the *Ace that Test* material. It does not seem to be a factor which has affected either group independent of the other. Because the training was done in the regular classroom, students likely perceived it as just another classroom activity about which they had little real choice (as, for example, compared to choosing to attend an after school program, at their own expense, to improve their MEAP scores, because they may value performing well).

Could the threats of selection or testing have influenced the findings? While random assignment of students to the experimental and control groups was not possible because of the class schedules of the students, the equivalence of the groups was established on measures of grade point average, DAT VR scores and the *Ace That Test* pre-test form DY. Therefore, selection can be ruled out as a contaminating factor.

Because the subjects were both pre-tested and post-tested on test-taking skills, a threat to validity is the possibility that familiarity with the test could enhance performance by students remembering items and responses. This threat was addressed by using alternate forms DY and EY for pre- and post-test experiences.

Results of Hypotheses Concerning Gender

The following section which includes Hypotheses 5 through 10, has been
organized to present the result of each hypothesis individually. The discussion of the findings on gender follows the presentation of Hypothesis 10.

In order to test Hypothesis 5 through 10, the ANOVA results related to interaction were examined. As noted in Table 6, the interaction was not statistically significant at the .05 alpha level for any of the hypotheses.

**Hypothesis 5: Type of Group, Gender, and Post-Test**

The null form of Hypothesis 5 tested for this study is: there is no interaction between gender and type of group (instructed or not instructed) for Ace That Test post-test scores. According to the ANOVA results, an observed significance level of .848, the interaction is not statistically significant at the .05 alpha level, and the null hypothesis is retained.

**Hypothesis 6: Type of Group, Gender, and MEAP Math Test**

The null form of hypothesis 6 tested in this study is: there is no interaction between gender and type of group (instructed or not instructed) for MEAP math test scores. According to the ANOVA results, an observed significance level of .682, the interaction was not statistically significant (p < .05) and the null hypothesis is retained.

**Hypothesis 7: Type of Group, Gender, and MEAP Story Reading Test**

The null form of Hypothesis 7 tested in this study is: there is no interaction
### Table 6

ANOVA Results of Interaction Between Gender and Type of Group

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>( F )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Five</td>
<td>.037</td>
<td>.848</td>
</tr>
<tr>
<td>Six</td>
<td>.170</td>
<td>.682</td>
</tr>
<tr>
<td>Seven</td>
<td>1.204</td>
<td>.279</td>
</tr>
<tr>
<td>Eight</td>
<td>.011</td>
<td>.916</td>
</tr>
<tr>
<td>Nine - performance</td>
<td>.962</td>
<td>.333</td>
</tr>
<tr>
<td></td>
<td>effort</td>
<td>1.842</td>
</tr>
<tr>
<td></td>
<td>interest</td>
<td>1.334</td>
</tr>
<tr>
<td>Ten - performance</td>
<td>.068</td>
<td>.795</td>
</tr>
<tr>
<td></td>
<td>effort</td>
<td>.046</td>
</tr>
<tr>
<td></td>
<td>interest</td>
<td>.009</td>
</tr>
</tbody>
</table>

\( p = .05 \)

between gender and type of group (instructed or not instructed) for MEAP story reading test scores. According to the ANOVA results, an observed significance level of .279, the interaction was not statistically significant, at the .05 alpha level, and the null hypothesis is retained.

**Hypothesis 8: Type of Group, Gender, and MEAP Information Reading Test**

The null form of Hypothesis 8 tested in this study is: there is no interaction between gender and type of group (instructed or not instructed) for MEAP
informational reading test scores. According to the ANOVA results, an observed significance level of .916, the interaction was not statistically significant at the .05 alpha level, and the null hypothesis was retained.

**Hypothesis 9: Type of Group, Gender, and Self-Report on Story Reading**

The null form of Hypothesis 9 tested for this study is: there is no interaction between gender and type of group (instructed or not instructed) for MEAP Students' Self-reports of performance, effort and interest on the story reading test. According to the ANOVA results, observed significance levels of .333 on performance, .182 on effort, and .255 on interest, the interaction was not statistically significant at the .05 alpha level, and the null hypothesis was retained.

**Hypothesis 10: Type of Group, Gender, and MEAP Self-Reports on the Informational Reading Test**

The null form of Hypothesis 10 tested for this study is: there is no interaction between gender and type of group (instructed or not instructed) for MEAP Students' Self-reports of performance, effort and interest on the informational reading test. According to the ANOVA results, observed significance levels were .795 on performance, .830 on effort and .925 on interest, the interaction was not statistically significant ($p < .05$), and the null hypothesis is retained.
Discussion of Findings Concerning Gender, Hypothesis 5 Through Hypothesis 10

No differences were found between males and females. While it was expected that students who received the training in test sophistication would achieve higher MEAP scores and report higher levels of performance, effort and interest on the story and informational reading tests, differences were also anticipated along gender lines. This is because of the findings in the review of the literature regarding females. Avasthi (1991) found that girls have lower self-esteem than boys do in adolescence, and girls feel less competent in math and science. If girls feel less competent in math, it can be considered that girls might perform more poorly than boys on the MEAP math tests. However, if self-confidence is one of the variables that test sophistication helps to improve, and if self-confidence has a positive impact on test performance, it is logical that increased self-confidence which could result from test sophistication could help girls improve performance, not only on the math test, but also on the reading tests.

Avasthi goes on to assert that girls rate themselves 19% lower than boys in school competence in high school. Girls' feelings of competence are likely to be related to their confidence in their ability to read well. This ability is critical to understanding and applying subject matter from various disciplines.

While research on intellectual functioning does not report consistent differences in superiority for either gender in any specific subject area, Good and
Brophy (1986) report that girls' socialization to their gender role is related to the loss of their early elementary school advantage over boys. This supports the notion that boys could be expected to outperform girls at the high school level.

The present study considered the possibility that females, then, because of their lack of self-confidence and their diminished actual high school performance, compared to males, might be more subject to the influence of test sophistication than males. This was not supported by the data. The main difference found in the study favored the students with test sophistication training, on a post-test of test sophistication. No differences were noted by gender.

**Effect Size**

Effect size is defined as the magnitude of difference or relationship in the sample or population. It is the standardized difference between treatment and control groups, expressed in standard deviations. Cohen's (1965) definition of effect size is the degree to which a phenomenon exists, and he further categorizes effect size as small (.25), medium (.50) and large (1.0), while acknowledging that such classifications are arbitrary.

For this study, it was predetermined as important to calculate effect sizes where there was the potential for important findings being obscured. The power of the test statistic with a sample of this small size may not produce statistical significance. The effect size, however, may be significant. Effect size can be critical in a high stakes testing environment, where, if one more student qualifies for the
state-endorsed diploma the outcome is meaningful. The arbitrariness of cut
scores, by which the student's response to a single question can impact upon
future opportunities becomes a key concern. Assuming a normal distribution with
the cut score as the mean, an effect size of .25 would result in 10% more students
passing.

Effect size data are helpful when the expected results do not occur. That
has been the case in this study. While there were few expected findings of sta­tis­tical significance resulting from this study, effect sizes were calculated for rele­vant findings. The findings of this study are such that the calculation of effect
size confirmed the findings of statistical significance. It cannot be concluded that
the treatment, the Ace That Test test-wiseness training, has made a statistically
significant difference between the experimental and control groups on the MEAP
math and reading tests. Test sophistication training had no discernible influence
upon MEAP scores of tenth grade students.

Summary of Findings

The following findings emerged from the examination of data resulting from
the study. Students who received the Ace That Test test sophistication training
achieved higher Ace That Test post-test scores than students who did not receive
the training. The results indicate that test-taking skills can be learned, but the
improvement in post-test scores did not carry over into improved MEAP scores
or Self-reports.
More specifically: (1) test sophistication was learned by tenth grade subjects; (2) test sophistication did not lead to higher scores on the criterion referenced MEAP subtests in math or in story reading or in informational reading, for students at the tenth grade level; (3) test sophistication did not influence tenth grade students' assessment of their performance, effort or interest on the story or informational reading subtests of MEAP; (4) test sophistication did not increase the proportion of students qualifying for state endorsement in math and reading on their diplomas; and (5) gender was not a factor in any of the study outcomes.

Possible interpretations of the results are discussed in Chapter V, Discussion and Recommendations.
CHAPTER V

DISCUSSION AND RECOMMENDATIONS

This study was designed to test the effectiveness of test sophistication training on tenth grade student scores in math and reading on the Michigan Educational Assessment Program test. This chapter draws conclusions and makes recommendations based on results of the study. The chapter has been divided into five major sections: (1) background, (2) interpretation of the results, (3) limitations of the study, (4) implications of the findings, and (5) recommendations.

Background

The MEAP is a criterion referenced standardized test which is used by Michigan's State Board of Education to assess student achievement of basic competencies. At the tenth grade level subtests include math and story and informational reading. Beginning with the class of 1994, Michigan students must achieve satisfactory scores on MEAP to qualify for a state-endorsed diploma.

Two intact classroom groups of tenth grade students (N = 46) comprised the sample. The groups were determined to be equivalent on measures of grade point average, Differential Aptitude Test (1982) Verbal Reasoning scores, and on pre-test measures of test-taking skills. Both groups received the pre-test, the experimental group received the training in test-taking skills, and both groups
took the post-test, and the MEAP math, story and informational reading subtests, and the MEAP Student's Self-report of Performance, Effort and Interest on the reading selections. The training provided students was the Ace That Test (Yarosz & Yarosz, 1991a) program developed by researchers at Hunter College.

Interpretation of the Results

Ten null hypotheses were tested in this study:

1. The difference between the post-test mean scores of students that experienced the Ace That Test instructional program and students who did not experience the program is zero. This null hypothesis was rejected.

2. The difference between the mean scores of students who experienced the Ace That Test program on the MEAP Math and Reading subtests, and students who did not experience the program is zero. This null hypothesis is retained.

3. The difference in the proportion of students who experienced the Ace That Test program and who reported high, moderate or low on the MEAP Self-report of Performance, Effort and Interest on Story reading and informational reading, and students who did not experience the program and who reported high, moderate or low is zero. This null hypothesis was retained.

4. The difference in the proportion of students who experienced the Ace That Test training and who qualified for state endorsement in Math and Reading on their diplomas and the proportion of students who did not experience the training and who qualified for state endorsement in Math and Reading on their
diplomas is zero. This null hypothesis was retained.

5. There is no interaction between gender and type of group (instructed or not instructed) for Ace That Test post-test scores. This null hypothesis was retained.

6. There is no interaction between gender and type of group (instructed or not instructed) for MEAP math scores. This null hypothesis was retained.

7. There is no interaction between gender and type of group (instructed or not instructed) for MEAP story reading scores. This null hypothesis was retained.

8. There is no interaction between gender and type of group (instructed or not instructed) for MEAP informational reading scores. This null hypothesis was retained.

9. There is no interaction between gender and type of group (instructed or not instructed) for MEAP Students' Self-reports of performance, effort and interest on the story reading test. This null hypothesis was retained.

