The Effects of Intensive, Structured Study on GRE Verbal and Quantitative Scores

Jan Marie Miller
Western Michigan University

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

Part of the School Psychology Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/1780

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 EFFECTS OF INTENSIVE, STRUCTURED STUDY ON GRE VERBAL AND QUANTITATIVE SCORES

by

Jan Marie Miller

A Dissertation Submitted to the Faculty of The Graduate College in partial fulfillment of the requirements for the Degree of Doctor of Philosophy Department of Psychology

Western Michigan University Kalamazoo, Michigan August 1995
THE EFFECTS OF INTENSIVE, STRUCTURED STUDY ON GRE VERBAL AND QUANTITATIVE SCORES

Jan Marie Miller, Ph.D.
Western Michigan University, 1995

This research served as a follow-up to previous research (Goodyear-Orwat and Malott, 1994) in which students engaged in self-study utilizing various study materials. Three courses (spring, summer, and fall) were offered to further examine the effects of intensive, structured study on Graduate Record Examination verbal and quantitative scores.

Twenty undergraduates enrolled in Graduate Record Examination preparatory courses studied from 66 to 105 hours. An attendance and participation contingency was established to encourage adequate study time. The courses were structured to allow students to assess and manage their own performance, utilizing standard, self-instructional texts and computerized drill programs. Students worked at their own pace based on their Individual Study Plan (ISP).

The results of the three courses are combined, because there seemed to be no reliable differences among them. The total GRE scores (quantitative and verbal combined) improved 96 points going from 842 (pretest) to 938 (posttest). The verbal GRE scores improved 39 points from 395 (pretest) to 434 (posttest). The quantitative GRE scores improved 56.5 points from 448 (pretest) to 504.5 (posttest).
The most methodologically sound published research on GRE preparation seems based on brief training (from 3 to 9 hours) and showed no significant improvements in combined GRE scores. The SAT research is ambiguous but more promising. On the other hand, the present research suggests that 66 to 105 hours of intensive, structured training can produce an average improvement of 96 points on the combined GRE scores—a socially significant accomplishment.
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.
ACKNOWLEDGMENTS

I give the honor and glory to my Lord and Savior, Jesus Christ, for His everlasting love and my reason for living.

I would like to thank my graduate advisor, Dr. Richard W. Malott for his invaluable feedback, direction, and support in the completion of this dissertation. I would like to thank the other members of my doctoral committee—Dr. Howard Farris, Dr. Dale Brethower, and Dr. Ollie Barnes, III. I would like to thank the following individuals: Diane Henderson, Brad Frieswyk, Jennifer Tairiol, Judy Honeywell, Steve Sundby, and Pam Vunovich for their involvement in this study. I extend my appreciation to all the students who participated in the study. I am grateful to my family and friends for their understanding and encouragement throughout my academic career.

I dedicate this dissertation to the memory of my mother, Margaret A. Miller, who always believed in me and never allowed me to lose sight of my goal.

Jan Marie Miller
**TABLE OF CONTENTS**

**ACKNOWLEDGMENTS** ................................................................. ii

**LIST OF TABLES** ................................................................. vi

**LIST OF FIGURES** .............................................................. vii

**CHAPTER**

**I. INTRODUCTION** .......................................................... 1

Definition of Coaching ................................................................. 1

Coaching for the GRE ................................................................. 2

Effectiveness of GRE Coaching Programs .................................... 4

Effectiveness of SAT Coaching Programs .................................... 13

Issues of Predictive Validity .................................................... 24

Issues of Social Inequality .......................................................... 38

An Intensive GRE Instructional Program ................................... 42

Summer, 1993 Course ............................................................... 43

Fall, 1993 Course ................................................................. 46

**II. SPRING, 1994 COURSE** ........................................ 47

Introduction ............................................................................... 47

Methods .................................................................................. 47

Setting, Participants, and Materials ......................................... 47

Procedure ............................................................................... 49

Results .................................................................................. 50

---

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

CHAPTER

III. SUMMER, 1994 COURSE .............................................. 53
   Introduction ................................................................. 53
   Methods ................................................................. 53
      Setting, Participants, and Materials .................. 53
      Procedure ......................................................... 54
      Results ............................................................... 54

IV. FALL, 1994 COURSE .................................................... 56
   Introduction ............................................................... 56
   Methods ................................................................. 57
      Setting, Participants, and Materials .................. 57
      Procedures ......................................................... 57
      Results ............................................................... 59

V. OVERALL RESULTS OF THE THREE 1994 COURSES ...... 63
   Student Attendance .................................................. 65
   Recruitment .......................................................... 66
   Value of the GRE Preparatory Courses .................... 67

VI. DISCUSSION .............................................................. 71
   Comparison With the Goodyear-Orwat and Malott (1994) Study ................................................. 71
   Practice Effects ...................................................... 73
## Table of Contents--Continued

### CHAPTER

- Practical Significance ................................................... 74
- Future Research ............................................................ 74

### APPENDICES

- A. Course Syllabi ............................................................ 76
- B. Informed Consent Form ............................................... 81
- C. Self-Recording Sheets ................................................. 83
- D. Daily Study Chart ....................................................... 87
- E. Sample Daily Study Sheet ........................................... 89
- F. Review Session Sheet ................................................. 91
- G. Course Evaluation Forms ............................................ 93
- H. Scores and Grades Point Averages of the 1994 Enrolled Students ......................................................... 99
- I. Human Subjects Institutional Review Board Approval ........ 102

### BIBLIOGRAPHY ............................................................... 104
LIST OF TABLES

1. Summary of Length of GRE Coaching Programs and Test Scores . . . . 3
2. Summary of Length of SAT Coaching Programs and Test Scores . . . 14
3. Length of GRE and SAT Coaching Programs ......................... 25
4. Summary of Early SAT Coaching Programs ............................ 32
5. Mean Pretest and Posttest Scores by Session for 1993 .............. 45
6. Mean Pretest and Posttest Scores by Session for 1994 .............. 51
7. Cliffs Score Increases by Area From Examination #1 to Examination #2 .................................................. 62
8. Summary of t-Test Analysis of Scores ................................. 64
9. Scores and Grade Point Averages of the 1994 Enrolled Students ............ 100
LIST OF FIGURES

1. Mean Change From Pretest to Actual GRE by Session for 1993 . . . . 45
2. Mean Change From Pretest to Actual GRE by Session for 1994 . . . . 52
3. Mean Posttest Minus Actual GRE Score ........................................... 52
4. Percentage of Student Items From Cliffs #1 to Cliffs #2 .............. 61
5. Percentage of Students With Pretest and Posttest Scores at or Above 500 for the Verbal and Quantitative Tests and at or Above 1,000 for the Combined Test ........................................... 65
6. Rating of Barron's Book ................................................................. 68
7. Rating of Princeton Book ............................................................... 68
8. Rating of Cliffs Program .............................................................. 69
9. Rating of Think Fast Program ...................................................... 69
10. Rating of ICT (Algebra) Program ............................................... 70
11. Value of GRE Course ................................................................. 70
CHAPTER I

INTRODUCTION

Definition of Coaching

The issue of "coaching" for standardized aptitude tests has gained much attention throughout history dating back to the early 1950s. Coaching programs have been developed and implemented in an effort to assist students in their preparation for standardized tests. The term coaching has been widely used to refer to a variety of test preparation activities to improve test scores. Such activities have varied in length, instructional method, and content (Wildemuth, 1983). More generally, coaching has been viewed as instructions given in preparation for taking a test in order to produce maximum performance by the individual coached (Cole, 1982). The various forms of coaching have been classified into six specific components of test preparation (Pike, 1978): (1) supplying correct answers to the test; (2) taking the test for practice; (3) maximizing motivation; (4) optimizing test anxiety; (5) instructing test wiseness; and (6) instructing test content.

In addition, a distinction has been made between (a) "instructions" (coaching) which improve scores on aptitude tests and (b) "nontest-specific learning experiences and cognitive growth," which also improve scores on aptitude tests. These latter nontest-specific learning experiences and cognitive growth refer to
the gradual development of skills over many years of experience in school and
nonschool settings (Messick, 1981). In this paper, "coaching" will refer only to
specific test preparation activities, rather than to nontest-specific instruction.
Powers (1985) contended that there is controversy and disagreement about the
extent to which formal coaching improves Graduate Record Examination (GRE)
scores. The few studies available on GRE coaching programs will be examined
to determine if such programs did result in the improvement of GRE scores.
And as a supplement, other studies will be examined to determine the effects of
coaching programs on Scholastic Aptitude Test (SAT) scores. The eight GRE
studies in the ERIC Database and the 24 SAT studies since 1968 in the ERIC
Database were selected for this analysis of the effects of coaching on test scores.

Coaching for the GRE

Of the eight GRE studies examined (Table 1), five reported improvements
in either one or two components of the GRE (Evans, 1977; Clark, 1986; Powers,
1985; Powers & Swinton, 1984; Swinton and Powers, 1983); but only three showed
statistically significant effects. The other three of the eight studies (Powers, 1987;
Powers, Clark, & Grandy, 1985; Swinton & Powers, 1985) reported no improve-
ments in GRE scores. This review will address only those five GRE studies
reporting improvement.
### Table 1

Summary of Length of GRE Coaching Programs and Test Scores

<table>
<thead>
<tr>
<th>Studies</th>
<th>Duration of coaching</th>
<th>Test Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evans (1977)</td>
<td>8 wks.</td>
<td>Small improvements for 12 of 15 classes (Q)</td>
</tr>
<tr>
<td>Clark (1986)</td>
<td>Not indicated</td>
<td>+31 pts. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+18 pts. (Q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+62 pts. (A)</td>
</tr>
<tr>
<td>Powers (1985)</td>
<td>Mean=8.0 hrs. (V)</td>
<td>-18 pts. below uncoached (V)</td>
</tr>
<tr>
<td></td>
<td>Mean=9.4 hrs. (Q)</td>
<td>-30 pts. below uncoached (Q)</td>
</tr>
<tr>
<td></td>
<td>Mean=8.0 hrs. (A)</td>
<td>-14 pts. below uncoached (A)</td>
</tr>
<tr>
<td>Powers, Clark, &amp; Grandy (1985)</td>
<td>Not indicated</td>
<td>+31 pts. (V)</td>
</tr>
<tr>
<td>Powers &amp; Swinton (1984)</td>
<td>2.9-3.37 hrs.</td>
<td>+.2 pts. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 pts. (Q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+22.1 pts. (A)</td>
</tr>
<tr>
<td>Swinton &amp; Powers (1983)</td>
<td>7 hrs.</td>
<td>-2.8 pts. below control (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-22.3 pts. below control (Q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+60.8 pts. above control (A)</td>
</tr>
<tr>
<td>Swinton &amp; Powers (1985)</td>
<td>2.67 &amp; 3.06 hrs. (V)</td>
<td>Not clearly indicated</td>
</tr>
<tr>
<td></td>
<td>2.68 &amp; 2.98 hrs. (Q)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.11 &amp; 2.33 hrs. (A)</td>
<td></td>
</tr>
</tbody>
</table>
Effectiveness of GRE Coaching Programs

GRE coaching programs varied in terms of their procedures and targeted area(s) of coaching. The five studies that will be described have been arranged based on their reports of statistically significant improvements in GRE scores. Evans (1977) examined the susceptibility of the GRE-quantitative test to short-term instruction. Additionally, the objective was to measure the differential effectiveness of the short-term instruction for black, white, and Chicano GRE candidates (total=263).