10. There is no interaction between gender and type of group (instructed or not instructed) for MEAP Students' Self-reports of performance, effort and interest on the informational reading test. This null hypothesis was retained.

The results of the data analysis indicated the rejection of one of the ten null hypotheses stated for this study. The null hypotheses rejected was number 1.

Null Hypothesis 1 was rejected based on a t-test analysis which found a statistically significant difference, with a probability level of .001, between the mean post-test scores of students who experienced the test sophistication training, and
students who did not. Therefore, this difference could not be interpreted as a result of chance ($\alpha = .05$).

Consequently, the evidence supports the Research Hypothesis 1 since students who experienced test sophistication training scored higher on a post-test of test-wiseness skills than students not experiencing the training.

The results of the study differed from what was expected. Based on a review of the literature, and a review of the Ace That Test test sophistication training program which indicated that its content is consistent with the body of knowledge contained in the literature review, it was expected that enhanced test-taking skills would result in enhanced MEAP scores, and in enhanced self-assessments by students about their performance on MEAP. It was not expected that all other null hypotheses concerning achievement would be retained. Neither was it expected that all null hypotheses regarding gender would be retained.

No differences were found between males and females. While it was expected that students who received the training in test sophistication would achieve higher MEAP scores and report higher levels of performance, effort and interest on the story and informational reading tests, differences were expected along gender lines. Because boys could be expected to outperform girls, the present study considered the possibility that females might be more subject to the influence of test sophistication than males. This was not supported by the data. The main difference found in the study favored the students with test sophistication training, on a post-test of test sophistication. No differences were noted by
That such results did not occur is worthy of note. It is possible that the MEAP test may not be subject to the influence of test-wiseness. It is important to recognize that MEAP is a criterion-referenced test, and that a review of the related literature indicated that little has been done to study the influence of test sophistication on criterion referenced tests (see p. 20).

Limitations of the Study

Three limitations in the study design need to be considered. First the sample was not randomly selected from the population. The population of this study was defined as the tenth grade English II classes from a rural Michigan school. The researcher selected only those tenth grade students who were enrolled in the fourth and fifth hour sections of English II. This criterion was used because it was possible to determine that the groups were equivalent on three measures: grade point average, DAT Verbal Reasoning test scores, and pre-test scores on the Ace That Test test of test sophistication. In an applied school setting, where students are assigned daily class schedules, random assignment of students to different groups was not possible.

A second limitation was lack of random assignment. One intact class was assigned to each of the two conditions: one group received the test sophistication training, and the other did not. While random assignment suggests that each student be assigned by chance to the treatment, because of the applied setting it was
necessary to assign treatment to members of an intact group. Thus it was decided to give one treatment condition per classroom, meaning that one group received the treatment and one group did not. This was done in order to protect the orderly functioning of the school. The notion of random assignment of classes to the treatment implies that the class would be the unit of analysis. However, with gender as a design variable, the analysis of the results had to be done using the student as the unit of the analysis.

A third limitation is non-generalizability of the results. This study was intended to inform faculty and administrators in a rural, culturally homogeneous Michigan school district, regarding strategies that could be employed to assist students in achieving MEAP scores that are reflective of the students' knowledge and abilities. It was not intended that this study could be generalized to the population of tenth grade students in the state of Michigan.

Implications of the Findings

This research can provide useful information related to test sophistication training and its impact on high school students' performance on mandated criterion-referenced tests. Among the consumers who can use this information are teachers, school administrators, persons developing test sophistication training programs, and persons involved with mandating criterion referenced tests.

The process of taking tests is well established within the public school environment. While students commonly experience testing in the classroom, in
Michigan they also routinely experience the MEAP tests. These are standardized, criterion-referenced tests designed to inform the State Board of Education about students' achievement of basic competencies. While it is logical that students should be familiar with the skills needed to take tests successfully, the findings of this research suggest that acquisition of test-wiseness skills does not impact performance on the MEAP tests at the tenth grade level, in a rural, culturally homogeneous setting. Teachers and school administrators should consider other strategies for helping students achieve representative scores.

The results of the study differed from what was expected. Millman et al. (1965) identified test-wiseness principles divided into two areas. First, elements independent of test purpose include time-using strategy, error-avoidance strategy, guessing strategy, and deductive reasoning strategy. Elements that depend on the test purpose include intent consideration strategy and cue-using strategy. Bangert-Drowns, Kulik, and Kulik (1983) and Samson (1985) found that positive gains for time spent on training lie in the ten to twenty hour range, with percentile point increases ranging from fourteen to seventeen points. Sarnacki (1979) cited a number of studies advocating that attention be given to lack of practice opportunities. McLellan and Craig (1989) reviewed the literature and identified strategies for use in the classroom, including practice forums, taking sample tests, and completion of a profile of their existing test-taking habits.

Based on a review of the literature, and a review of the Ace That Test test sophistication training program which indicated that its content is consistent with
the body of knowledge contained in the literature review, it was expected that enhanced test-taking skills would result in enhanced MEAP scores, and in enhanced self-assessments by students about their performance on MEAP. That such results did not occur is worthy of note.

It is important to consider that the MEAP test is a criterion-referenced test, not a timed achievement test. A review of the related literature indicated that little has been done to study the influence of test sophistication on criterion referenced tests (see p. 20).

Other possible interpretations of the results are that the MEAP test is not subject to influence by test sophistication training of students, or that the Ace That Test program does not work with the MEAP test, and possibly, other criterion-referenced tests as well. For this type of test, it may be more important to examine whether the curriculum of the school in question is aligned and implemented so that the skills tested on MEAP have been taught prior. If this has not happened, it is likely that students may lack the skills tested, and, therefore, test sophistication training would have little impact.

Recommendations

What can be recommended as a result of the findings and conclusions of the study? If testing continues to be a force with which public school educators must cope, test-wiseness programs will continue to be marketed. The continued pressure to increase student scores, coupled with the responsibility educators have, to
spend limited resources in the manner most effective for students, has implications for decision-makers. These decisions require thoughtful consideration of all ethical alternatives which can lead to scores which accurately represent the ability of the student tested. Test sophistication training is one of those options. Clear and consistent data with regard to student outcomes as a result of such programs is needed.

Many possibilities can be suggested as factors to be examined in the future. Perhaps students have not had the appropriate instruction to be able to perform the objectives tested on the MEAP. Perhaps test sophistication training is more effectively transferred to standardized test experiences when employed by students who are highly motivated to improve their scores (e.g., by paying a fee to participate in an after-school training program).

Perhaps, even in spite of the diploma endorsement, students do not value the MEAP test. Perhaps teachers do not value the MEAP test, and this message is communicated to students.

Does an endorsed diploma make a difference to students? Does an endorsed diploma make a difference to teachers? Do teachers value their own curriculum, and respond negatively to state mandates which threaten to interfere? These questions illustrate possible topics to be explored in order to reach greater understanding of the value of test sophistication with regard to the MEAP test.
General Recommendations

It is recommended that prior to any large scale purchase of test sophistication training materials, significant research and development of options be carried out by the school district considering them.

It is recommended that test sophistication researchers study the differences between criterion-referenced tests, and norm-referenced tests, and implement studies to determine program efficacy for students facing them.

It is recommended that schools wishing to achieve MEAP scores they feel are representative of the ability and learning of the students align curriculums so that the objectives tested are taught and learned on an outcomes basis prior to MEAP testing.

Future Study

It is recommended that a future study be carried out in which two groups of students receive different treatments. One group would receive the test-wiseness instruction, and the second group would be engaged in a review of the skills needed to accomplish the required objectives. Results of the group scores on MEAP could then be the dependent variable.

It is recommended that test-wiseness researchers undertake studies in which all subjects are in-school students who are required to participate in training (rather than being those who are somehow 'selected' or who volunteer) to closely
parallel the applied conditions in public schools.

It is recommended that researchers examine the role that effort plays in test-taking. Does test sophistication training enable students to take tests at a higher comfort level? Could test sophistication training make a difference for anxious students?

Research on test sophistication and criterion-referenced tests could be helpful to school officials in their efforts to help students achieve representative test scores. The present findings are not consistent with those reported from studies of test sophistication and norm-referenced tests. There are a variety of possible avenues of research which have not yet been explored with regard to the MEAP test. The growth of the body of research on this topic may contribute to the future planning, execution and evaluation of local school district curriculum. Furthermore, the findings of such research may help state-level policy makers to consider anew the tendency of state mandated testing to drive curriculum in local schools.
Appendix A

MEAP Administration Manual Information
<table>
<thead>
<tr>
<th>Table of Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction .................................................................</td>
</tr>
<tr>
<td>Students To Be Tested .......................................................</td>
</tr>
<tr>
<td>Test Scheduling, Sequence and Setting ...................................</td>
</tr>
<tr>
<td>For Tested Special Education Students .................................</td>
</tr>
<tr>
<td>Testing Materials ...............................................................</td>
</tr>
<tr>
<td>Arrangements Before Test Administration ...............................</td>
</tr>
<tr>
<td>Recording Student Identification Information ........................</td>
</tr>
<tr>
<td>Monitoring the Assessment ....................................................</td>
</tr>
<tr>
<td>Beginning the Test Administration: Grade Ten ..........................</td>
</tr>
<tr>
<td>Essential Skills Reading Test: Topic Familiarity ....................</td>
</tr>
<tr>
<td>Essential Skills Reading Test: Story Selection ........................</td>
</tr>
<tr>
<td>Essential Skills Reading Test: Informational Selection ..............</td>
</tr>
<tr>
<td>Essential Skills Mathematics Test ..........................................</td>
</tr>
<tr>
<td>Mental Arithmetic ...............................................................</td>
</tr>
<tr>
<td>Non-Calculator Computation ..................................................</td>
</tr>
<tr>
<td>Basic Skills Mathematics .....................................................</td>
</tr>
<tr>
<td>Conceptualization and Problem Solving (Session One) ...............</td>
</tr>
<tr>
<td>Conceptualization and Problem Solving (Session Two) ...............</td>
</tr>
<tr>
<td>Ending A Testing Session .....................................................</td>
</tr>
<tr>
<td>Science Test Administration: Grade Eleven ...............................</td>
</tr>
<tr>
<td>Subsequent Sessions .............................................................</td>
</tr>
<tr>
<td>Arrangements After Testing ...................................................</td>
</tr>
<tr>
<td>Test Preparation Standards ...................................................</td>
</tr>
<tr>
<td>Test Administration Standards ................................................</td>
</tr>
</tbody>
</table>
Introduction

MEAP Activities 1992-93

**MEAP Essential Skills Reading and Mathematics Tests.** To be administered to all students in grades 4, 7, and 10 except Excluded Students. The Basic Skills Mathematics Test will be available on a voluntary basis.

**MEAP Science Test.** To be administered to all students in grades 5, 8, and 11 except Excluded Students.

**MEAP Health Test.** To be administered to grades 5, 8, and 11 in a statewide sample of schools.

**Diploma Endorsement Retest Option.** Grade 11 students will have the opportunity to retake the Essential Skills Reading and/or Mathematics Test in order to satisfy the diploma endorsement requirement.