The study employed four phases: (1) feasibility, (2) exploration, (3) development, and (4) operation. The feasibility phase entailed the administration of a survey to 12 colleges with black, white, and Chicano students, in order to determine the need and interest for a mathematical instruction program. The exploratory phase included the administration of a 15-item test to examine the specific problems encountered on GRE-quantitative items. The developmental phase entailed the development of a curriculum that included testing and the utilization of lesson plans covering basic mathematical concepts. This curriculum consisted of eight two-hour sessions. Each two-hour session focused on a different component. An anxiety-reduction session focused on strategies to alleviate anxiety associated with taking admissions tests. Four sessions were devoted to instruction and strategies in basic mathematics. Instruction included lesson plans and exercises focusing on the mastery of mathematical concepts. Two of the three testing
sessions consisted of the administration of a test during the first and last sessions that were used as pretest and posttest measures. A testing session following anxiety-reduction training entailed a short (40-item) mathematics test. And finally, the operational phase involved offering the course to volunteer GRE candidates on 12 campuses.

The results suggested a variety of effects. The eight week program (total = 16 hours) resulted in positive differences between posttest and pretest scores. The type of tests used, a retired version of an actual GRE or a test specifically developed for the study, was not indicated. Twelve of the fifteen courses (treatment groups) showed consistent positive effects due to the four sessions devoted to mathematics instruction alone. The mean positive intrasubject differences from pretest (a 60-item test) to posttest (a 40-item test) ranged from 0.0078933 to 0.13959. There were no statistically significant effects attributed to the eight hours of mathematical instruction. Six of the eleven courses (treatment groups) that were represented showed positive effects due to the one anxiety reduction session alone. The mean positive intrasubject effects for the treatment groups from the anxiety reduction to posttest ranged from 0.016219 to 0.190515. There was no indication of the assessment of statistically significant effects.

There were, however, twelve courses that served as the control groups. Ten of the twelve courses showed positive effects due to the one anxiety reduction session alone. The mean positive intrasubject effects from anxiety reduction to posttest ranged from 0.0029531 to 0.10402. The effects attributed to the anxiety
reduction session were statistically significant. These small effects of both groups, however, failed to reveal much about the extent of improvement. There is no clear indication of the magnitude of those intrasubject gains. Furthermore, there was a 20-point difference in the number of items covered on the tests, which may actually account for the mean differences rather than the mathematical instruction or anxiety session. The author noted, without presenting any data, that no single sex or ethnic group showed greater gains as a result of instruction. These results lend little support to the effects of the coaching program.

Powers and Swinton (1984) focused on whether GRE test takers benefited as much from independent study of instructional material as from formal instruction from a teacher ("instructor-based test preparation"). The authors also looked at the effects of encouragement on the preparation of tests. Encouragement entailed mailing a letter encouraging candidates from the treatment group to use the package of test preparation materials that had accompanied the letter. The control group received no letter of encouragement with their test preparation materials.

They studied three analytical types: (1) analysis of explanations, (2) logical diagrams, and (3) analytical reasoning. The three specific test preparation features used were: (1) test practice, (2) feedback or knowledge of results, and (3) test-taking strategies for each item type. Practice tests included the administration of two 50-minute analytical portions of the GRE that had been included in the 1980-81 GRE Information Bulletin. Feedback or knowledge of results
entailed explanations of the answers of the two practice tests. The test-taking strategies feature included a set of tips and strategies for answering the three GRE analytical item types noted above. Participants were mailed a questionnaire following the administration of the GRE to assess the use of the materials sent.

There were no statistically significant differences in the mean score for the group that received no encouragement versus the group that did for the verbal test (477.0 vs. 477.2), or the quantitative test (494.5 vs. 492.5). However, there was a small (22.1 point) difference in mean analytical scores (509.7 vs. 531.8) (this was the area addressed by the training materials). There were no statistically significant differences between both groups in the time spent preparing for the verbal and quantitative sections. The mean number of hours spent on the verbal section for the group that received encouragement compared to the group that did not was 2.90 versus 2.89 hours, and the mean number of hours spent on the quantitative section was 3.14 versus 3.04 hours. The difference in the mean number of hours spent on the analytical section between the two groups was statistically significant. The mean number of hours spent on the analytical section for the group that received encouragement compared to the group that did not was 3.37 versus 2.80 hours.

The authors said the effect of the letter of encouragement was due to improvements on: (a) analysis of explanations, and (b) logical diagrams. Based on their results, the authors, provided a further explanation for the increases in those two analytical item types by stating that such increases were due to
"heightened test sophistication" because of the examinees' increased familiarization with those specific item formats, rather than to improvements in their analytical abilities.

Though the authors received information from the test takers on their use of materials, the reliability of their responses is in question as is the generality of those responses because those who replied (78.1%) may have been a non-random sample of the population. Nonetheless, information from the questionnaire was used to conduct an analysis to determine the use of the materials sent.

The overall objective was to determine if independent study of instructional materials could be as beneficial as instructor-led study. However, in the absence of comparative data from an instructor-led study group, we can make no conclusions.

Swinton and Powers (1983) engaged in a nonrandomized study to develop and test the effects of a brief curriculum of special preparation for the analytical portion of the GRE. Their goal was to determine if the special preparation, focusing on practice and familiarization with analytical strategies and techniques, rather than on the development of specific analytical abilities, could improve performance on three analytical item types. The structure of the course entailed a total of 7 contact hours focusing on the analytical section of the test prior to the October test administration. Two analytical practice tests were used. One was obtained from the GRE Information Bulletin and the other was especially designed for the participants. A set of explanations was also developed to assist the
participants in arriving at the correct answers.

The analytical differences between the treatment and control groups were statistically significant. The mean analytical score of the treatment group (self-selected GRE candidates) was 591.5 compared to 530.7 for the control group (those who took the test on the same day as the treatment group). Those students in the treatment group (total=25) scored significantly higher than the control group (total=415) on two of the analytical sections: (1) analysis of explanation (mean= 28.6 compared to 24.2 for the control group), and (2) logical diagrams (mean= 12.1 compared to 10.7 for the control group). The authors suggested that the significantly higher overall score could be attributed to those two analytical sections.

The mean scores of the treatment group were slightly lower than for the control group on both the verbal section (495.5 vs. 498.3) and the quantitative section (513.2 vs. 535.5). Therefore, the authors said that the lower verbal and quantitative scores for the treatment group suggested that the treatment group would also have performed more poorly on the analytical test, if they had not had the special coaching. Presumably, they volunteered for the coaching because they were most in need of it.

The authors conclude that the improvements on the analytical section were possible with a short (7 hour) intervention with a primary focus on practice and familiarization with analytical strategies and techniques. The authors noted that other factors such as the pacing and guessing strategies as well as the instructions
outlined in the GRE Information Bulletin could be learned by any candidate if he or she actually took the time to prepare in this manner.

Clark (1986) compared data from two different groups of GRE test takers. One group consisted of all test takers (total=210,000) who took the GRE during the 1979-80 school year and the other group consisted of a primary study sample (total=433) of test takers who had taken the GRE before and who repeated the GRE in a 1980 administration of the test. Data were based on information obtained from the GRE administration files, data provided by institutional users, and questionnaires completed by those who repeated the GRE. Those who repeated the test within three years (the younger group) of the first administration did so to improve their scores; whereas, those who repeated the test after three years (the older group) did so at the request of a graduate school that required more recent scores. Rather than assess the effects of a specific coaching program, the objective was to determine what factors affected the performance of the test repeaters.

The mean scores of the 210,000 test takers in the 1979-80 group were 487 for the verbal, 516 for the quantitative, and 508 for the analytical. The mean change in scores of the test repeaters from the first to the second test was +31 points, from 453 to 484 (verbal), +18 points, from 460 to 478 (quantitative), and +62 points, from 438 to 500 (analytical). The group of repeaters attempting to improve their scores reported spending more time in review and more use of study aids. Those who repeated the test at the request of a graduate school
reported spending less time preparing for the test even though they did report spending some time reading to increase particular skills. No specific study times were indicated nor was there any statistically significant effects reported by the author.

Therefore, this analysis is of little value because there was no indication of statistically significant effects, no indication of the magnitude of the differences in preparation times between the two groups of test repeaters, and no indication of differential improvement between the two groups of test repeaters. Moreover, the accuracy of the information obtained from the questionnaire presents a confound in the results. There was no indication of the accuracy or whether the questionnaires returned were representative of the group being assessed. The possibility of a test-retest effect on the scores received may also have confounded the results making it impossible to accurately determine the factors that affect test performance. The information provided by the author does not achieve the goal of determining any specific factors that may have affected the performance of the test repeaters.

Powers (1985) surveyed GRE test takers to determine whether they had participated in any formal coaching program and to determine the length of such coaching programs. Of those surveyed, only a relatively small proportion of GRE candidates (about 3%) obtained formal coaching. He compared the uncoached (total=3,162) with the coached (total=259) examinees who took the GRE during the 1979-80 testing year.
In spite of the fact that the coached examinees reported more hours than
the uncoached examinees in preparation for the test, they obtained lower average
scores on each section of the GRE than did the uncoached test takers. However,
the reliability of their involvement was not determined as the information was
based on self reports. The mean scores of the coached and uncoached groups
were 474 vs. 492 for the verbal, 475 vs. 505 for the quantitative, and 526 vs. 540
for the analytical.

The mean number of hours the coached examinees spent preparing was 8.0
for the verbal, 9.4 for the quantitative, and 8.0 for the analytical. Although the
mean scores for the coached examinees were lower than the uncoached exami­
nees, the effects of the length of coaching in relation to the analytical scores were
statistically significant. There were, however, no statistically significant effects
of the length of the program in relation to the overall verbal and quantitative
scores.

But an analysis of item types revealed that the duration of the coaching
programs appeared to affect performance on some specific types of items: (a) the
discrete quantitative items (those covering basic arithmetic operations, elementary
algebra, and plane geometry); and (b) the analysis of explanation and logical
diagrams (fixed-format response options found in the analytical section).

Powers suggested that other variables, such as the actual amount of time
preparing, regardless of the nominal length of the program, should also be taken
into consideration. Moreover, Powers suggested another explanatory variable for
the differences in quantitative scores of examinees from one of the commercial schools; the author offered two possible explanations: (1) that the enrollees were a highly selected group, or (2) that the increase in the quantitative sections were actually attributed to the coaching programs.

The manner of self-selection of whether or not to be coached establishes a confound. It is possible that those who would have done poorly anyway were the ones who participated in the coaching. The author did acknowledge that there may have been other examinee variables that could have affected test scores rather than longer coaching programs as opposed to shorter ones, thus affecting his confidence in the validity of the findings. Pretest scores as well as any differences between pretest and posttest scores as a result of coaching, were not determined, nor were there treatment and control groups, thus, making it impossible to conclude what the effects of the length of the coaching program were on test scores.

Effectiveness of SAT Coaching Programs

Of the 24 SAT studies examined, 21 reported improvements in overall scores or scores on specific components; whereas, two reported no improvements on SAT scores as a result of training, and the results of one were not clearly indicated (Table 2). As shown, most of the studies reported improvements, some statistically significant, on specific components rather than on overall SAT scores. Only four of the SAT studies reporting statistical significance were
### Table 2

Summary of Length of SAT Coaching Programs and Test Scores

<table>
<thead>
<tr>
<th>Studies</th>
<th>Duration of Coaching</th>
<th>Test Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alderman &amp; Powers (1980)</td>
<td>3-10 wks. (5-45 hrs.)</td>
<td>+10.53 pts. (V)</td>
</tr>
<tr>
<td>Chaika (1985)</td>
<td>9 days</td>
<td>+200 pts. above '84 SAT scores</td>
</tr>
<tr>
<td>Coffin (1987)</td>
<td>2 1/2 wks. 6 wks.</td>
<td>-22 pts. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+51 pts. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+20 pts. (Q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+40 pts. (Q)</td>
</tr>
<tr>
<td>Evans &amp; Pike (1973)</td>
<td>21 hrs.</td>
<td>+29 &amp; +57 pts. (QC)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+37 &amp; +52 pts. (DS)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+43 &amp; +68 pts. (RM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+25 pts. (Q)</td>
</tr>
<tr>
<td>Harris &amp; Rohfeld (1983)</td>
<td>4 wks.</td>
<td>Not indicated</td>
</tr>
<tr>
<td>Johnson &amp; Wallace (1989)</td>
<td>7 wks. (30 hrs.)</td>
<td>+74.85 pts. 1st gp. (Q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+19.36 pts. 2nd gp. (Q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+22.75 pts. 3rd gp. (Q)</td>
</tr>
<tr>
<td>Kouzakanani, Llabre, &amp; Baldwin (1989)</td>
<td>6 wks.</td>
<td>Not clearly indicated</td>
</tr>
<tr>
<td>Lai &amp; Saka (1993)</td>
<td>1 hr.</td>
<td>-18 pts. (V)</td>
</tr>
<tr>
<td>McPhail (1975)</td>
<td>15 wks.</td>
<td>Not clearly indicated</td>
</tr>
</tbody>
</table>
Table 2--Continued

<table>
<thead>
<tr>
<th>Studies</th>
<th>Duration of Coaching</th>
<th>Test Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payne (1992)</td>
<td>Not indicated</td>
<td>&quot;Ability&quot; variable greatly affected V &amp; Q scores; no test scores provided</td>
</tr>
<tr>
<td>Pennock-Roman, Powers, &amp; Perez (1989)</td>
<td>18 wks. (1 sem.)</td>
<td>Not indicated</td>
</tr>
<tr>
<td>Powers &amp; Alderman (1983)</td>
<td>Median =3.2 hrs.</td>
<td>+6.9 pts. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+3.1 pts. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-3.2 pts. (Q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+7.3 pts. (Q)</td>
</tr>
<tr>
<td>Reynolds &amp; Oberman (1987)</td>
<td>63 hrs.</td>
<td>-0.4 pts. below comparison gp. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+4.6 pts. above comparison gp. (Q)</td>
</tr>
<tr>
<td>Reynolds, Oberman &amp; Perlman (1988)</td>
<td>64 hrs.</td>
<td>-1.9 pts. below comparison gp. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+4.6 pts. above comparison gp. (Q)</td>
</tr>
<tr>
<td>Sesnowitz, Bernhardt &amp; Knain (1982)</td>
<td>10 wks. (40 hrs.)</td>
<td>+28 pts. School A (V)</td>
</tr>
<tr>
<td></td>
<td>10 wks. (24 hrs.)</td>
<td>+1.85 pts. School B (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+24 pts. School A (Q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+4.2 pts. School B (Q)</td>
</tr>
<tr>
<td>Smyth (1989)</td>
<td>Not indicated</td>
<td>+6 pts. (V)</td>
</tr>
<tr>
<td>Smyth (1990)</td>
<td>Not indicated</td>
<td>+6 pts. (V)</td>
</tr>
<tr>
<td>Teague (1992)</td>
<td>12 wks.</td>
<td>+101 pts. (V)</td>
</tr>
</tbody>
</table>
### Table 2--Continued

<table>
<thead>
<tr>
<th>Studies</th>
<th>Duration of Coaching</th>
<th>Test Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whitla (1988)</td>
<td>Not indicated</td>
<td>+11 pts. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+16 pts. (Q)</td>
</tr>
<tr>
<td>Wiggins (1992)</td>
<td>6 hrs.</td>
<td>+88 pts. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+83 pts. (Q)</td>
</tr>
<tr>
<td>Wing, Childs &amp; Maxwell (1989)</td>
<td>2 wks. 3 wks.</td>
<td>+26.37 pts. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+25.43 pts. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+55.37 pts. (Q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+49.13 pts. (Q)</td>
</tr>
<tr>
<td>Worsham &amp; Austin (1983)</td>
<td>3 sem.</td>
<td>+42 pts. (V)</td>
</tr>
<tr>
<td>Zuman (1988)</td>
<td>8 wks. 9 wks.</td>
<td>+41 pts. regular gp. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+31 pts. scholar gp. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+40 pts. regular gp. (Q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+71 pts. scholar gp. (Q)</td>
</tr>
</tbody>
</table>

Methodologically sound in that they involved randomly assigned participants and specified the duration of coaching (Alderman & Powers, 1980; Evans & Pike, 1973; Johnson & Wallace, 1989; Zuman, 1988). Therefore, only these four will be reviewed here.

Alderman and Powers (1980) conducted a study to determine the effects of special preparation on the SAT verbal section. The schools where this study
was conducted were selected based on their existing special preparation programs that were at least five hours or longer and that utilized commercial review books or teacher-made materials to improve SAT verbal scores. Students from each of the schools (5 public and 3 private high schools) were randomly assigned to either the treatment or the control group. The special preparation programs ranged in duration from 5 to 45 hours (3 to 10 weeks) focusing on four verbal item formats: (1) analogies, (2) antonyms, (3) reading comprehension, and (4) sentence completion.

The students’ (total=559) Preliminary Scholastic Aptitude Test (PSAT) verbal scores were used as their pretest scores. The average PSAT verbal score for the treatment groups was 43.86 and 43.76 for the control groups. The mean posttest SAT verbal score was 455.90 for the eight treatment groups and 445.37 for the eight control groups, equating to an overall treatment effect of 10 points. Statistically significant effects were determined for each of the four item formats. The effects of analogies and antonyms were statistically significant, while the effects of reading comprehension and sentence completion were not. Therefore, the authors attributed the small overall SAT verbal increases to students’ performance on the analogies and antonym items. Moreover, they suggested that the schools’ regular practices may be more effective in increasing SAT scores than special preparation programs. But in spite of the statistically significant effects on two of the four item formats, the practical significance of a 10-point gain is of questionable practical value.
Evans and Pike (1973) conducted a study to determine the susceptibility of performance on three mathematics aptitude item formats to special instruction. The study entailed the solicitation of volunteers from 12 high schools to participate in an instructional program. Subjects (total=417) were randomly assigned to an experimental or control group. During the experimental period for each school, there were two experimental groups and one control group. One experimental group received special instruction in the Quantitative Comparisons (QC) format, while the other group received special training in either the Regular Mathematics (RM) or Data Sufficiency (DS) format. The one control group received no special instruction. The experimental period entailed 7 instruction sessions totaling 21 hours. Special instruction included six workbooks with 3 or 4 mathematical drills in the appropriate item format.

Participants took a pretest, posttest, and a third SAT during an April test administration. As will be seen, scores by the instructed groups were higher than those obtained by the uninstructed group in all three item formats. The mean SAT-quantitative pretest score of the QC-trained group was 415 with a mean change of +29 and +57 points on the second and third SAT administrations. The mean SAT-quantitative pretest score of the DS-trained group was 401 with a mean change of +37 and +52 points on the two subsequent tests. The mean SAT-quantitative pretest score of the RM-trained group was 426 with a mean change of +43 and +68 points on the following two tests. The mean quantitative pretest score of the control group was 392 with a mean change of +18 and +47
points on the second and third tests. The improvements from the pretest to each of the two posttests for all three item formats were statistically significant; this shows the susceptibility of the SAT quantitative section to special instruction.

Though the results showed statistically significant improvements for all three item formats, the authors noted that the differences in the posttest scores between the three format-training groups could have been due to the differences in the nature of each of the three item formats. However, they stated that the SAT quantitative section contained twice as many RM as DS items and contained no QC items. Therefore, it should be of no surprise that the RM-trained group scored higher than the other two groups, and that the QC-trained group obtained the lowest mean changes. The differences could also have been attributed to practice effects, as the examinees took the SAT a total of three times.

It should be noted, however, that the authors do not indicate whether these group differences are statistically significant. Also, there was no reported assessment of statistical significance of the scores received by the control group on its three successive tests; between the treatment and control groups; or between the three instructional groups. Thus, special preparation may not have been the only reason for the score increases; it could have been simply a result of repeated testing.

Johnson and Wallace (1989) conducted an exploratory study to assess the effects of coaching on the SAT quantitative items among black urban students. The study entailed the recruitment of black high school juniors and seniors from
low-income families in three urban areas. Students from only one of the three areas were randomly assigned to groups. The details of this random assignment were not provided.

The instructional program consisted of sessions lasting 2 1/2 to 3 hours over a period of 7 weeks (total=30 hours). Students (total=116) used materials developed by the National Association of Secondary School Principals (NASSP). The sessions were devoted to the review of both the quantitative and verbal sections, test anxiety reduction, the politics of testing, and test-taking strategies and skills.

The results were only provided for the quantitative subtest scores. The subtest contained 120 items in the four retired SAT test forms used. Students were categorized into three groups based on their pretest quantitative scores. The mean gain in points for students (n=33) whose pretest score was less than 300 points was 74.85. The mean gain in points for students (n=47) whose pretest score was from 300 to 400 points was 19.36. The mean gain in points for students (n=36) whose pretest score was greater than 400 points was 22.75. The gains for all three groups were statistically significant. The largest gain was obtained from the group with pretest scores that were less than 300 points. The authors contend that performance on quantitative SAT items can be improved, and that such programs should be extended to major urban areas with large at-risk populations.