**Non-Tested Students Report.** Demographic information will be gathered on non-English speaking and Special Education students who were purposefully excluded from MEAP testing and each student who was eligible for testing, but was not tested because of parent/guardian request, student refusal, absence or other reasons.

The 1992 MEAP fall testing activities are to be conducted between September 21 and October 23. All testing and make-ups must be completed during that time. The five-week testing period is provided to allow adequate time for all test activities. Special provisions have been made for districts that will not open as scheduled. In this case, the School Coordinator will notify you when to administer the test(s).

Answer documents and Non-Tested Students Reports for complete districts will be processed upon receipt by the Scoring Service and results will be shipped within two weeks. Districts that complete testing early and submit materials promptly will receive results before those that do not complete testing until the end of the testing period. Guidelines for interpreting and using the test results will be provided when the results are returned.

This manual describes for test administrators the students to be tested, the preparations to be made prior to testing, the procedures for administering the tests, and the steps to be taken after testing to assure accurate and efficient scoring and reporting.

To assure best results, MEAP test administration procedures should be followed precisely. The test administrator is responsible for:

- coordinating the distribution, monitoring, and collection of test materials;
- administering the tests in a definite "critical" order;
- directing accurate completion of several sections (some required, others at district option) of the student answer document; and
- following the specific step by step test administration procedures described in this manual.

For best results, the test administrator should be familiar with the test booklets, answer documents, and administration directions prior to the first test session. If you have questions about test materials or test administration that are not addressed in this manual, contact the School Coordinator for clarification.

Page 4
### Students To Be Tested

All students have the right to take the MEAP tests. Students may NOT be excluded if they choose to participate or their parents'/guardians' desire them to participate.

Every student in grades 4, 5, 7, 8, 10, and 11 is to be given the opportunity to be tested. Those who meet the Students Eligible for Exclusion standards can be tested with the option of having their results excluded from the School, District and State Summary Reports. See your School Coordinator for further details.

### Test Scheduling, Sequence and Setting

**Test Scheduling.** The scheduling of testing sessions is a district/school responsibility. The schedule should be developed in keeping with MEAP options and district constraints (e.g., length of class periods, number of students involved, available resources/facilities).

The Essential Skills Mathematics and Reading Tests should be administered in the order in which they occur in the test booklets. All MEAP tests except the Mathematics Mental Arithmetic subtest are untimed and student paced. The following estimate of the testing time required is provided for scheduling purposes.

#### Grade 10

<table>
<thead>
<tr>
<th>Essential Skills Reading</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic Familiarity</td>
<td>20-30</td>
</tr>
<tr>
<td>Story Selection</td>
<td>50-60</td>
</tr>
<tr>
<td>Informational Selection</td>
<td>50-60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Essential Skills Mathematics</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental Arithmetic (Timed)</td>
<td>10</td>
</tr>
<tr>
<td>Computation</td>
<td>20</td>
</tr>
<tr>
<td>Basic Skills Mathematics*</td>
<td>30</td>
</tr>
<tr>
<td>Conceptualization and Problem Solving</td>
<td></td>
</tr>
<tr>
<td>Session 1</td>
<td>40-50</td>
</tr>
<tr>
<td>Session 2</td>
<td>40-50</td>
</tr>
</tbody>
</table>

*Optional in 1992

#### Grade 11

<table>
<thead>
<tr>
<th>Science</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>80-90</td>
</tr>
</tbody>
</table>

Students will need approximately 2.5 hours to complete testing for reading and 2.5 hours for mathematics in grade 10 and approximately 1.5 hours to complete the science test in grade 11.

### Critical Testing Sequence

The order in which the grade 10 reading test is given is critical. The Topic Familiarity items assess prior knowledge by asking the student to identify characteristics of, examples of, and relations among key concepts or ideas later encountered in the reading selections. To obtain useful information from these questions, they must be administered before either of the reading selections and students MUST NOT be permitted to return to these questions after they have begun reading either the story selection or the informational selection. It is also important that students be given enough time to read each selection and answer the related questions in one session. Attempts to complete testing quickly will undoubtedly affect student performance adversely. Time and care spent on scheduling and administration will enable students to perform at their best.
Test Setting

Arrangements for rooms and seating should be completed well in advance of test administration. The tests may be given in the regular classroom or other group setting. If more than thirty-five students will be tested in one room, the Assessment Administrator should have the assistance of a proctor.

Testing Conditions That Benefit Students

- Schedule Tuesday through Thursday sessions when possible because students tend to perform best on these days. Schedule Monday afternoon or Friday morning testing sessions only if absolutely necessary. If possible, the tests should be administered by the person(s) responsible for instruction in the subject area being tested.

- To encourage optimum student performance, NO SESSION SHOULD BE LONGER THAN ONE HOUR. Test sessions of approximately 1 or 2 hours per day over several days are preferred, using two or three sessions to administer the reading test and as many as four sessions to administer the mathematics test. Remember, Topic Familiarity must be entirely completed before either of the reading selections.

- In the past, some schools completed testing in a single day. With the increased length of the tests, this will no longer be possible. Some schools may decide, therefore, to complete testing in only two days. MEAP does NOT recommend this procedure. Student performance is likely to be affected adversely.

To help students perform well, the test administrator is expected to establish a pace that is appropriate for the students being tested. It is important that the students not rush through the sections nor compete to see who can finish first.

The following conditions should be guaranteed.

- Provide one proctor for every 35 students.
- Motivate students appropriately.
- Provide adequate lighting.
- Prevent interruptions.
- Provide appropriate work areas.
- Remove disruptive students.

NOTE: Test administrators are NOT allowed to read to students any part of the MEAP tests except designated instructions and directions.

For Tested Special Education Students

Permit a student who is receiving services in a Special Education program to be tested either in a regularly scheduled session with general education classmates, or by the Special Education teacher in the Special Education classroom to minimize any effects of the test setting. The following accommodations should be considered for handicapped students:

- Extended time (MEAP tests are untimed)
- Administer testing over more days
- Allow more frequent breaks
- Provide magnifying devices (including special microscopes)
- Provide Braille or Large Print editions of the test
- Provide special seating and/or lighting
- Test in a separate room
Large-print editions of the 1992 reading, mathematics and science tests, and Braille editions of the reading, mathematics, and science tests are available from the Library of Michigan Services for the Blind and Physically Handicapped (Phone: 517-373-1590, Monday through Friday, 8:00 a.m. - 5:00 p.m.). If needed, these materials should be ordered before September 11.

Test Administration Materials

Assessment Administration Manual, Grades 10 and 11. Each test administrator and test proctor should have an Assessment Administration Manual.

Teacher Identification Sheet. One Teacher Identification Sheet is needed for each group of students to be identified with a teacher.

No. 2 Pencils. At least one No. 2 pencil with an eraser should be available for each student to be tested, plus a few extras.

Coding Lists. Each test administrator and test proctor should have copies of all coding lists. The lists will be provided by the School Coordinator if your district has elected to use these scoring options. You may have as many as four coding lists, including:

- a Feeder School List, which will contain all the schools in your district and the official four-digit code assigned to each school by the State; and

- one or two Research Coding Lists which contain lists of research questions and the corresponding "answer" codes developed to collect district information

- a Teacher Coding list which assigns a unique two-digit number to each teacher in the school and to each class section, if results are to be returned by teacher and by section.

Reading and Mathematics Test Materials for Grade Ten

Student Assessment Booklets. The grade 10 assessments have been reorganized this year into separate test booklets for reading and mathematics. You will need one Essential Skills Reading Test booklet and one Essential Skills Mathematics Test booklet for each student to be tested plus a few extra copies.

Answer Sheets. You will need one reading answer sheet and one mathematics answer sheet for each student to be tested. Each sheet includes spaces to record student identification information and responses to questions for either the Essential Skills Reading Test or the Essential Skills Mathematics Test and the Basic Skills Mathematics Test.

Metric Ruler. You will need one ruler for each student to be tested.

Acetate grid and protractor. You will need one grid and protractor printed on clear acetate for each student to be tested.

Science Test Materials for Grade Eleven

Student Assessment Booklet. You will need one booklet for each student to be tested, plus a few extra copies.

Answer Sheets. You will need one answer sheet for each student to be tested. Each sheet contains spaces to record student identification information and responses to questions on the science test.
### Shortage of Materials

Materials should be counted and checked upon receipt to be sure that nothing was omitted and that the quantities received are sufficient. If additional materials are needed, notify your School Coordinator immediately.

### Arrangements Before Test Administration

#### Setting

Arrangements for rooms and seating should be completed well in advance of administering the tests. In most cases, the students' regular classrooms will provide the most favorable testing environment. Testing should be conducted in a room with good lighting, adequate ventilation, and freedom from noise and interruptions. The room should be large enough so that the students are not crowded. Seating should be arranged so that the students are not tempted to look at the answers of others. Comfortable seats with smooth, hard, writing surfaces large enough to accommodate a test booklet and an answer document should be provided.

#### Proctors

If it is necessary to test more than thirty-five students at one time, proctors should be used to help in the following ways: 1) distributing and collecting materials; 2) ensuring that students are on the correct page and are marking their responses properly; 3) answering questions about directions; 4) checking to make sure that students are recording their responses in the appropriate sections of their answer documents; and 5) checking to make sure that students are not making stray marks on their answer documents. Students may, however, use blank spaces in their assessment booklets for scratch paper.

Proctors should become familiar with the materials and procedures prior to testing. During testing, proctors should monitor all students in the room.

The test administrator should caution proctors against helping students find the right answers.

#### Notification

Announce to students in advance that they will participate in the testing. Tenth grade students should be informed that performance on the MEAP tests may determine their eligibility for a diploma endorsement. Announcements similar to regular school announcements should provide sufficient emphasis. Because a student's test results may have substantial consequences, however, the school may decide to notify parents of the diploma endorsement requirements prior to testing.

Teachers should be sure students understand that these tests will not affect their grades and that they may not be able to answer all of the questions. The test directions that are read to the students includes a statement encouraging students to do their best.
Identification

The following information should be entered on Side 1 of both the reading and mathematics answer sheet:

1. School name
2. District name
3. Teacher name
4. Student name
5. Student gender
6. Student birth date

The information for items 1, 2, and 3 should have been printed on the front of this manual by your School Coordinator.

Be sure to "grid" student name, birth date, and male or female.

If your district has selected an alternative procedure (other than student gridding) to complete the identification section of the student answer documents, follow the directions provided by the School Coordinator.

If your district has elected to use any of the Optional Services (Student No./Other, Research Reports I and II, Feeder School, and/or Racial-Ethnic Code), follow the directions provided by the School Coordinator. Be certain to "grid" the appropriate oval(s) in each field. If none of these options apply, leave the corresponding sections blank. An example of a properly gridded answer document appears on the back cover of the Student Assessment Booklets.

The "Special Education Category" section of the answer document should be completed only for a Special Education student who takes the test and submits an answer document to be scored. Follow the directions provided by the School Coordinator.