The effectiveness of the program was limited in that it did not provide any pretest means other than ranges (e.g., below 300, 300-400, and 400+ points) that
could have been compared with posttest means. The absence of random assignment of participants from all three urban areas may have contributed to the differences in the scores. In other words, there may have been examinee variances not controlled for that affected the differences in scores rather than the effects of special instruction.

Zuman's (1988) study focused on the effects of a commercial coaching program on the performance of two different groups of participants. One group consisted of a sample of the "regular" clientele of the coaching school, while the other group consisted of economically and ethnic minority students. The former group paid to attend a coaching school and the latter group received scholarships to attend.

The regular and scholarship students were randomly assigned to treatment and control groups. A total of 88 students participated in the study. Training for both groups differed in terms of the administration of tests and the duration of instruction. There is no indication of the control groups' involvement during the time of the treatment groups' training. The regular students, who met for 9 weeks, were administered three practice tests during the course in a simulated SAT-type environment, received instruction in Mathematics and English SAT items, and discussed techniques and strategies. The scholarship students met for only 8 weeks and were not given practice tests in a simulated SAT-type environment. Instead, they were administered three practice tests, two of which were taken outside of the course, while one was taken during the course. But like the
regular students, they also received instruction in Mathematics and English SAT items, and discussed techniques and strategies. The authors noted that the scholarship students did not receive as many hours of instruction as the regular students due to their late arrival times. These students had to commute from other Brooklyn and Manhattan locations.

The regular students from the treatment group improved their mean verbal score by 41 points (from 520 to 561) and their mean math score by 40 points (from 580 to 620). The scholarship students from the treatment group improved their mean verbal score by 31 points (from 344 to 375) and their mean math scores by 71 points (from 374 to 446). The mean verbal and math scores of the regular students from the control group dropped (-7 and -16 points, respectively); whereas, the mean verbal and math scores of the scholarship students from the control group improved (+31 and +15 points, respectively). Both control groups were subsequently coached and improved their scores in both areas. The regular students improved their mean verbal score by 39 points (from 503 to 542) and their mean math score by 87 points (from 549 to 636). The scholarship students improved their mean verbal score by 4 points (from 385 to 389) and their mean math score by 80 points (from 383 to 463).

Statistically significant effects were also determined. The coaching effects of the regular students from the treatment group were statistically significant for the verbal and math, but could not be determined for the control group that subsequently received coaching. The coaching effects of the scholarship students in
both groups were statistically significant for math but not for the verbal. The author attributed the differential effects of the scholarship students to the lack of training experience to teach and meet the needs of the minority students.

There were other factors that may have contributed to the differences in scores obtained between the regular and scholarship students. One major factor was the differential availability of instruction received by the scholarship students. The scholarship students had less instruction time and were tested under different conditions than the regular treatment group. Differences were already apparent before making any comparisons in scores between the two groups.

Though students were randomly assigned to treatment and control groups, comparisons continued to be made between the regular and scholarship students as if examinee differences had been controlled for by way of random assignment. The author noted that the instructors were not adequately prepared to teach these scholarship students, and that the instructors had to change their method of teaching, perhaps attesting to the fact that participant differences such as skill levels or deficits were already apparent prior to enrolling in the commercial coaching program. Therefore, although improvements within groups are clearly attributable to the coaching, it is difficult to analyze the differential effectiveness of coaching between the regular and scholarship groups.

The studies just reviewed focused primarily on the examination of coaching programs in raising students' GRE and SAT scores. The majority of the researchers reported improvements, while a few (not analyzed here) reported no
significant improvements as a result of the coaching programs. Table 3 shows duration of coaching and test scores for the GRE and SAT studies analyzed in this document. Data from the last three studies listed suggest that if a substantial number of hours is spent preparing, a substantial increase in test scores will occur; whereas, data from the first six studies do not as effectively address the relationship between the number of hours spent preparing and the effects of coaching.

Despite the large body of research showing various types of coaching programs and various amounts of improvements in GRE and SAT scores, there remains much controversy on the issue of whether the improvements affect the predictive validity of the test scores, that is, the extent to which the improved test scores reflect what the tests purport to measure or predict. Opponents have noted that coaching programs are ineffective in raising scores that purport to measure "fixed characteristics" developed over an individual's lifetime. Supporters, on the other hand, have contended that coaching can, indeed, raise a student's scores also and that the scores are not a good indicator of a student's ability to succeed in college. These opposing views of test validity will be summarized in terms of the potential impact of coaching programs on students' abilities and academic success.

Issues of Predictive Validity

Cole's (1982) concerns about test validity are paraphrased as follows: (a) students have a set of verbal, quantitative, and analytical abilities that are fixed
<table>
<thead>
<tr>
<th>Studies</th>
<th>Duration of Coaching</th>
<th>Test Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evans (1977) (GRE)</td>
<td>8 wks.</td>
<td>Small improvements for 12 of 15 classes (Q)</td>
</tr>
<tr>
<td>Clark (1986) (GRE)</td>
<td>Not indicated</td>
<td>+31 pts. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+18 pts. (Q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+61 pts. (A)</td>
</tr>
<tr>
<td>Powers (1985) (GRE)</td>
<td>Mean = 8.0 hrs. (V)</td>
<td>-18 pts. below uncoached (V)</td>
</tr>
<tr>
<td></td>
<td>Mean = 9.4 hrs. (Q)</td>
<td>-30 pts. below uncoached (Q)</td>
</tr>
<tr>
<td></td>
<td>Mean = 8.0 hrs. (A)</td>
<td>-14 pts. below uncoached (A)</td>
</tr>
<tr>
<td>Powers &amp; Swinton (1984) (GRE)</td>
<td>2.9-3.37 hrs.</td>
<td>+.2 pts. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 pts. (Q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+22.1 pts. (A)</td>
</tr>
<tr>
<td>Swinton &amp; Powers (1983) (GRE)</td>
<td>7 hrs.</td>
<td>-2.8 pts. below control (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-22.3 pts. below control (Q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+60.8 pts. above control (A)</td>
</tr>
<tr>
<td>Alderman &amp; Powers (1980) (SAT)</td>
<td>3-10 wks. (5-45 hrs.)</td>
<td>+10.53 pts. for trmt. gp. (V)</td>
</tr>
<tr>
<td>Evans &amp; Pike (1973) (SAT)</td>
<td>21 hrs.</td>
<td>+29 &amp; +57 pts. (QC) gp.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+37 &amp; +52 pts. (DS) gp.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+43 &amp; +65 pts. (RM) gp.</td>
</tr>
</tbody>
</table>
Table 3--Continued

<table>
<thead>
<tr>
<th>Studies</th>
<th>Duration of Coaching</th>
<th>Test Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnson &amp; Wallace</td>
<td>7 wks. (30 hrs.)</td>
<td>+74.85 pts. 1st gp. (Q)</td>
</tr>
<tr>
<td>(1989) (SAT)</td>
<td></td>
<td>+19.36 pts. 2nd gp. (Q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+22.75 pts. 3rd gp. (Q)</td>
</tr>
<tr>
<td>Zuman (1988)</td>
<td>8 &amp; 9 wks.</td>
<td>+41 pts. regular gp. (V)</td>
</tr>
<tr>
<td>(SAT)</td>
<td></td>
<td>+31 pts. scholar. gp. (V)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+40 pts. regular gp. (Q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+71 pts. scholar. gp. (Q)</td>
</tr>
</tbody>
</table>

and stable and are developed throughout a lifetime; (b) these abilities are important determinants of academic success; (c) standardized tests such as the GRE and SAT measure these abilities; (d) even if coaching programs can improve the test scores to a level of practical significance, such programs cannot affect those fixed and stable abilities; (e) therefore, these coaching programs cannot improve academic success; and (f) therefore, effective coaching programs would harm the predictive validity of those tests.

Cole seems to represent the testing establishment in her assumptions and concerns. Coaching programs are a major concern of the testing establishment.
Their assumption is that if those programs are effective, their use would hurt the predictive validity of the standardized tests. And those tests are the bread and butter of this establishment. With this in mind, we will look at efforts to assess the impact of coaching on predictive validity.

The efforts to assess predictive validity have been mainly in terms of attempts to show that coaching has no practically significant effects and therefore cannot harm predictive validity. No one seems to have empirically addressed the issue of whether coaching-induced improvements in standardized test scores correlate with improvements in academic success.

The SAT literature contains an ongoing debate about whether test-score gains attributed to coaching programs can affect the predictive validity of SAT scores. The seven SAT studies conducted in the 1950s and 1960s (as cited by the College Entrance Examination Board, 1968) were examined to illustrate the assertion that score gains reportedly attributed to SAT coaching programs were too small to affect the predictive validity of the SAT scores. Moreover, the Educational Testing Services (ETS) and the College Board have consistently been cited throughout the literature as being adamantly opposed to the notion that coaching programs can raise test scores in any practically significant manner. Coaching has, according to the College Board, resulted in gains averaging 10-25 points on the SAT 200-800 point scale. These point gains equate to only 1 to 3 test items and are thus relatively small.

However, in considering these studies, it should be noted that all seven
studies entailed non-randomized assignment of participants to coached and uncoached groups. Each of the seven studies will be described briefly.

The Dyer study involved two independent preparatory schools for boys; the goal was to examine the effects of practice exercises on students' quantitative and verbal SAT scores. English teachers did the coaching for the verbal sections and mathematics teachers did the coaching for the mathematical sections. The average SAT-verbal score increase was 4.6 points higher for the coached boys than for the uncoached boys. The coached group received a score increase of 12.9 points higher on the math section than the uncoached group. The difference in the verbal score increases between the coached and uncoached groups was not statistically significant. The difference in the quantitative score increases between the coached and uncoached groups was statistically significant.

The first experiment conducted by French examined the effects of SAT coaching on public school students. Three public schools were selected to participate in the study. One school was selected as the control group that received no coaching. The other two schools served as the treatment groups. Students from one treatment group received coaching on the verbal section only, while students from the second treatment group received coaching on both the verbal and mathematical sections. The largest average gains of the coached groups were about 18 points for both the verbal and mathematical sections. The College Board equated this increase to a little more than 1 point on a standard classroom scale of 60 to 100 (this equation from the SAT to the classroom scale is unclear).
The average score gain differences between the students from the school that received coaching on the verbal section only was statistically significant. The school that received coaching on both the verbal and mathematical sections were statistically significant in both sections.

In a second study conducted by French, an "identical item" coaching program was employed to examine the effects of coaching students on the same questions that would appear on the SAT. Half of the items on both practice tests contained questions that were on previous editions of the SAT. The identical item coaching program produced an average increase of 47 points higher on the verbal section and 15 points on the mathematical section than the averages obtained by the coached groups that did not receive the identical items format of coaching. There was no reported assessment of statistical significance.

Dear conducted a study on the effects of an SAT coaching program for students from nineteen secondary schools. Six students from ten of the schools received coaching on both the verbal and mathematical sections. The other students from those ten schools served as the first control group. Students from the remaining nine schools served as the second control group. Both control groups received no coaching. Exercise materials prepared by ETS were used by the verbal and mathematics tutors. The coached group received an average verbal score of 2.91 less than the first control group and 2.13 points less than the second control group. The mathematical section showed an average of about 26 points higher than the control group that took the March SAT and 21 points higher than
the control group that took the January administration of the SAT. The average gain in mathematical scores of the coached group was statistically significant.

Frankel examined the effects of a commercial coaching program on 90 high school seniors. A total of 45 pairs were selected, each pair consisted of one student who received coaching and one who had not received coaching. The coached students spent a total of 30 hours preparing for the SAT. The coached students received an average verbal increase of 8.38 points greater than the uncoached students. The coached students received an average mathematical increase of 9.07 points higher than the uncoached students. There was no statistically significant effects in the average increases of the coached group in both the verbal and mathematical sections.

Whitla conducted a study to examine the effects of a coaching program on the scores of 52 students enrolled in the Reading Institute of Boston. A control group was selected and was comprised of 52 students from three area high schools. The coached group received tutoring on the verbal and mathematical sections. The coached group received an average SAT-verbal score of about 11 points higher than the uncoached group. The average SAT-mathematical score of the coached group was 7 points lower than the uncoached group. There was no statistically significant effects in the differences in score increases between the coached and uncoached groups in either the verbal or mathematical sections.

Lass conducted an informal study of the effects of a coaching program in one high school. Three groups were selected for the study. The first group
consisted of students who received no coaching or special preparation. The second group consisted of students who received coaching prior to taking the SAT. The third group consisted of students who had participated in a high school orientation program that was designed to acquaint students with the way the SAT is administered and the kinds of questions asked, rather than involving any extensive training. The average gain in verbal scores from the first administration to the second test administration of the coached group was 3 points higher than the uncoached group and 9 points lower than the orientation group. The average gain in quantitative scores from the first administration group to the second test administration of the coached group was 11 points higher than the uncoached group and 12 points higher than the orientation group. There was no reported assessment of statistical significance. As shown in Table 4, the duration of coaching varied as well as the effects of coaching.

The conclusion reached by the College Board’s examination of the previous seven studies is that the small gains reported as a result of coaching programs were fewer than 10 points—too small to affect college admissions decisions. Moreover, it was stated that the small gains occurred as a result of students and scores changing with the passage of time rather than as a result of any coaching program. Therefore, the predictive validity of the SAT remains undisturbed.

There have been others (DerSimonian & Laird, 1983; Jackson, 1980; Snedecor, 1989) who have also argued against the effectiveness of coaching programs and their impact on the predictive validity of the test scores obtained.
<table>
<thead>
<tr>
<th>Studies Cited</th>
<th>Duration of Coaching</th>
<th>Test Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dyer (1953)</td>
<td>6 hrs. (V) 4 hrs. (Q)</td>
<td>+4.6 pts. above uncoached (V) +12.9 pts. above uncoached (Q)</td>
</tr>
<tr>
<td>French (1955)</td>
<td>Approximately 8 hrs.</td>
<td>+18 pts. above uncoached (V &amp; Q)</td>
</tr>
<tr>
<td>French (1955)</td>
<td>Approximately 8 hrs.</td>
<td>+47 pts. above coached w/o item-format training (V) +15 pts. above coached w/o item-format training (Q)</td>
</tr>
<tr>
<td>Dear (1958)</td>
<td>36 hrs. (12 wks.)</td>
<td>-2.91 pts. below 1st control gp. (V) -2.13 pts. below 2nd control gp. (V) +26 pts. above 1st control gp. (Q) +21 pts. above 2nd control gp. (Q)</td>
</tr>
<tr>
<td>Frankel (1960)</td>
<td>30 hrs.</td>
<td>+8.38 pts. above uncoached (V) +9.07 pts. above uncoached (Q)</td>
</tr>
<tr>
<td>Whitla (1962)</td>
<td>10 hrs.</td>
<td>+11 pts. above uncoached (V) -7 pts. below uncoached (Q)</td>
</tr>
</tbody>
</table>
DerSimonian and Laird (1983) conducted a meta-analysis to evaluate the effect of coaching on SAT scores. Their analysis was based on published uncontrolled and controlled studies that had examined the effects of coaching programs. The authors acknowledged that there was evidence to support a positive effect of coaching; however, they noted that large score increases could not be attributed to coaching because of insufficient evidence. They reached the same conclusion reached by ETS and the College Board, that the score increases that could be attributed to coaching were too small to be of any practical significance.

Slack and Porter (1980) attempted to discredit the utilization of the SAT as a measure unaffected by training or learning experiences. But they were, in turn, criticized by Jackson (1980) for predimating their argument on an unfair representation of facts. Jackson supported the claims made by the College Board
and ETS. Slack and Porter were criticized for not making a clear distinction between the duration of programs such as short-term programs offered by high schools and commercial organizations and more longer-term educational programs. The College Board's statement addressed the ineffectiveness of short-term drill and practice, not longer-term educational programs. Jackson noted that coaching programs have had small effects on SAT scores. Additionally, the author stated that to suggest to students, parents, and schools that coaching programs could normally result in score increases, would be considered "deceptive and educationally unsound."

Snedecor (1989) also argued against coaching programs affecting the predictive validity of SAT scores. Snedecor cited an article published by Whitla (1988) in which SAT scores were obtained from different samples of students who had attended a coaching school compared to students who had not enrolled in a coaching school. The combined score increase on the verbal and quantitative sections of the coached group was 94 points compared to a combined score increase of 64 points for the uncoached group. The point differences amounted to an 11 point increase on the verbal score and a 16 point increase on the quantitative score. These results were used to support the argument that SAT coaching programs were not as effective as had been claimed.

Snedecor (1989), in turn, administered a questionnaire to 535 seniors from 10 high schools to determine if coaching programs were effective in increasing SAT scores. The results reported by the coaching firms showed that some
coaching firms did better than others. However, it was reported that the actual score gain (an average of only 15 points) was insignificant for practical purposes, and that SAT coaching schools seemed to make little difference in score gain. Snedecor noted flaws in the results of the questionnaire in that it did not allow for the measurement of any score gain nor the separate effects of coaching on the verbal and quantitative scores.

On the other hand, some have asserted that coaching programs can result in an increase in a student's scores with the potential of affecting the predictive validity of those scores (Anderson, 1981; Cole, 1982; Messick, 1981; Slack & Porter, 1980). Anderson (1981) provided a summary of coaching issues and noted that there is a positive relationship between amount of contact time and score increases associated with coaching. The author provided non-data-based examples of the number of contact hours needed in order to increase scores. Anderson pointed out that coaching seems to work best when it is used as a "refresher" tool. Cole (1982) strongly believed that coaching could produce detectable, but small, improvements in students' scores and that the small improvements should not pose a serious threat to the predictive validity of those scores. Cole advised students who were seeking to obtain their maximal possible score to participate in a coaching program if the small gains of 10, 20, or 30 points were important in their admission process. This is considered important to some, as it may raise the probability of achieving admission in a highly competitive situation. Cole had two other concerns in addition to whether coaching can increase test scores: (1)
the effectiveness of the different types of coaching programs (e.g., commercial vs. free coaching programs); and (2) the identification of the different components of coaching (which was noted as being poorly identified).

Messick (1981) addressed the issue of coaching effectiveness by pointing to the long-term stability of the improvements. Messick states that "if effective coaching does improve abilities, this might occur via the development of new skills or, more likely, by the strengthening, honing, and refining of existing skills through exercise and challenge" (p. 42). It was also noted that coaching might be effective in teaching answer-selection tricks, thus reducing the "operative difficult level of some test items" (p. 42).

Slack and Porter (1980) also contend that training for the SAT can effectively help students to raise their scores. They further maintained that the SAT scores are not a good predictor of college performance, despite the claims made by ETS and the College Board. They noted that well-designed coaching materials combined with large amounts of study time could result in score increases. They also maintained that the more time spent preparing, the higher their score will be. The authors used studies that had been cited and some not cited in a report by ETS to support their claim. Contrary to ETS and the College Board, they found that seven of the studies cited as being insignificant were indeed statistically significant. The conclusion that Slack and Porter reached was that the SAT is a third-rate predictor of college performance, that it has less relevance than a student's high school grades in the prediction of college performance.
No one seems to have addressed the effects of coaching on predictive validity in light of the somewhat greater effectiveness of some of the coaching programs reviewed in the immediately preceding section of this document. This increased effectiveness might give the testing establishment more reason for concern. At this point, an alternative view to that of Cole and the testing establishment is presented. This view best characterizes that of the training establishment: (a) students have a set of verbal, quantitative, and analytical skills (not abilities) that are relatively stable (fixed is too strong) and that are learned (not developed) over a lifetime; (b) these skills may be important determinants of academic success; (c) standardized tests such as the GRE and SAT sample these skills (not measure these abilities); (d) instructional programs (not the pejorative coaching programs) can improve the scores on these tests (the only question is how long and intensive do these programs need to be. And those instructional programs can do so by helping the students learn the skills they had previously failed to learn (rather than if they cannot affect those fixed and stable abilities); (e) therefore, those programs can (not cannot) affect and improve academic success; and (f) therefore, effective coaching programs need not (rather than would) harm the predictive validity of those tests. We might summarize these differences by saying that the testing establishment categorizes people in terms of their unchanging worth and the training establishment proposes to help individuals improve their worth.

In addition to the ongoing debate about whether coaching affects the
predictive validity of test scores, there has been another important issue raised in relation to the inequitable opportunities for students to participate in coaching programs. The issue pertains to the social implications in relation to coaching opportunities.

**Issues of Social Inequality**

The social implications stem from the concern that there are individuals whose economic status may place them at an advantage on admissions or professional certification tests simply because they can afford the expensive coaching schools. This issue has been viewed as a serious threat to testing in terms of whether such expensive coaching is superior to other types of test preparations.

Anderson (1981) addressed this issue by proposing that public schools become more involved in providing appropriate opportunities to prepare for the tests. The author strongly believed that this could help avoid the inequity associated with students who could afford expensive commercial coaching courses and those who could not. Cole (1982) provided a historical perspective by stating that this issue has gained much attention primarily due to the civil rights movement of the 1960s. She supported the notion that such inequity due to one's socioeconomic background runs contrary to the goal of traditional testing. The premise of this goal has been to extend the opportunity to the most capable regardless of economic background. Cole asserts that there is little reason, based on existing studies, to prefer commercial coaching to free, school-based coaching.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Messick (1981) recommended that all test candidates be given the opportunity to engage in preparation focusing on test-format familiarization and test-taking strategies. Messick noted that equity becomes an issue if there are differences in test taking skills as a result of coaching programs and if those differences affect test scores.