Recording Student Identification Information

Grades Ten and Eleven

The usefulness and validity of the assessment results depend on accurate recording of the required identification information. It is important to be sure that students have entered the required information on this portion of the answer document accurately and completely. It may be necessary for the test administrator or proctor to complete the required information for a few students. An example of a properly gridded answer document appears on the back cover of the Student Assessment Booklets.

If the identification information on the answer document has already been completed for the students, omit the directions that follow and turn to the appropriate instructions for test administration.

Passages printed in boldface and preceded by SAY are instructions to the pupils and are to be read aloud exactly as they are written. Instructions to the teacher are not in boldface and should not be read aloud. A pause is indicated by ***. Read the directions in a normal voice. Adjust your pace as necessary. Although the directions refer to an "answer folder," each test actually has a separate answer sheet.

Before beginning, print the following information on the chalkboard. The students should transfer this information to their answer documents.

School Name _____________________________________
District Name ____________________________________

Your School Coordinator should have printed this information on the front of this manual exactly as it is supposed to appear on the students' answer sheet.
Monitoring the Assessment

Allow students to work at their own pace. Walk around the room during testing to make sure that the students are progressing through the test and are not confused about what they are doing. Watch for the following things.

- As students begin taking the test, are they marking their answers in the correct section of the answer document and have all of their questions been answered?
- Is every student working in the correct section of the test booklet? Students should not be permitted to return to previously administered sections of the test.
- Is any student making a large number of erasures? Is the student confused about the directions or having some other problem?
- Is any student randomly marking answers to finish the test quickly? Remind him/her that the results are important and, if possible, indicate how they will be used.
- Is any student paying little attention to his/her test and distracting others? Perhaps he/she should be tested at another time or in a separate room.

NOTE:
- Test administrators may NOT read any part of the MEAP tests except designated instructions and directions.
- Students are encouraged to use calculators for the Conceptualization and Problem Solving portion of the Essential Skills Mathematics test. Calculator use is prohibited for the Mental Arithmetic and Computation portions of the test, and for the Basic Skills Mathematics Test.
- Students MAY mark and write in their test booklets (e.g., mathematical calculations, underlining, or highlighting reading selections, etc.), except on the Mental Arithmetic part of the Essential Skills Mathematics Test.

Please forward any comments or problems concerning the tests or test administration directions to your School Coordinator.

Beginning the Test Administration: Grade Ten

Please read all of the directions carefully before the day of the test, even if you have previously administered the MEAP tests.

These directions are to be read aloud after the students (or school staff) have finished gridding the answer sheet. If gridded in an earlier session or by school staff, distribute the appropriate answer sheet to the students now and check to be sure that each student has his/her answer sheet. Also, distribute a No. 2 pencil and test booklet to each student.

Passages that are printed in boldface and preceded by SAY are instructions to students and are to be read aloud exactly as they are written. This is essential to ensure standard administration of the tests throughout the state. Instructions to the teacher are not in boldface and are not to be read aloud. **** means to pause briefly. Read the directions in a normal voice. Adjust your pace as necessary.

SAY: Write your name on the cover of the test booklet in the upper-right hand corner.
**** Now, listen carefully as I read the directions to you. There are several parts in your test booklet. You will be given directions before each part. Read the directions carefully and follow them exactly.
Inform students whether this test will or will not be used by the district to qualify students for the Michigan diploma endorsement.

SAY: It is important that you do your best. Your performance on this test will (will not) be used to meet the Michigan diploma endorsement requirement. Choose the answer you think is best, then go on to the next item. Do not worry if there are some questions that you cannot answer. Take your time and do as well as you can.

Directions for the Essential Skills Reading Test begin on page 13 (below). Directions for the Essential Skills Mathematics Test begin on page 17.

Essential Skills Reading Test:
Topic Familiarity

The Topic familiarity section of the reading test and the Story Selection may be administered during the same session. You will need to make sure that there is enough time for the students to complete both sections during one session. Most students will need about 20-30 minutes to complete the Topic Familiarity section and 50-60 minutes to read and answer the questions for the story selection. If this is not available, end the session after the students have finished the Topic Familiarity section and administer the Story Selection in a later session.

Passages that are printed in boldface and preceded by SAY are instructions to students and are to be read aloud exactly as they are written. This is essential to ensure standard administration of the tests throughout the state. Instructions to the teacher are not in boldface and are not to be read aloud. **** means to pause briefly. Read the directions in a normal voice. Adjust your pace as necessary.

SAY: We are going to begin the reading testing. Make sure you have the Essential Skills Reading Test Booklet. Make sure your name is in the upper-right hand corner of the cover. **** Open your test booklet to page 3. **** Be sure that the words “Topic Familiarity” are printed at the top of the page.

Remember that you are to mark only one answer for each question. You should try to answer each question on the test. If you are not sure of an answer, make the best choice you can and go on to the next question. Be sure to do your best.

Read the directions for the Topic Familiarity section silently while I read them out loud.

There are three parts to this reading test and Topic Familiarity is the first part. How much you already know about a selection that you will read influences how well you can read it. The purpose of these questions is to determine how familiar you are with the topics of the selections that you will read later. Be sure you understand all the directions before you begin. You will have as much time as you need to complete these 24 Items.

Begin marking your answers on the answer folder in the area labeled “TOPIC FAMILIARITY.” Use only a Number 2 pencil to mark your answers. If you change an answer, be sure to erase the first mark completely. Mark only one answer for each question.

Now look at the Sample Items below. Sample Item 1: “Does self-governing help to describe Independence?” **** The answer choices are “yes” and “no.” **** Since self-governing is a characteristic of Independence, you would have filled in oval A for Sample Item 1.
On your answer folder, find the area labeled "Topic Familiarity" Sample Item 2. "Is a person keeping his mail to himself an example of privacy?" **** Yes. When a person wants to keep something to himself, it is considered private. Therefore, the correct answer is A. You should have filled in oval A for Sample Item 2 on your answer sheet.

Look at Sample Item 3. "Independence must come before government reform. Is this correct?" **** Since government reform could happen without Independence, you should have filled in oval B on your answer folder for Sample Item 3.

Remember, mark only one answer for each numbered question. Make a dark mark that fills the oval. If you change your answer, erase your first answer completely. Use a number 2 pencil, NOT a pen.

Now look for the answer spaces in your answer folder. The answer spaces are grouped so that whenever you start a new page in your test booklet you should also start with a new group of answer spaces. This should help you keep your place in the answer folder.

When you are done answering the Topic Familiarity questions, close your test booklet and put your pencil down so your teacher will know you have finished.

Are there any questions? ****

You may begin.

IF YOU PLAN TO CONTINUE WITH THE READING TEST FOLLOWING THE TOPIC FAMILIARITY TEST, BE SURE THAT STUDENTS WILL HAVE ENOUGH TIME TO READ THE STORY SELECTION AND RESPOND TO ALL OF THE QUESTIONS DURING THIS SESSION. Most students will need about 50 - 60 minutes to read the story and respond to the questions. If adequate time is not available, end the session and administer the story selection in the next session.

Essential Skills Reading Test:
Story Selection

If this is a new testing session, redistribute the test booklets, answer sheets and No. 2 pencils.

SAY: Check to be sure that the test booklet and answer sheet I gave you have your name on them. We are going to continue with the testing you began earlier. Please do not talk until I tell you to do so.

Open the test booklet to page 6. **** The word "Directions" should be printed on the top of the page. Read the directions silently while I read them out loud.

In this test, you will use your reading abilities. You will have as much time as you need to finish the test. You will read two reading selections in this test, one now and the other at a later time. It is important to read each selection completely. The reading selections are divided into numbered "Sections." These section numbers will help you look back to answer questions quickly and easily. If needed, you may write in your test booklet.

Begin marking your answers on the answer sheet in the area labeled "STORY SELECTION." For each question, choose the BEST answer. If you are not sure of the answer to a question, make your BEST choice and go on to the next question. If you change an answer, be sure to erase the first mark completely. Mark only one answer for each item.
Please read the story below. After you are done reading, we will go over the Sample Items on the next page. 

Allow time for the students to read "The Dip."

SAY: Look at Sample Item 1 on page 7. "How are Tick and the other kid alike?" The answer choices are: A) Both want a place of their own, B) Both like to fight, C) Both are twins, D) Both are boys. Since both Tick and the other kid wanted a place of their own, the correct answer is A. You would have filled in oval A for Sample Item 1.

Now look at Sample Item 2. "Section 2 describes the characters by" The answer choices are: A) telling about their problems, B) giving the causes of their actions, C) comparing what they look like, D) describing how they solved their problems. Since Section 2 compares the way the two characters look, you should have filled in oval C for Sample Item 2 on your answer folder.

You may look back to the reading selection any time you need to. Some of the questions will be like Sample Item 2 and will ask you to go back and reread a section to find the answer.

Another type of Item that you will answer will ask how you feel about the reading selection and answering the questions. Read each sentence and decide whether you strongly agree, agree, disagree, or strongly disagree.

Look at Sample Item 3. "It was easy for me to read the words in "The Dip." If you feel that it was very easy for you to read all the words in "The Dip," then you would select answer choice A, you strongly agree. If you feel that it was mostly easy for you to read "The Dip" and there were only one or two words which were hard for you, then you would select answer choice B, you agree. If you feel that it was a little difficult for you to read the words in "The Dip", then you would select answer choice C, you disagree with the sentence that says "The Dip" was easy for you to read. If you had a lot of difficulty reading the words in "The Dip," then you would select answer choice D, you strongly disagree with the sentence that says "The Dip" was easy for you to read.

Since answers to this type of Item will be different from one person to another, you should have marked the letter that BEST shows how you felt about reading "The Dip."

If you do not understand the directions, please raise your hand.

You will have as much time as you need to complete the story selection. When you have finished answering the story selection questions, close your test booklet and put your pencil down. Do not go on. Are there any questions?

Remember, the answer spaces are grouped so that whenever you start a new page in your test booklet you should also start with a new group of answer spaces. This should help you keep your place in the answer folder.

You may turn to page 8 and begin.

Allow about five minutes to collect and count the booklets and answer folders at the end of the test session. Turn to "Ending a Testing Session" on page 22 and follow the procedure there.

MEAP recommends that the Informational Selection portion of the Reading Test be administered in a separate test session.
Essential Skills Reading Test: Informational Selection

Students should begin the informational reading selection together. This part of the test should be administered at the beginning of a new testing session. Redistribute the students' test booklets, answer sheets and the No. 2 pencils.

SAY: Now it is time to read the second selection in the reading test. Please do not talk until I tell you to do so. I will hand out the materials. Check to be sure that I have given you the test booklet with your name on it. **** Check to see that you have your own answer sheet. **** Now I will review the directions. Do not open your test booklets until I tell you to do so. ****

The test questions you will answer today are similar to those you answered for the story selection. This time, though, you will read an informational selection.

You may wish to review the sample questions preceding the story selection on pages 6 and 7, particularly if you had students absent during the story selection test session. When the students are ready to begin the informational selection, read the following directions.