Five of the GRE studies discussed earlier (Powers, 1985; Powers, 1987; Powers & Swinton, 1984; Swinton & Powers, 1983) also raised the issue of the inequity of coaching programs. Powers (1985) contended that coaching was not uniformly available to all test takers, thus supporting the notion that some examinees have an advantage as a result of attending expensive coaching programs. Powers (1987) reiterated the statement made by Cole (1982) that those who stand to benefit the most are those who can least afford to purchase effective preparation.

Powers and Swinton (1984) addressed this inequity issue by noting that there may be differential effectiveness in raising test scores due to the various methods of preparing for the test. According to the authors, methods could vary from very expensive coaching programs to less expensive books and materials, perhaps resulting in the most expensive programs being the most effective. They suggested providing self-study test familiarization materials and encouraging their use in an effort to reduce the inequity problem. It is their opinion, based on their study, that inequality is more related to the candidates' awareness of the need for and their willingness to devote time to test preparation, rather than to the cost.
of materials. Swinton and Powers (1983) recognized the potential for inequality due to examinees’ time to devote to preparation in addition to availability of financial resources. Swinton and Powers, in fact, state that examinees’ awareness, willingness, and time spent preparing were the major contributors to the differences over and above the financial means.

The issue of inequities in coaching was also addressed in the SAT literature. Roughly one third of the studies cited addressed the issue of unfair availability of coaching programs, many expressing different views on the issue. The most prevalent position taken was that the socioeconomically disadvantaged students are not afforded the same opportunities to participate in coaching programs requiring course material fees.

Clients of services such as private counseling, coaching, and tutoring come predominantly from middle class white suburbs, thus widening the gap between the suburban middle class and the urban poor (Coffin, 1987). Smyth (1989) noted that the typical clients of major coaching firms are college-bound students whose family incomes are high enough to pay prep course fees, which could range from $300 to $600. The lack of equal availability of instruction would result in an unfair difference in how adequately students are prepared to take the test (Evans & Pike, 1973). The Federal Trade Commission (1981) reported that the students most in need of training are those who have the least access to it. Students at schools in affluent neighborhoods also have more exposure to the multiple-choice standardized test format prior to taking the SAT.
Some researchers addressed the issue in relation to additional factors such as the geographic location and profile of students. Harris and Rohfeld (1983) pointed to three reasons why inner-city youths rarely participate in coaching programs: (1) insufficient money to pay the fees, (2) inconvenient locations, or (3) lack of awareness. Johnson and Wallace (1989) argued for the continuation and broadening of preparation programs throughout the major urban areas in an effort to reach large at-risk populations of minority youths.

Factors other than the family's financial resources and the geographic location of students were also cited as being contributors to the unfairness of coaching opportunities. Wing, Childs, & Maxwell (1989) contended that gains in SAT scores of students from above-average socioeconomic families could be attributed to a number of factors, such as, gender, racial/geographic background, the year the program was offered, the prep agency, and the verbal and mathematical sections of the SAT (though they did not elaborate sufficiently on these issues).

Inequality of opportunities was also addressed in relation to college admissions policies. Zuman (1988) pointed out one of the problems in the interpretation of the findings of a Federal Trade Commission study. According to this report, the findings revealed significant gains in verbal and quantitative SAT scores of students who had attended a coaching school, but lacked a comparison control group, thereby confounding the interpretation of their results. He contended that the students who attend the commercial test coaching schools tend to be more affluent than the general population of SAT test-takers, thus affecting
the economically, disadvantaged students’ chances of college admissions. Moreover, he emphasized that there are differences in the accessibility to effective equality of educational opportunities.

Considerable concern has been expressed that all students should have equal access to training programs. But even if all socioeconomic barriers were removed, social justice would not be achieved because almost all of the training programs are ineffective in achieving practical significance. Therefore, our major concern should be first to develop effective instructional programs that can compensate for inadequate training received during a student’s lifetime. Only then should we be concerned to ensure equal participation in those programs regardless of the socioeconomic level of the students.

It is evident that there is a wide range of issues related to the social implications of coaching programs. Many factors were cited as being contributors to the unfair accessibility of coaching programs in preparation for standardized tests. Though much attention has been given to coaching programs, views expressed by many on other relevant issues warrant further examination. Those issues are coupled with the need for the continuous development of coaching programs.

An Intensive GRE Instructional Program

Goodyear-Orwat and Malott (1994) provided structured study sessions at a university to help undergraduate students prepare for the GRE. Their objective was to determine whether structured self-study sessions would generate significant
amounts of study time, and whether the time spent preparing would affect students' performance on the verbal and quantitative sections of the GRE. A detailed account of their two instructional programs is provided as a base for systematic replication of their work.

**Summer, 1993 Course**

The GRE preparation course of 1993 was conducted in two classrooms at Western Michigan University during the university's summer session. Ten students participated. Students used the following materials: *Barron's How to Prepare for the GRE* (Brownstein, Weiner, Green, & Hilbert, 1992), *Cliffs Study-Ware for the GRE* (Bobrow, Orton, Covino, 1992), and *Think Fast* (Parsons, 1993). The *Barron's* book contained strategies for taking the GRE, practice drills for each section, commonly used words on the GRE, and five practice tests. The *Cliffs* computerized program contained test-taking strategies, and three practice tests. *Think Fast*, a computerized flashcard program, contained the 300 most commonly used words on the GRE, prefixes and suffixes, and mathematical terms and formulas.

The students participated in this workshop as a regular academic course for 3 credit hours. The course met on Monday through Friday from 8:00 a.m. to 12:00 p.m., totaling 140 hours over 7 weeks. Students who participated in 92% or more of the workshop hours, received a course grade of A.

On the first day of class, students took the first practice test in the
Barron's workbook, where they completed the verbal and quantitative components of the examination. Each section of the practice test had a 30-minute time limit. The first practice scores served as the pretest measures of the students' performance in both areas.

The course was structured to allow students to study in a classroom and a computer room. Since each student had different skill deficits that needed improving, each student was allowed to work at his or her own pace, utilizing the materials of his or her choice. Students decided when to work in the workbook and when to work in one of the computer tutorial programs. In addition to choosing what materials to use, students were allowed to determine the time spent on each.

A posttest—another practice test in the Barron's workbook, was administered to the summer students on the final day of class to determine their progress and whether they met the 500-point score criterion (the most common admission cut off) on both the verbal and quantitative sections. Students also responded to a course evaluation at the completion of the posttest.

All 10 students completed the course, 9 with an attendance of at least 92%, and one with an attendance between 87% and 92%. The mean change from pretest to posttest was an improvement of 206 points (Table 5). The mean change from the pretest to the actual GRE was an improvement of 100 points (Figure 1). Twenty percent of the summer students scored at or above 1,000 on the pretest total. Ninety percent of these students scored at or above 1,000 on
Table 5
Mean Pretest and Posttest Scores by Session for 1993

<table>
<thead>
<tr>
<th>Session</th>
<th>Pretest Mean</th>
<th>Posttest Mean</th>
<th>Difference</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer</td>
<td>906</td>
<td>1,112</td>
<td>206</td>
<td>10</td>
</tr>
<tr>
<td>Fall</td>
<td>865</td>
<td>1,025</td>
<td>160</td>
<td>8</td>
</tr>
</tbody>
</table>

Figure 1. Mean Change From Pretest to Actual GRE by Session for 1993.

the posttest total. Seventy one percent of the summer students received total scores at or above 1,000 on the actual GRE. The mean GRE verbal score was
464 and the mean GRE quantitative score was 587.

Fall, 1993 Course

The second course was offered during the first 6 weeks of the university's fall semester when students were carrying a full load of other classes. Students used the same materials that were used by the summer students. The eight students participating in the fall workshop did not receive course credit, nor did they pay for the workshop; instead, the costs were paid for by the College of Arts and Sciences. They attended class Monday through Thursday from 7:00 p.m. to 10:00 p.m. totaling 66 hours over 6 weeks. Students who, at any time, had failed to attend 92% or more of the workshop were not allowed to continue in the workshop. Otherwise the course procedures were the same as the summer, 1993 course.

All eight students met the 92% criterion. The mean change from pretest to posttest was an improvement of 160 points (Table 5). The mean change from the pretest to the actual GRE was an improvement of 126 points (Figure 1). Twenty percent of the fall students scored at or above 1,000. Seventy one percent of these students scored at or above 1,000 on the posttest total. Sixty two point five percent of the fall students received total scores at or above 1,000 on the actual GRE. The mean GRE verbal score was 478 and the mean GRE quantitative score was 490.
CHAPTER II

SPRING, 1994 COURSE

Introduction

In their GRE preparation courses, Goodyear-Orwat and Malott (1994) provided more hours of training (66 to 140 hours) than any other GRE or SAT course found in the literature. And probably as a result of these larger amounts of training, their courses produced greater increases in combined GRE scores (means of 160 to 206) than any other GRE or SAT course found in the literature.

The following three experimental GRE preparation courses were attempts at systematic replications of these earlier successes. However, the maximum number of hours of study per day and, thus, per course was decreased at student request. In addition, increased guidance was provided as to the topic and materials to be studied at any given time. And, in contrast to the 7-week summer course of Goodyear-Orwat and Malott (1994), this 7 1/2 week spring GRE preparation course was conducted 3 hours per day rather than 4, used a less intensive main text, and used a different attendance and participation policy.

Methods

Setting, Participants, and Materials

The GRE preparation course was sponsored by the Psychology
Department and the Office of Conferences and Institutes at Western Michigan University. It was conducted in two classrooms at Western Michigan University during the university's spring session. One room was equipped with IBM-compatible computers and the other contained only tables, chairs, and desks. Students were recruited from the faculty supervisor of this course as well as from other advanced undergraduate psychology courses. A mass mailing to all juniors and seniors (the total was about 1,400) in the College of Arts and Sciences included information on the GRE preparation course and the registration form. Five out of the eleven students who enrolled completed the course. The cumulative grade point averages of the five spring students ranged from 2.88 to 3.42.

The following materials were used in this course: The Princeton Review: Cracking the GRE, (Robinson & Katzman, 1992), Cliffs StudyWare for the GRE (Bobrow, Orton, & Covino, 1992), Intelligent Computer Tutorial (ICT) for Basic Algebra (1993), and Think Fast (Parsons, 1993). The Princeton book contained easy-to-read strategies and techniques for achieving high scores on the GRE, and limited practice drills for the verbal and math sections, and a vocabulary list containing the most frequently tested words on the GRE. The Cliffs computerized program contained test-taking strategies and drills, and three practice tests. ICT, a computerized flashcard program, contained Algebra tutorials and practice drills. Think Fast, a computerized flashcard program contained the 300 most commonly used words on the GRE, prefixes and suffixes, and mathematical terms and formulas.
Procedure

Because of university constraints, the students paid to participate; and also because of university constraints, they did not receive academic credit. Based on feedback from the summer, 1993 students, the current students attended class Monday through Friday, from only 9:00 a.m. until 12:00 p.m., rather than 8:00 a.m. to 12:00 p.m., thus totaling only 105 hours over 7 1/2 weeks rather than 140 hours over 7 weeks. Students who missed three or fewer classes and actively participated in at least 92% of the total class hours received a $25 rebate, rather than a grade of A or the mere opportunity to stay in the course, as in the earlier courses.

On the first day of class, students took an official 1992 version of the GRE (Educational Testing Service, 1992), rather than an examination from the Barron’s book. A random assignment of students to two different test forms was made; some students were administered the GR92-1 form while the others were given the GR92-2 form. Students completed two sections of both the verbal and quantitative components. Each section of the pretest had a 30-minute time limit. This timed examination served as a pretest measure of student performance in both areas. Following the pretest, the syllabus, course materials, and an informed consent form were given to all students (Appendixes A and B).

Monday, Wednesday, and Friday were normally designated computer days, and Tuesday and Thursday were normally designated classroom days, due to limits
in classroom availability. On days 2, 3, and 4, students read course materials. These students began work on the Cliffs computer program on the 5th day. They took the computer-based Cliffs StudyWare examination #1 to determine their level of performance. At the completion of the examination, the computer generated an Individual Study Plan (ISP). The ISP displayed the students’ scores in two domains—verbal and quantitative, which were subdivided into a total of 16 different topic areas. The scores were categorized into five different achievement levels: (1) very weak, (2) weak, (3) fairly strong, (4) strong, and (5) very strong. Each topic area listed page numbers for review of those specific topics in the Cliffs supplemental preparation book. Students used their ISP recommendations pertaining to the verbal and quantitative sections as a guide for further review.

Spring students were given a sheet to self-record their activities each day (Appendix C). These sheets were turned in at the end of each week for review by the instructor and given back to the students at the beginning of the following week.

A posttest was administered to the spring students on the final day of class, to determine their improvement in both the verbal and quantitative domains. Students responded to a course evaluation at the completion of the posttest. Copies of the pretest and posttest scores were mailed to each student.

Results

Five of the eleven enrolled students actually completed the course. So the
data analysis will be restricted to those five. Of the five students, the mean number of days the students attended class was 29.40 out of 35. Two of those five students met the criteria of missing no more than three days and actively participating in at least 92% of the class hours and thus received the $25 award.

The combined quantitative and verbal pretest scores ranged from 730 to 960 points. The posttest scores ranged from 760 to 1100 points. The mean of the combined scores from pretest to posttest changed by +86 points (Table 6). The mean change from the pretest to the actual GRE was +78 points (Figure 2). The change in scores ranged from -10 to +240 points. The mean difference between the posttest and the actual GRE was -8 points (Figure 3) indicating the posttest was a fairly reliable predictor of the actual GRE score, at least on the average. The difference in scores from posttest to the actual GRE ranged from -60 to +30 points.

Table 6

Mean Pretest and Posttest Scores by Session for 1994

<table>
<thead>
<tr>
<th>Session</th>
<th>Pretest Mean</th>
<th>Posttest Mean</th>
<th>Difference</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring</td>
<td>852</td>
<td>938</td>
<td>86</td>
<td>5</td>
</tr>
<tr>
<td>Summer</td>
<td>793</td>
<td>887</td>
<td>94</td>
<td>6</td>
</tr>
<tr>
<td>Fall</td>
<td>870</td>
<td>973</td>
<td>103</td>
<td>9</td>
</tr>
</tbody>
</table>
Figure 2. Mean Change From Pretest to Actual GRE by Session for 1994.

Figure 3. Mean Posttest Minus Actual GRE Score.
CHAPTER III

SUMMER, 1994 COURSE

Introduction

This course entailed the use of additional materials in an attempt to increase its impact on the GRE score. Also, a change in the structure of their classroom and computer time was made to provide students with a variety of daily activities.

Methods

Except where noted the method was the same as in the previous experiment.

Setting, Participants, and Materials

The GRE preparation course was conducted during the university's summer session. These students were recruited in the same manner as the spring, 1994 course, without the inclusion of a mass mailing. Six out of the ten students who enrolled completed the course. The cumulative grade point averages of the six summer students ranged from 3.01 to 3.73.

These students used the same course materials as used by the spring, 1994
students, with the addition of *Barron's How to Prepare for the GRE*, (Brownstein, Weiner, Green, & Hilbert, 1992). The *Barron's* book contained six complete model examinations and an in-depth review covering all test areas.

**Procedure**

The six students were required to meet the same attendance and participation criteria as described in the previous course. These students worked in the classroom for the first hour, and on the computer for the last two hours of each day. This change provided more daily variety as suggested by the spring, 1994 students. Students took the pretest on the first day and read course materials on days 2 to 6. These students began work on the *Cliffs* computer program on the 7th day. They took the computer-based *Cliffs StudyWare* examination #1 and generated an ISP categorizing their achievement levels as in the spring course. Summer students were given a more in-depth sheet to record their activities each day (Appendix C). In addition to recording their activities, they were also required to record the time spent on each activity.

**Results**

This analysis will be restricted to the six of the 10 students who completed the course. The mean number of days these six students attended class was 22.5 out of 35. None met the attendance criteria to receive the $25 award.

The combined quantitative and verbal pretest scores ranged from 680 to
910 points. The posttest scores ranged from 760 to 1020 points. The mean difference of the combined scores from pretest to posttest was +94 points (Table 6). The mean change from the pretest to the actual GRE was -20 points (Figure 2). The change in scores ranged from -120 to +90 points. The mean difference between the posttest and the actual GRE was -114 points (Figure 3). This time the validity of the posttest as a predictor of the actual GRE was not too good. The change in scores from posttest to the actual GRE ranged from -70 to -180 points.
CHAPTER IV

FALL, 1994 COURSE

Introduction

The fall GRE preparation course met for a shorter period of time (5 1/2 weeks) than the spring and summer courses (7 1/2 weeks) due to the October administration of the GRE. The course ended two days prior to the examination, thus changes were made to allow students to make the most optimal use of their time. Students were required to study the specific areas outlined in their study sheets, and they had the opportunity to review practice items with the instructor. Students took a second Cliffs examination that was used to compare any changes in their performance in both the quantitative and verbal domains. The monetary award for meeting the attendance and participation criteria was increased in an attempt to improve class attendance and participation. In addition, there were only 22 days to prepare, rather than the 35 days for the spring and summer courses; and changes were made to collect more data on student performance and to more effectively structure the study time during this shortened GRE preparation course.
Methods

Setting, Participants, and Materials

The GRE preparation course was sponsored by the two departments previously noted in the first two courses. Similarly, the course was conducted in the classrooms noted in the previous two courses. These students were recruited from advanced psychology courses as with the previous two courses. Also, an advertisement was placed in the school's newspaper prior to the fall, 1994 course. All nine of the students completed the course. The cumulative grade point averages of the nine fall students ranged from 2.16 to 4.00.

The fall students used the same materials used by the summer, 1994 students. An updated version of two of the books was made available prior to this course. These students used the Princeton book (Robinson & Katzman, 1993) and the Barron's book (Brownstein, Weiner, Green, & Hilbert, 1994).

Procedure

The students paid to participate. They attended class Monday through Thursday, from 7:00 p.m. until 10:00 p.m. totaling 66 hours over 5 1/2 weeks. Students received a $50 rebate for meeting the 92% attendance and participation criteria. Students took the pretest on the first day and began work on the Cliffs computer program on the 2nd day. They took the computer-based Cliffs Study-Ware examination #1 and generated an ISP categorizing their achievement levels...
in the five areas previously noted. In addition, these students took the Cliffs examination #2 on the 20th day of class. The objective was to provide students with more information on their performance and to enable them to determine if they had made progress in any of their areas of weakness.

These students also recorded the time spent on each activity (Appendix C). Furthermore, they were required to follow their ISP. The objective was to structure the students' time more optimally. They were given a Daily Study Chart (Appendix D) that contained a list of all 16 Cliffs topic areas pertaining to the verbal and quantitative domains, in addition to their corresponding reading and practice materials. They were instructed to record the rating of their skills in each topic area from very weak to very strong as indicated in their ISP generated by the Cliffs program. For each area, there was a corresponding Daily Study Sheet with corresponding readings and exercises from the other course materials (Appendix E). Each student selected the appropriate study sheet, starting with his or her weakest rating, and completed all readings and exercises for that particular area. Students signed and recorded the date of task completion on each sheet before turning it in to the instructor. Students could not proceed to the next area until they completed all tasks outlined in the previous study sheet. This continued for the duration of the course.

Fall students were also encouraged to review their practice test questions with the instructor (Appendix F). The goal was to enable them to talk through the steps involved in the exercises or practice problems, and to give students
additional contact with some of the materials. The involvement of the instructor entailed assisting the students in locating relevant sections in their reading materials.

Results

The fall session entailed 22 three-hour class days. All nine enrolled students completed the course. The mean number of days these nine students attended was 19.5 out of 22. Six of these nine students met the attendance and participation criteria and thus received the $50 award.

The combined quantitative and verbal pretest scores ranged from 600 to 1100 points. The posttest scores ranged from 820 to 1190 points. The mean difference of the combined scores from pretest to posttest was +103 points (Table 6).

The mean change from the pretest to the actual GRE was +68 points (Figure 2). The change in scores ranged from -30 to +170 points. The mean difference between the posttest and the actual GRE was -35 points (Figure 3). Again the validity of the posttest as a predictor of the actual GRE was not too good. The change in scores from posttest to the actual GRE ranged from -160 to +100 points.

As few as 7 (78%) and as many as 9 (100%) of the fall students turned in their self-recording sheets for a given week. Eight of the nine students showed 100% adherence to their ISP recommendations. That is, they worked from and
recorded their items in the exact order outlined in their ISP. One student worked and recorded the items outlined in the ISP but chose to complete his work in a different order; this student worked on five of the top nine "very weak" items in his ISP.

The fall, 1994 students, unlike the spring and summer students, also took the second *Cliffs* examination near the end of the course. Student performance was analyzed according to the 16 topic areas—9 focusing on the verbal skills and 7 on the quantitative skills. To assess the improvement in the topic areas, the number of students (9) was multiplied by the number of topic areas (16), so that there were 144 "opportunities for improvement." Each "opportunity for improvement" was then categorized according to where it fell among the five achievement levels ("very weak" to "very strong"). Figure 4 shows the percentage of the 144 "opportunities for improvement" distributed among the 5 achievement levels for examination #1 and examination #2. The majority of the students' scores remained in the very weak level even after 5 1/2 weeks of preparation. The number of actual improvements of the 81 "opportunities for improvement" in the verbal topic area was 32 (40%). In other words, the *Cliffs* examination showed little evidence of reliable improvement in the verbal topic areas. Similar calculations were made for the seven quantitative topic areas with 67% of the "opportunities for improvement" showing improvements, a slightly more encouraging number. Table 7 shows the percentage of fall students whose scores increased from the first to the second examination by individual topic areas.
A total of 43 topic areas were completed. All completed topic areas were from the "very weak" achievement level. Of those completed, 22 topic areas were from the verbal domain and 21 were from the quantitative domain.