SAY: Listen carefully to the directions. Read the Informational selection carefully, then answer the questions. Notice that the selection has been marked into sections. Be sure to look back in the selection any time you want to check your answers.

Mark only one answer for each question. You should try to answer each question on the test. If you are not sure of an answer, make the best choice you can and go on to the next question. If you make a mistake or wish to change an answer, be sure to erase your first answer completely. Remember to make all of your answer marks heavy and black. Be sure to do your best.

When questions ask you to select answers "strongly agree," "agree," "disagree," and "strongly disagree" in any terms that are meaningful to the students.

The Informational selection is 6 pages long. Be sure to read the complete selection before answering the test questions that start on page 28. As you answer the questions, you may look back through the selection anytime you like.

Begin marking your answers on the answer folder in the area labeled "INFORMATIONAL SELECTION."

When you finish answering the Informational selection questions, close your test booklet and put your pencil down. DO NOT GO ON.

Remember, the answer spaces are grouped so that whenever you start a new page in your test booklet you should also start with a new group of answer spaces. This should help you keep your place in the answer folder.

Are there any questions? **** Turn to page 22 and begin reading.

Students should finish the informational reading selection at their own pace.

At the end of the test session, allow about five minutes to collect and count the assessment booklets and answer folders. Turn to "Ending a Testing Session" on Page 22 and follow the procedure there.


**Essential Skills Mathematics Test**

The directions for the Essential Skills Mathematics Test are to be read aloud after the students have finished gridding their answer sheet (directions on page 9) and the directions for "Beginning the Test Administration" (on page 12) have been read to the students.

SAMPLE ITEMS: Two Sample items are included in the directions at the beginning of each part of the Mathematics Test. These should be reviewed prior to administering each part of the test booklet.

Passages that are printed in boldface and preceded by SAY are instructions to students and are to be read aloud exactly as they are written. This is essential to insure standard administration of the tests throughout the state. Instructions to the teacher are not in boldface and are not to be read aloud. **** means to pause briefly. Read the directions in a normal voice. Adjust your pace as necessary.

**Mental Arithmetic**

SAY: We are now going to work on mathematics. Find the part of your answer sheet that is labeled MENTAL ARITHMETIC. **** Mark your answers for this section of the test in that area. Do not make any other marks on your answer folder. Be careful when you record your answers on the answer folder. If you skip a question on your test, you also have to skip it on your answer folder. Be sure that you mark the correct oval for the question you are answering. If you mark your answer in the wrong row, your answer will be scored as wrong.

Now look for the answer spaces in your answer folder. The answer spaces are grouped so that whenever you start a new page in your test booklet you should also start with a new group of answer spaces. This should help you keep your place in the answer folder.

For this section, you may NOT write in the test booklet or on the answer folder except to mark your answer. When you finish answering all the questions in the MENTAL ARITHMETIC section, close your test booklet and sit quietly. Now, turn to the directions for the MENTAL ARITHMETIC section on page 3 of your test booklet. **** Read the directions silently as I read them aloud.

This part of the test contains mental arithmetic and estimation items. You must do these problems in your head. Do not work these problems on scratch paper or in your test booklet. Use your pencil only to mark your answer on the mathematics answer sheet. If you have a calculator, please put it away. You may not use a calculator in this part of the test. Begin marking your answers on the answer sheet in the area labeled "MENTAL ARITHMETIC."

Now look at the sample items below and find the matching row of ovals on the answer folder. Sample Item 1: "Twenty-four plus five plus one minus thirteen equals" **** Repeat the question if necessary. The correct answer is seventeen. You should fill in oval A on the answer sheet because that is the letter in front of the correct answer. ****

SAY: Try Sample Item 2: "Aaron bought a pen for twenty-eight cents and a pencil for eleven cents. ABOUT how much did he spend all together?" **** Repeat the question if necessary. The correct answer is forty cents. You should mark oval D on the answer folder because that is the letter in front of the correct answer. **** Are there any questions? **** Answer questions, if any.
SAY: Do not turn the page until I tell you to do so. You will have a short time to answer the questions, so do not spend too much time on any one problem. If you are not sure of an answer, you may skip it and come back to it later if there is enough time. If you skip a question, be sure to mark the answer to the next question in the right place on the answer folder. When you finish answering the questions in this part, close your test booklet.

Remember, you may not write in the test booklet or on the answer folder except to mark the answer ovals. You will have 10 minutes to work on this part. You may turn the page and begin now.

TIME THIS SECTION. After EXACTLY 10 minutes.


Proceed with administration of the NON-CALCULATOR COMPUTATION section, next.

Non-Calculator Computation

Hand out one paper ruler and one acetate grid and protractor to each student.

SAY: Find the area of your answer sheet labeled "COMPUTATION." Mark your answers for this part of the test in that space on the answer folder. Be sure to mark the space of the letter that is your answer to the question. You may work ONLY on the questions in the NON-CALCULATOR COMPUTATION section of the test booklet.

Turn to the directions for the NON-CALCULATOR COMPUTATION section on page 7 of your test booklet. Read the directions silently as I read them aloud.

In this part, you will use paper and pencil and other tools such as the paper ruler, transparent grid and protractor to answer the questions. You may use the space in your test booklet for scratch paper. Answer all the problems by yourself. If you are not sure about the answer to a question, make your best choice and go on to the next question. You will have as much time as you need to finish this part. Work as quickly as you can, but don’t rush through this part of the test. If you do not understand the directions, raise your hand. Mark your answers on the answer sheet in the area labeled "COMPUTATION."

Answer questions, if any.

SAY: If you have a calculator, please put it away. You may not use a calculator in this part of the test. Mark only one answer for each item. Use only a pencil to mark your answers and be careful to mark each answer in the correct space on the answer folder. If you change an answer, be sure to erase the first mark completely.

Now, look at the sample items below. Sample Item 1: “One-half equals n fourths. Find n.” Repeat the question if necessary. Because the correct answer is two, you should mark oval B on the answer sheet.

Now look at Sample Item 2: “Nineteen times seventy eight equals” Repeat the question if necessary. The correct answer is one thousand, four hundred eighty-two. You should mark space C.

Do you have any questions about how to mark your answer sheet? Answer questions, if any.
SAY: Do not begin until I tell you to do so. After you answer the questions in this part, check your work. Close your test booklet when you finish.

You may begin.

After students have finished answering questions in the NON-CALCULATOR COMPUTATION part of the test, you may proceed with administration of the BASIC SKILLS MATHEMATICS section below, if your district has decided to administer this optional test. Students will need to retain the ruler for use in the BASIC SKILLS MATHEMATICS section.

Basic Skills Mathematics

In addition to base testing, districts may voluntarily administer to fourth-, seventh-, and tenth-graders a subset of the previous Basic Skills Mathematics test, at no cost to the district. This will provide local-district trend data on the basic skills test in the transition years from the previous mathematics test to the new mathematics test. Unless you are told not to do so by the School Coordinator, you must administer this part of the test.

SAY: Find the area of your answer folder labeled "BASIC SKILLS MATHEMATICS." Mark your answers for this part of the test in that area of the answer sheet. Be sure to mark only the oval containing the letter that is your answer to the question. You may work ONLY on the questions in the Basic Skills section of the test booklet.

Now, turn to the directions for the "BASIC SKILLS MATHEMATICS" section on page 15. **** Read the directions silently as I read them aloud. This part of the test is not required for all students. All of the students in our school will take this part of the test this year. You are to work these items by yourself. You will have as much time as you need to finish this part of the test. If you are not sure about the answer to a question, make your best choice and go on to the next one. Work as quickly as you can, but do not rush through this test. If you have any questions about the directions, raise your hand. **** Answer questions, if any. If you have a calculator, please put it away. You may not use a calculator in this part of the test.

SAY: If you want to work a problem you may use the space in the test book for scratch paper. Mark only one answer for each item. Use only a number 2 pencil to mark your answers. If you change an answer, be sure to erase the first mark completely. Although most items are followed by four choices marked A, B, C, D, a few are not. For these items, the extra ovals on the answer sheet should be ignored. Begin marking your answers on the answer sheet in the area labeled "BASIC SKILLS MATHEMATICS." Now look at the first sample item below.

Sample item one. "How much will three radios cost if they are priced at fifteen dollars each?" **** Since three radios would cost forty-five dollars, you would have marked oval C on your answer sheet because the letter in front of forty-five dollars is C.

Now try sample item two. "What whole number is greater than the product of six times three?" **** The correct answer is twenty, since twenty is the only number greater than eighteen. You would have filled in oval D on your answer sheet for Sample Item 2. If you have any questions about the directions, please raise your hand. **** Answer questions, if any.

SAY: Close your test booklet and put your pencil down when you finish this part.

Go on to the next page.
If you are NOT going to administer another section of the mathematics test at this time, allow approximately 5 minutes to collect and count the booklets and answer folders. Turn to "Ending a Testing Session" on page 22, and follow the procedure there.

**Essential Skills Mathematics Test: Conceptualization and Problem Solving (Session 1)**

Please allow sufficient time (approximately 40 to 50 minutes) to administer Session 1 of Conceptualization and Problem Solving in its entirety. Return to each student his/her test booklet and answer sheet. Distribute calculators or instruct students to take out their own calculator. Students may NOT share a calculator.

SAY: Check to be sure that you have your own answer folder and your own test booklet.

SAY: Be careful as you mark your answers on your answer folder. If you skip a question on your test, you also have to skip it on your answer sheet. If you mark your answer in the wrong row, the scoring machine will score your answer as wrong. Never fill in two ovals to answer one question.

Turn to the directions for CONCEPTUALIZATION AND PROBLEM SOLVING on page 21 of your test booklet. You will only work on the first part of CONCEPTUALIZATION AND PROBLEM SOLVING, THROUGH PAGE 30. Read the directions silently as I read them aloud.

In this part, you will use paper and a pencil to solve several types of problems. You may use the space in the test booklet for scratch paper. Work all the problems by yourself. If you are not sure about the answer to a question, make your best choice and go on to the next question. You will have as much time as you need to finish this part. Work as quickly as you can but do not rush through this part of the test. If you do not understand the directions, raise your hand.

Answer questions, if any. Mark your answers on the answer sheet in the area labeled "CONCEPTUALIZATION AND PROBLEM SOLVING."

SAY: Mark only one answer for each question. Be careful to mark each answer in the correct space on the answer sheet. If you change an answer, be sure to erase the first mark completely.

You may use a calculator to help you answer any of the problems in this part. A few questions ask you to find an answer using a calculator. If you do not have a calculator, you may work these problems by hand.

Now look at the sample items below and find the matching row of ovals on the answer folder.

Sample Item 1: "What is the perimeter of a rectangle measuring 10 meters by 15 meters?" Repeat the question if necessary. The correct answer is fifty meters. You should mark oval C for Sample Item 1.

Look at Sample Item 2: "If Mark scored 85% correct on his last spelling test, how many of the 20 spelling words did he get wrong?" Repeat the question if necessary. The correct answer is three. You should mark oval A for Sample Item 2.
When you reach the "STOP" page in this part, go back and check your work. Then, close your test booklet and wait for further instructions.

You may begin.

You should allow approximately 5 minutes after testing to collect completed test materials. Before collecting student answer folders and test booklets,

SAY: Check your answer folders and erase any stray marks. Place your answer folder inside the front cover of your test booklet, and pass them forward.

Conceptualization and Problem Solving (Session 2)

Please allow sufficient time (approximately 40 to 50 minutes) to administer the second portion of CONCEPTUALIZATION AND PROBLEM SOLVING in its entirety. Return to each student his/her test booklet, student answer sheet, and (if collected earlier) calculator.

SAY: Check to be sure that you have your own answer sheet and your own test booklet. **** Be sure to mark your answers for this session in the area labeled CONCEPTUALIZATION AND PROBLEM SOLVING SESSION 2 on the answer sheet, beginning with question number 44.

Be careful as you record answers on your answer sheet. If you skip a question on your test, you also have to skip it in your answer sheet. If you mark your answer in the wrong row, the scoring machine will score your answer as wrong.

For this section, you may write in the test booklet. You may use a calculator to help you answer questions in this part of the test. A few questions ask you to find an answer using a calculator. If you do not have a calculator, you may work these problems by hand. You are NOT allowed to share calculators.

The last page of the test booklet includes a question that asks you to write an explanation. Be sure to answer this question to the best of your ability. Don't forget to fill in your name at the bottom of that page.

Now turn to page 32 and begin.

When students have completed Session II,

SAY: It is time to stop now. Before we collect the test booklets, look over your answer sheet to be sure you have marked each of your answers darkly. Also be sure you have completely erased your mistakes. **** Carefully tear off the back cover from your test booklet, write your name on it and hand it in with your booklet and answer folder. Please remain quietly in your seat until all materials are collected.

The mathematics essay item handed in with the answer folder should be retained by the teacher for local scoring UNLESS the school has been notified in advance to submit their student responses for scoring. If student responses are to be sent to the scoring contractor, the teacher may wish to make a copy of the students' papers and retain them for classroom discussion.

Allow about five minutes to collect and count the booklets and answer sheets at the end of the test session. Turn to "Ending a Test Session" on page 22 and follow the procedure there.
Ending A Testing Session

About five minutes before the end of the test session, the students should stop working to allow time to collect and count the assessment booklets and answer sheets.

SAY: It is time for us to stop working on the tests for now. (We will work on them again at another time if you have not finished.) Before we collect the test booklets, look over your answer sheet to be sure you have marked each of your answers darkly. Also be sure you have completely erased your mistakes. **** Put your answer sheet inside the test booklet where you left off. Please remain quietly in your seat until all materials are collected.

After the students have put their answer documents inside their test booklets, collect the booklets and count them. The number must match the number of students. Collect the No. 2 pencils.

If the students have completed this test, set the answer documents aside for forwarding to the School Coordinator.

Science Test Administration: Grade Eleven

Please read all of the directions carefully before the day of the test, even if you have previously administered the MEAP tests.

These directions should be read aloud after the students have gridded the student identification information on their answer sheets. Passages that are printed in boldface and preceded by “SAY” are instructions to the students and are to be read aloud exactly as they are written. This is essential to ensure a standard administration of the assessment instruments throughout the state. Instructions to the teacher are not in boldface and are not to be read aloud. **** means to pause briefly. Read the directions in a normal voice. Adjust your pace as necessary.

Inform students whether this test will or will not be used by the district to qualify students for the Michigan diploma endorsement (see page 13).

Distribute an assessment booklet and a No. 2 pencil to each student.

SAY: Before we begin the test, write your full name on the front of the test booklet. **** You should try to answer each question on the test. Mark the answers by filling in the oval that contains the letter that is your answer to the question.

Be sure to make all of your answers on your answer sheet in the section labeled “SCIENCE.” Make all your answer marks heavy and black. Use a No. 2 pencil. Mark only one answer for each question. If you make a mistake or wish to change your answer, be sure to erase your first answer completely. You may write in your test booklet, but be sure to mark your answer on the answer sheet. Do not make any other marks on your answer sheet.

You will not be graded on the results of this test. However, it is important that you do your best. Work as quickly as you can, but do not rush through this test. Now, turn to the directions on page 3 of your test booklet so that you can practice marking items. **** Read the directions silently as I read them aloud.

In this test you will be using your science knowledge. Be sure you understand the directions before you begin. You will have as much time as you need to finish the test. If you are not sure about the answer to a question or an item make your best choice, and go on to the next one.
Be sure to start at number 1 on side 2 of your answer sheet. Remember to put your answer on the answer sheet. If you want to change an answer, be sure to erase the first mark completely. Now look at the sample items below.

SAY: Sample Item 1: "At what temperature does water boil?" **** Repeat the question if necessary. Since water boils at one hundred degrees Celsius, you would have marked space "B" because the letter in front of one hundred degrees Celsius is B.

Now try this sample. Mark your answer on the answer sheet in the row of spaces under Sample Item 2.

Sample Item 2: "The sun's energy provides A) water and light; B) air and heat; C) water and air; D) heat and light." **** Repeat the question if necessary. The correct answer is heat and light, so you should have filled in space D on your answer sheet for Sample Item 2.

Remember, mark only one answer for each question. If you change your answer, erase your first answer completely. Make a dark mark that fills the space. When you have finished, go back and check your work. Then close your test booklet and wait quietly for further instructions. You may begin.

About five minutes before the end of the test session the students should stop working to allow time to collect and count the assessment booklets and answer sheets.

SAY: It is time for us to stop working on the test now. We will collect the test booklets first and then the answer sheets. While we are collecting the booklets, look over your answer sheet to be sure you have marked each of your answers darkly. Also be sure you have completely erased any mistakes. **** Please remain quietly in your seat until all materials are collected.

Collect the test booklets first and then collect the answer sheets. Count the booklets and answer sheets. The number must match the number of students tested. Collect the pencils.

If students do not finish the test in one session, use the directions for "Subsequent Sessions," to start a later session.

Subsequent Sessions

Read these directions when beginning any test session after the initial session.

SAY: We are going to continue the test you began earlier. Please do not talk until I tell you to do so. We will now hand out the materials. Do not open your booklets until you are told to begin work.

Distribute the test booklets, answer sheets and No. 2 pencils. Be sure each student receives the answer sheet and test booklet with his/her name on it.

SAY: Check to be sure the test book and answer sheet I gave you have your name on them. Did each of you get your own materials? **** You should try to answer each question on the test. If you are not sure of an answer, make the best choice you can and go on to the next question. Mark only one answer for each question. If you make a mistake or wish to change an answer, be sure to erase your first answer completely. Remember to make all of your answer marks heavy and black. Be sure to do your best. Are there any questions? ****

Answer any questions the students have and then instruct them to begin work.
Arrangements After Testing

• Separate the answer documents from the test booklets.

• Inspect each answer document to be sure that the identification grids are properly completed. See the example on the back cover of the Student Assessment Booklets.

• Be sure that all response ovals are filled with a dark mark.

• Be sure that the applicable “Optional Grids” are properly completed.

• Inspect each answer document for stray marks and doodles. These must be completely erased.

• DO NOT alphabetize the answer documents. This will be done when they are processed.

• Complete a Teacher Identification Sheet. Be sure to fill in the teacher name, school name, and district name from the front of this Assessment Administration Manual. Mark grade 10 or 11 and indicate the number of answer documents you are returning. Be sure to enter and grid the teacher’s name.

• Prepare a list of all enrolled students who were not tested. List each student who was purposefully excluded (non-English speaking or special education) and every student who was eligible for testing, but not tested because of parent/guardian request, student refusal, absence, or other reason. Include the student’s name and the reason he/she was not tested. The list will be used by the School Coordinator to complete the Non-Tested Students Report(s).

• Place the completed Teacher Identification Sheet on top of the list of students who were not tested and the completed answer documents. Secure the bundle with a paper band. DO NOT use string, rubber bands or tape to secure the bundle as these may tear the materials. Return all student answer documents, including those in which students did not complete all sections of the tests.

• Keep one copy of this administration manual and a Student Assessment Booklet for each test given. These will be needed to interpret the results of the assessment when they are returned, and helpful in explaining results to students and parents.

• Return all used and unused testing materials (except your reference copies) to your School Coordinator, including unused answer documents and assessment booklets.
The purpose of the Michigan Educational Assessment Program is to promote meaningful learning by identifying essential skills that all students should possess. Quality instruction and student learning of MEAP tested objectives should be emphasized in the grades prior to MEAP testing. The test taking process then measures how well students have learned the skills.

There are several different ways to prepare students to do well on the MEAP tests. Some test preparation activities are considered more appropriate than others when judged by the standards of fairness and the students' long-term retention of skills. In general, activities designed to promote quality, long-term learning are appropriate, while those designed just to improve test scores are considered inappropriate and unethical. Appropriate and inappropriate test review practices are listed below.

It is appropriate to....

- review with all students at the beginning of the school year skills and concepts taught in previous years.
- review reading, mathematics, and science skills and concepts along with other learning areas.
- review MEAP tested objectives as part of a general overall review.
- limit the review schedule. The optimum time allotment is two or three hours.
- teach or review test-taking skills.
- have fourth graders complete the practice test. Filling out the name grid and gridding the ovals correctly is the main purpose of the practice test.

It is not appropriate to....

- teach MEAP test content which has not been taught in previous grades in the period just preceding MEAP test administration.
- review in isolation the MEAP tested mathematics and reading skills and concepts with only fourth, seventh, and/or tenth graders.
- review in isolation the MEAP tested science skills and concepts with just fifth, eighth, and eleventh graders.
- limit review to ONLY reading, mathematics, or science.
- select for review only those MEAP tested objectives on which students did poorly last year.
- call students' attention to the fact that a similar question will be on the MEAP test.
- use any current, past, or parallel MEAP test items as drill or test materials during review.
- make a few changes in the foils of MEAP test items, then use them as a "practice test."
- develop and use elaborate MEAP review materials (workbooks, worksheets, etc.).
- use computer software packages of MEAP-tested skills and concepts.
- teach MEAP tested objectives in one block of time (i.e., last few weeks of previous grade or the first few weeks of the grade being tested).
- administer mathematics, reading, or science "practice tests" in the fall just before MEAP testing.
Test Administration Standards

You should

- know which students you are responsible for testing.
- have test booklets and student answer documents ready by the first test session.
- give students sharpened No. 2 pencils with erasers, and have extras on hand.
- be sure students are comfortable but can't easily see the answers of others.
- be sure you can be heard clearly by all.
- provide students with the opportunity to have their questions answered before testing begins.
- tell students why the test is being given and how the results will be used.
- tell the students they will be given as much time as needed to finish the test, except for the timed Mental Arithmetic part of the Mathematics test.
- develop and provide some form of positive motivation for students to do their best on the test.
- encourage students to use calculators for the Conceptualization and Problem Solving portion of the Essential Skills Mathematics test. Calculator use is prohibited, however, for the Mental Arithmetic and Computation portions of the test, and the Basic Skills Mathematics Test.
- allow students to mark and/or write in their test booklets (e.g., mathematical calculations, underlining or highlighting reading selections, etc.), except for the Mental Arithmetic part of the Essential Skills Mathematics test.
- allow grade four students to mark responses in their test booklets and transfer their responses to their answer document during the same testing session (if deemed more effective).
- have schoolwork available for students who finish early.
- return all completed and incomplete answer documents to your School Coordinator.
- keep student test booklets long enough to review test results with students and parents—then return to your School Coordinator.

You should NOT

- coach students or indicate in any way (facial expression, gestures, body language, etc.) that their answer choices may be wrong, should be reconsidered, checked, or changed either during or after testing.
- answer students' factual questions from the test or about the test.
- give students the definition of terms or words used in the test, even if asked.
- read any parts of the test to students.
- give special help to poor readers during the test.
- allow anyone other than the student tested to transfer responses from a test booklet to an answer document, unless prior approval is obtained from MEAP, or a Braille or large-print test has been used.
- change any student answers.

If you have questions about these standards, please contact a MEAP staff member at (517) 373-8393.
Appendix B

Ace That Test Informational Materials
Ms. Susan Jones  
3929 Martin Road  
Kalamazoo, MI 49002  

Dear Susan:  

ACE THAT TEST is pleased to present you with the results of the Pre-Test. The enclosed report shows your level of test sophistication; it also shows your strengths and weaknesses in ten important test-taking areas:  

- Reading and following test directions  
- Understanding answer sheets and responding to them correctly  
- Using test-taking time wisely  
- Making "intelligent" answer choices on true-false, multiple-choice (many types) and matching questions  
- Understanding the guidelines for paragraph reading  
- Knowing how to approach unusual question formats  
- Knowing when and how to guess and to change answers  
- Knowing how to approach essay exams  
- Knowing how to approach teacher-made tests  
- Knowing how to prepare the evening and hours before an exam  

The fact that you took the Pre-Test shows that you know the importance of good test-taking skills. If you are not a good test taker now, don't be discouraged. Test-taking skills can be learned. Through the ACE THAT TEST training program, thousands of people have improved their test-taking skills. We have found without exception that students who take ACE THAT TEST and who attend all classes and participate actively, become better test takers!  

At this point you may be a little skeptical. You may be asking yourself if this is really true. Maybe you even believe that test-taking skills are a gift given to the chosen few. Nothing could be further from the truth!  

Test-taking skills can be learned and practiced with great success. One of the senior researchers of the ACE THAT TEST program discovered this at Newark College of Engineering in 1961 when he taught the "how to take a test" lesson in a course on how to study.  

The researchers found that students who did poorly on tests, lacked certain specific skills. Once students had acquired these skills they were able to do better on all types of tests.  

Through their research, they found that many people shared the need to improve their test-taking skills and that the ACE THAT TEST methods worked equally well whether the students were in high school, college or graduate school or were preparing for civil service or other professional
exams. Since 1972 more than five thousand people have become test wise, using the methods they learned in the ACE THAT TEST training program.

How many times have you taken a test then said to yourself afterwards, "I knew the material. I should have done better." Well, now there is a program where you can learn the skills you need. ACE THAT TEST is a comprehensive course that covers all the dynamics of test-taking skills. If you are committed to increasing your test-taking skills, we will teach you those skills.

Don't let your lack of test sophistication keep you from getting ahead in school or on the job. Join the more than 5,000 who have improved their test-taking skills through the ACE THAT TEST training program.

We will be pleased to accept your enrollment in the next ACE THAT TEST program which starts soon. Call 616/344-1234 today to register.

Sincerely,

Carol Decker
President

P.S. Get the success habit with ACE THAT TEST! When you have learned the skills taught in ACE THAT TEST, you will feel confident and your test scores will more accurately reflect what you have studied and learned. You will do better on exams. You will feel better about yourself because you will be able to do your best!
This part of the report is intended for those Pre-Test scores are 30 TOTAL SCORE or % correct.

Your test score indicates a very competent level of test sophistication. On most exams your scores probably reflect accurately all that you have learned and know. You might benefit from a brief review of test-sophistication principles and tips, but neither a teacher-directed review nor complete ACE THAT TEST program participation is necessary. Although this test did not directly measure confidence in test taking, you probably are confident in your test-taking knowledge and skills. Only if you need to develop more self assurance when taking tests, should you spend time completing the ACE THAT TEST program.

You may want to review any of the ten areas where a moderate or major focus is recommended. A teacher-directed review would be especially helpful prior to a major exam. In addition to strengthening your excellent test-taking skills, a review will reinforce and add to what you already know. These efforts may help you pick up a few extra points.

This part of the report is intended for those with Pre-Test scores of 24-29. Your test score indicates a competent level of test sophistication. Although your test scores probably accurately reflect most of what you have learned and know, a brief review of test-sophistication tips and principles could help you improve somewhat. A teacher-directed review would be very helpful, but complete ACE THAT TEST program participation is optional. Although this test did not directly measure confidence in test taking, you may need to develop more self assurance when taking tests. If you find you do not do well even when you prepare well for exams, full ACE THAT TEST program participation is recommended. Completion of the entire program should increase your confidence. You may need to repeat the teacher-directed review in the future, in order to maintain and/or increase any gains.

You may want to review any of the ten areas where a moderate or major focus is recommended. A teacher-directed review would be especially helpful prior to a major exam. In addition to strengthening your test-taking skills, which are above average, a review will reinforce what you already know. These efforts may help you pick up a few extra points, especially on long and complex competitive exams.

This part of the report is intended for those with Pre-Test scores of 19-24. Your test score indicates a fairly competent level of test sophistication. Although your test scores probably reflect accurately most of what you have learned and know, you could improve somewhat. A teacher-directed review would probably be helpful, however, complete ACE THAT TEST program participation is optional. Although this test did not directly measure confidence in test taking, you may need to develop more self assurance when taking tests. If you find you do poorly even when you prepare well for exams, full ACE THAT TEST program participation is strongly recommended. The test-taking principles and tips which you will learn should prove helpful.

If you attend all sessions of the full ACE THAT TEST program, do all exercises and ask questions when you do not understand what is presented, we can assure you that you will improve on your ACE THAT TEST Post-Test score. In the future, you may find that you need to review the materials learned in this course, if you are to maintain and/or increase your gains.

When you participate in the ACE THAT TEST program, you should review all ten areas and discover where a moderate or major focus is recommended. Doing this will help you know where
you need to pay extra attention and will help you be prepared to ask questions about whatever is not clear to you.

This part of the report is intended for those with Pre-Test scores of 18 or less. Your test score indicates that you need to improve your level of test sophistication. Although your test scores probably reflect accurately most of what have learned and know, you could improve a lot. A teacher-directed review would be of some help, but complete ACE THAT TEST program participation will be very helpful to you. Although this test did not directly measure confidence in test taking, you probably need to develop more self assurance when taking tests.

Full ACE THAT TEST program participation is strongly recommended. If you complete the entire program, you will learn test-taking tips and principles which will be very helpful to you. If you attend all sessions, do all exercises, ask questions when you do not understand what is presented, we can assure you that you will improve on your ACE THAT TEST Post-Test score. In the future, you may find that you need to review the materials learned in this course, if you are to maintain and/or increase your gains.

When you participate in the ACE THAT TEST program, you should review all ten areas and discover where a moderate or major focus is recommended. Doing this will help you know where you need to pay extra attention and will help you be prepared to ask questions about whatever is not clear.

This section will show you how you scored in the ten specific areas covered by the PRE-TEST OF TEST SOPHISTICATION.

DIRECTIONS (5 questions) 3-5 questions missed ... The test sophistication questions in this area measure what you should know about following directions. The questions you missed indicate that you have a difficult time following test directions. Learning more about following test directions will help you in obtaining exam scores that better reflect what you have studied and learned. When you cover this area in your ACE THAT TEST training, give it a major focus. Also, be alert to suggestions on ways you can continue to develop skills in this area after you have completed the program.

2 questions missed ... The pre-test questions in this area measure what you should know about following directions. The questions you missed indicate that you know some of what is necessary to follow test directions well. Learning more about following test directions would probably help you obtain scores that better reflect what you have studied and learned. When you cover this area in your ACE THAT TEST training, give it at least a moderate focus. Also, be alert to suggestions on ways you can continue to develop skills in this area after you have completed the program.

0 to 1 questions missed ... Congratulations! The test results in this area indicate that you know a lot about how to follow test directions. This is an area where you need little or no focus. However, you may be able to help others taking the program by being an active class participant when this area is covered in the ACE THAT TEST training program. By helping others you will come to master this area. This could lead to better test scores, especially on those exams for which you are prepared.
TIME USE (5 questions) 3-5 questions missed...

Your test scores in this area indicate that you need to know more about using time wisely. Although this test did not directly measure speed in taking tests, you probably are someone who at times either does not finish exams on time or rushes to finish near the end. Learning more about how to save and use time well on all types of tests could help you get higher test scores. Be fully attentive to exercises in this area of the ACE THAT TEST training and give them a major focus. Also, be sure to follow the suggestions on ways you can continue to develop skills in this area after you have completed the program.

2 questions missed...

Your test scores indicate that you have some knowledge of how to use time wisely. Although this test did not directly measure speed in taking tests, you are probably someone who could increase your pace in test taking by following ACE THAT TEST tips on time use. These suggestions can be applied when answering a wide variety of question types (from unusual multiple choice questions to paragraph reading tests to essay exams and more). When you cover this area in your ACE THAT TEST training, be attentive to the suggestions of how you can continue to improve in this area after you have completed the program. Give this area at least a moderate focus.

0-1 questions missed...

Congratulations! This is an area of test taking which you understand well. You seem to know how to use time wisely on a wide variety of exams. You probably almost always use almost all of the time available to complete an exam, but move at an optimum pace for you. This is not an area where you need help in the ACE THAT TEST training program. You need little or no focus in this area, but you can help others taking the program by passing along some of your techniques on time use to them.

MULTIPLE-CHOICE KEY WORDS (11 questions) 6 -11 questions missed...

This test measures your awareness of key words as tips to some correct and some incorrect answer choices. Your results in this area indicate that you could improve your skills in recognizing and using key words in order to make better choices in test-taking. By developing a better knowledge of key words and learning to use them as clues to the correct answers, you could learn to make better answer choices on exams. This area requires a major focus of your attention. Once you learn the key words, practice using them in the tests that you take. This can develop into a good test-sophistication habit for you.

4-5 questions missed...

This test measures your awareness of how to use key words as tips to some correct and some incorrect answer choices. Your results in this area indicate that you do not know as much as you should about recognizing and using key words to make better answer choices. You failed to recognize enough of them in the pre-test and this lowered your score. This area will require moderate focus. Once you learn these key words, practice using them in tests that you take. This can develop into a good test-sophistication habit for you.

0 - 3 questions missed...

Congratulations! This is an area of test taking which you understand well. You seem to know how to identify key words on a wide variety of exams. This is not an area where you need help in the ACE THAT TEST training program. You need little or no focus in this area, but you can help others taking the program by passing along some of your techniques to make good answer choices.
MUSTLKE-CHOICE PARAGRAPH-READING QUESTIONS (4 questions) 3
- 4 questions missed... The way you answered the questions in this area indicate that you need to know more about how to approach paragraph-reading questions. This part of the pre-test does not measure your skill in answering these types of questions. It does measure what you know about strategies for answering multiple-choice-paragraph reading questions. Since many tests you will be taking will (at least in part) contain these types of questions, you should make this area a major focus of your test-sophistication training. It is equally important that you apply what you have learned in this area to all types of tests so that you can develop better test-taking skills.

2 questions missed... The way you answered the questions in this area indicates that you know some of the important aspects of how to approach paragraph reading questions. The pre-test does not measure your skill in answering these types of questions. It does measure what you know about strategies for answering multiple-choice questions. Since many tests that you will be taking will (at least in part) contain these types of questions, learning even more than you already know about them will help. Maintain a moderate focus when you cover this area in the ACE THAT TEST program.

0 - 1 questions missed... Congratulations! This is an area of test taking which you understand very well. You seem to understand the best strategies for answering multiple-choice questions that follow reading a paragraph. This is an area where you need little or no focus in reviewing.

MUSTLKE-CHOICE-UNUSUAL QUESTION FORMATS (8 questions) 5 - 8 questions missed... The questions you missed in this area indicate that you are not as skillful as you could be in answering multiple-choice questions and that you need to learn how to approach them correctly. These questions are not like the ones found on most exams you take. You probably missed most questions posed in unfamiliar formats and you probably wasted time trying to answer them. Becoming familiar with these question formats and learning the best strategies for answering them will definitely help you in your future test taking. Pay full attention to the exercises in this area, they will require a major focus.

3 - 4 questions missed... The questions you missed in this area indicate that you do not know as much as you should about answering unusual question formats, and that you need to learn how to approach them correctly. These questions are not like the ones found on most exams you take. You probably missed many questions posed in these unfamiliar formats and you probably wasted time in trying to answer them. Becoming more familiar with question formats and learning the best strategies for answering them will definitely help you in your future test taking. Pay full attention to the exercises in this area, they will require a moderate focus. Be sure to follow the suggestions on how to identify unusual question formats and be sure to continue to practice on unusual question formats after completing the ACE THAT TEST program.

0-2 questions missed... Congratulations! This is an area of test taking which you understand well. You probably have had a number of test taking experiences where there were unusual questions. You seem to have become skilled in answering unusual multiple-choice questions without losing time and probably know how to approach many types of them. Even though you need little or no focus in this area, you probably can be helpful to others in the ACE THAT TEST training program by sharing some of the test-taking experiences where you first encountered unusual question formats.
ESSAYS (5 questions) 3 - 5 questions missed... These test questions measure your familiarity with approaches to taking essay exams. The results indicate that you still do not know the best strategies for approaching essay exams. You need to maintain a major focus when reviewing this area. However, you should realize that the review of this area in the ACE THAT TEST program will not make you a better essay writer. The program focuses on test-taking tips and strategies, and a review will not improve your writing skills. If during the course of the ACE THAT TEST program, you find that you need to improve your writing skills in order to get higher grades on essay exams, follow the suggestions for improvement that will be given to you. The primary way to improve your writing is to write, write, write.

2 questions missed... These test questions measure your familiarity with and approaches to taking essay exams. The results indicate that you probably are familiar with and know some of the best ways to approach essay exams. You probably could do better on essay exams, if you were more familiar with them and were able to approach them with a better test-taking strategy. This test does not directly measure how well you write answers to essay questions. Although this ability will be assessed in the ACE THAT TEST training program, there is not enough time in it to devote to your writing skills. If your writing needs improvement, suggestions on how to do so with the resources available to you should be followed. The primary way to improve your writing is to write, write, write.

GUESSING AND CHANGING ANSWERS (5 questions) 3 - 5 questions missed... You need to know more about when and how to guess when taking a test and about how and when to change an answer. Your test scores would be higher if you knew more about the strategies for guessing and changing answers. While the test did not directly measure your skills in these areas, it is likely that you have more often than not lowered your test score when guessing or lost points by not guessing when you should have. When you change answers, you probably change them from right to wrong more often than wrong to right. The ACE THAT TEST program will help you learn when to guess and what techniques to use such as "blind" and "hunch" guessing. Also you will have a chance to improve your skill in changing answers. When you cover this area in your ACE THAT TEST training, make it a major focus. Be sure to note how you can begin to develop your guessing strategies for exams as you prepare for them.

2 questions missed... You know some of the key aspects of when and how to guess and when and how to change answers when you are not sure of them. But you might be able to learn even more. If you want to get the highest possible score when taking a test, this area always needs your careful attention. When you take the ACE THAT TEST training program, this area will require a moderate focus. This will assure you that you will gain more points than you lose when guessing and that you will change more answers from incorrect to correct than vice versa. This test did not directly measure your skills in these related areas, but it is likely that your test scores will improve if you know more about guessing and changing answers.

0 - 1 missed... Congratulations! This is an area of test taking which you understand well. You are probably a person who gains more points than you lose by guessing even when there is a penalty for wrong answers. You are also likely to change more answers from incorrect to correct than vice versa. Apparently, this is an area where you need little or no focus. However you may be able to help others taking the program by being an active class participant when this area is covered.
TEACHER-MADE TESTS (3 questions) 2 -3 questions missed... Your test score for this section indicates you would do much better on teacher-made tests if you were able to recognize how various teachers go about making up questions for their own tests and if you learned the best ways of approaching them. If you maintain a major focus in this area during the ACE THAT TEST training program, you will learn about this and other valuable test taking tips. Since this is not a teacher-made test, it does not directly measure how well you do on teacher-made tests. In the ACE THAT TEST training program, you will learn to apply test sophistication techniques to many of the tests that are given by teachers in schools, colleges, and training programs.

1 question missed... You know most of the important aspects of how to approach tests that teachers make up for you and other students. You might even do better on these tests, if you place a moderate focus on how various teachers go about making up question for their own tests. Since this is not a teacher-made test, it does not directly measure how well you do on teacher-made tests. In the ACE THAT TEST training program, you will learn to apply test-sophistication techniques to many of the tests that are given by teachers in schools, colleges, and training programs.

0 -1 questions missed... Congratulations! You seem to know many of the important aspects of how to approach tests that your teachers make up for you and other students. However, because this is not a teacher-made test, it does not directly measure how well you actually do on teacher-made tests. In the ACE THAT TEST training program, you will learn to apply test-sophistication techniques to many of the tests given by teachers in schools, colleges, and training programs.

PREPARING (3 questions) 2 -3 missed... You could learn more about how to prepare for an exam the evening and the hours before you take it. You might do better on the tests, if you follow the ACE THAT TEST suggestions in this area. You should review with a major focus the best ways to use your time preparing for tests in terms of whether to "cram" or not, how to make sure you get "proper" food and sleep, etc. You should begin to use these techniques as soon as possible, so that they become second nature to you. If you are a student in a school setting, you should review this area immediately, as it affects all of your test-taking.

1 question missed... You know most of the important aspects of how to prepare for an exam the evening and hours before you take it. If you follow all of the suggestions for this area which you will learn in the ACE THAT TEST program, you may do better on tests. You should review with a moderate focus the best ways to use your time when preparing for tests in terms of whether to "cram" or not, how to make sure you get "proper" food and sleep, etc. You should begin to use these techniques as soon as possible, so that they become second nature to you.

0 questions missed... Congratulations! This is an area of test taking which you understand very well. You seem to understand the best strategies for preparing for tests. This is an area where you need little or no focus in reviewing.
Appendix C

Parental Consent Form
Dear Tenth Grade Student and Parents:

Our superintendent, who is doing advanced study at Western Michigan University, is looking for ways to help our students do better on tests that they must take. We are asking you to help us with her work.

We plan to have all 10th graders who take Mrs. deGraaf's English II class take a test which will show them what they already know about how to do well on tests. Then, one group of students will be given sixteen hours of instruction about how to take tests. We expect that the second group will be given this instruction at a later time. Both groups will be tested again about what they know about taking tests, and then they will be given the required Michigan Educational Assessment Program tests. We are asking permission to use test information from your student's file, which will include scores on the Differential Aptitude Test, and this year's MEAP test, as well as the tests about how to do well on tests. We will compare the results to see if the new program helps our students. If the results are positive, we will make the instruction available to all 10th graders.

We do not foresee any risks to our students. We think there will be benefits for them. These would include less worrying about taking tests, more confidence in their abilities, a higher desire to do well, and higher test scores.

We will use a coding system to protect the privacy and confidentiality of your student's test scores and records. Persons having questions about the project, and/or the rights of the students should contact Superintendent Iris Williams. The Board of Education is pleased to make this opportunity available to our students this year.

We do need your approval before your student can participate in this effort to improve instruction in the Martin Public Schools. If you agree that it is permissible for your student to take part please sign this form and return it to Mrs. deGraaf by September 4. Please be certain that there will be no penalty for any student who chooses not to allow his/her test results to be considered in the project.

I agree to allow my 10th grade student's test score information to be used as a part of the effort to improve test taking skills in the Martin Public Schools. I understand that his/her confidentiality will be protected.

Date: ______________________________________________
Signed: _____________________________________________
Student's Name: ____________________________________
Appendix D

Human Subjects Institutional Review Board
Letter of Approval
Date: September 3, 1992

To: Iris Williams

From: Mary Anne Bunda, Chair

Re: HSIRB Project Number: 92-08-06

This letter will serve as confirmation that your research protocol, "The Impact of Test Sophistication on Tenth Grade Student MEAP Scores" has been approved under the exempt category of review by the HSIRB. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the approval application.

You must seek reapproval for any changes in this design. You must also seek reapproval if the project extends beyond the termination date.

The Board wishes you success in the pursuit of your research goals.

xc: Thompson, Ed Leadership

Approval Termination: September 3, 1993
Appendix E

MEAP Copyright Permission Letter
Iris Williams  
Box 131  
Martin, Mi 49070

Dear Ms. Williams:

I am responding to your request to include the MEAP Mathematics, Reading and Science materials, as appropriate, in your dissertation. Permission is granted to include the copyrighted MEAP materials in the appendix of your dissertation. Copies maybe supplied on demand.

Don't hesitate to call if you need something else from our office. Good Luck!

Sincerely,

Diane L. Smolen
Supervisor
Michigan Educational Assessment Program

DS:wf
Appendix F

ACE THAT TEST Copyright Permission Letter
March 8, 1993

Ms. Iris Williams
Superintendent of Martin Schools
961 Lee
Martin, Michigan 49070

Dear Iris:

As the owner of the copyright of Ace That Test I understand that UMI may produce single copies of your dissertation that includes some of the Ace That Test materials.

Sincerely,

Carol Decker