The fall, 1994 students were encouraged to review their practice test items with the instructor who assisted the students in locating items in the reading materials that were related to those test items the students had trouble with during their practice. Eight of the 9 students chose to meet with the instructor on at least one occasion.
Table 7

Cliffs Score Increases by Area From Examination #1 to Examination #2

<table>
<thead>
<tr>
<th>Cliffs Items</th>
<th>% of Students Showing Increases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rdng Comp-Inf</td>
<td>0%</td>
</tr>
<tr>
<td>Rdng Comp-Org</td>
<td>89%</td>
</tr>
<tr>
<td>Rdng Comp-Main</td>
<td>67%</td>
</tr>
<tr>
<td>Rdng Comp-Tone</td>
<td>56%</td>
</tr>
<tr>
<td>Rdng Comp-App</td>
<td>0%</td>
</tr>
<tr>
<td>Rdng Comp-Det</td>
<td>22%</td>
</tr>
<tr>
<td>Antonyms</td>
<td>56%</td>
</tr>
<tr>
<td>Analogies</td>
<td>44%</td>
</tr>
<tr>
<td>Sent Completion</td>
<td>22%</td>
</tr>
<tr>
<td>Quant Comp-Alg</td>
<td>100%</td>
</tr>
<tr>
<td>Quant Comp-Geom</td>
<td>44%</td>
</tr>
<tr>
<td>Quant Comp-Arth</td>
<td>78%</td>
</tr>
<tr>
<td>Math Abil-Arth</td>
<td>89%</td>
</tr>
<tr>
<td>Math Abil-Geom</td>
<td>56%</td>
</tr>
<tr>
<td>Math Abil-Alg</td>
<td>56%</td>
</tr>
<tr>
<td>Graphs/Charts</td>
<td>44%</td>
</tr>
</tbody>
</table>
CHAPTER V

OVERALL RESULTS OF THE THREE 1994 COURSES

For the remainder of this analysis of the GRE scores, the results of the three 1994 courses were pooled. Fifteen of the 20 students who completed the course took the GRE. Seven percent took the examination in August, 73% took the examination in October, and 20% took it in December. One student chose to take the computerized version of the GRE. A correlated-sample t-tests was used to evaluate these pooled results (Table 8). The combined verbal and quantitative scores improved 96 points from pretest to posttest ($M=96$, $SD=93.0$), $t(20)=4.61$, $p < .05$. The verbal scores improved 39.5 points from pretest to posttest ($M=39.5$, $SD=66.1$), $t(20)=2.67$, $p < .05$. The quantitative scores improved 56.5 points from pretest to posttest ($M=56.5$, $SD=50.4$), $t(20)=5.01$, $p < .05$. Though the difference in combined scores improved 56 points from the pretest to the actual GRE, there was no statistically significant effect. The posttest over predicted the actual GRE by a mean of 43 points.

Figure 5 shows the combined results of the 15 students who completed the three 1994 courses. Six percent of the students received a score of 500 or above on the verbal pretest compared to 26% on the posttest. Twenty six percent of the students received a score of 500 or above on the quantitative pretest compared to 53% on the posttest. Twenty percent of the students received a score of 500
Table 8
Summary of t-Test Analysis of Scores

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>20</td>
<td>842.5</td>
<td>121.6</td>
<td>96*</td>
</tr>
<tr>
<td>Posttest</td>
<td>20</td>
<td>938.5</td>
<td>119.7</td>
<td></td>
</tr>
<tr>
<td>Pretest-Verbal</td>
<td>20</td>
<td>394.5</td>
<td>66.4</td>
<td>39.5*</td>
</tr>
<tr>
<td>Posttest-Verbal</td>
<td>20</td>
<td>434</td>
<td>77.5</td>
<td></td>
</tr>
<tr>
<td>Pretest-Quantitative</td>
<td>20</td>
<td>448</td>
<td>88.4</td>
<td>56.5*</td>
</tr>
<tr>
<td>Posttest-Quantitative</td>
<td>20</td>
<td>504.5</td>
<td>80.9</td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>15</td>
<td>836</td>
<td>124</td>
<td>56</td>
</tr>
<tr>
<td>GRE</td>
<td>15</td>
<td>892</td>
<td>158.5</td>
<td></td>
</tr>
</tbody>
</table>

*p<.05

or above on the actual GRE verbal compared to 46% percent on the GRE quantitative. The mean change in performance from pretest to posttest on the verbal section ranged from -80 to +140 points and the quantitative section ranged from -30 to +140 points. Figure 5 also shows the percentage of 1994 students with combined verbal and quantitative scores at or above 1,000 points. Six percent of the students received a pretest score above 1,000 compared to 26% on the posttest and 26% on the actual GRE.
An attendance and participation policy was established for all three courses to encourage adequate study time. The spring and summer students could receive a $25 rebate for meeting the 92% attendance and participation criteria, but only two of the 21 students starting the spring and summer courses, met the criteria to receive the $25. So $50 was offered to the fall students for meeting the 92% criteria; and this time, six of the nine students starting the course met the criteria to receive the $50 award.
While the fall's increased attendance and participation may have been due
to the larger award, it might also have been due to other factors, most probable
of which might be the smaller number of hours and days required during the fall--
66 hours rather than 105 hours. In other words, it was easier to meet the criteria
for the shorter fall course.

However, the spring and summer attrition data suggest otherwise: The
mean number of days attended by the 10 students who dropped out of the spring
and summer courses was only 12.7, ranging from 6 to 23. This low number of
attended days suggests that generally they did not drop out because of the longer
length of the spring and summer courses. This suggestion is supported by the
observation that 8 of those 10 dropouts completed fewer days of the course than
the fall students' mean of 19.5. So the ($50) award appears to be a more likely
cause of improved attendance.

Recruitment

Several efforts were made to recruit students for the GRE courses. These
efforts entailed presentations given to undergraduate psychology classes, a mass
mailing to all junior and senior students in the College of Arts and Sciences, and
a campus newspaper advertisement. The most effective recruitment approach was
the classroom presentations. Of the thirty students who enrolled in the courses,
29 heard about the course via class presentations. The other one heard about the
course through the university's Graduate College. Perhaps further efforts should
be made to recruit students from other psychology and non-psychology classes, rather than through mailings and advertisements; this might be a more cost-effective approach.

Value of the GRE Preparatory Courses

Figures 6 to 10 show students' ratings of the materials used in all three courses. The majority of the students' ratings were in the range of 3 (the middle rating) to 5 (the highest rating). In addition to all other course materials, the summer and fall students used the Barron's book as a result of feedback received from the spring students. The majority of the students gave the book a rating of 3 or 4.

All students were asked to give their rating on the value of the course (Appendix G). The fall students gave the overall course the highest rating (Figure 11). As a follow-up to this question, students were asked to comment on how the course could be improved. Some common suggestions included: (a) having more practice and timed tests throughout the course (5), (b) the availability of a math instructor/tutor (7), and (c) tighter attendance and higher monetary contingencies (9).
Figure 6. Rating of Barron's Book.

Figure 7. Rating of Princeton Book.
Figure 8. Rating of Cliffs Program.

Figure 9. Rating of Think Fast Program.
Figure 10. Rating of ICT (Algebra) Program.

Figure 11. Value of GRE Course.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
CHAPTER VI

DISCUSSION

Comparison With the Goodyear-Orwat and Malott (1994) Study

The earlier study in this thematic series conducted by Goodyear-Orwat and Malott (1994) involved two courses during 1993. Those 1993 and the present 1994 courses were similar in that both sets provided structured study sessions to help undergraduates prepare for the GRE. The objective was the same—determination of whether structured self-study sessions and the time spent preparing would affect student performance on the verbal and quantitative sections of the GRE. However, there were differences in the student performance, attendance policy, how the students spent their time, the materials used, and the number of hours per day during the spring and summer courses.

The 1993 students were allowed to work at their own paces utilizing the Barron's, Cliffs, and Think Fast materials. The summer and fall, 1994 students used the same materials and in addition used the Princeton and ICT materials. The spring, 1994 students used all materials used by the other 1994 students with the exception of the Barron's book.

The mean improvements from the pretest to posttest for the 1993 students were 206 and 160 points for the summer and fall courses. The mean
improvements from the pretest to posttest for the 1994 students were 86, 94, and 103 points for the spring, summer, and fall courses—considerably less. Pretest and posttest data for the students in the 1994 courses, along with their grade point averages at the time of enrollment, are presented in Appendix H. The differences between the summer, 1993 course and the spring and summer, 1994 courses could be attributed to the decreased number of hours of study offered by the 1994 courses (105 hours rather than 140 hours). However, this does not account for the difference between the fall, 1993 course and the fall, 1994 course, as the hours were the same for both (66 hours).

A second factor could have been the textbooks. The 1993 courses used the Barron's book as their main text. The 1994 courses either did not use the Barron's book or put less emphasis on it. And concentrating on the extensive drills of the Barron's book might have been crucial.

A third, and perhaps more probable factor could have been the type of tests used to assess pretest and posttest scores. The 1993 courses used practice tests contained in the Barron's book; whereas, the 1994 courses used two retired versions of the 1992 administration of the GRE. The Barron's tests might have been more sensitive to the training effects than the retired GRE tests. Therefore, this 1994 study might more realistically reflect the impact of this extensive training than did the 1993 study.

Different attendance and participation policies were used for the 1993 and 1994 courses. The summer, 1993 course offered a grade of A for students
meeting the 92% attendance and participation criteria. The fall, 1993 course allowed students to remain in the course only if they continued to meet the 92% criteria. Both policies seemed effective. The 1994 courses offered a monetary award for meeting the 92% attendance and participation criteria. Seventeen of the 18 students (94%) enrolled in the 1993 courses met the attendance and participation criteria; whereas only 8 of the 20 students (40%) completed the 1994 courses met the attendance and participation criteria and thus received the monetary award. One could speculate, with caution, that those monetary incentives (especially the $25 award) were not as effective as the incentives offered in 1993.

Practice Effects

How likely is it that the large 1994 statistically significant increases in verbal and quantitative scores resulted from the confounding of taking the exam twice (pretest and posttest) rather than from the training received in the GRE courses? This might be answered by looking at the earlier GRE studies that involved only a few training hours (2.9 to 9.4 hours). None produced improvements in verbal and quantitative GRE scores near the magnitude of the current study. And none produced statistically significant results (the largest improvement was 0.2 points). No relevant data from the SAT research were found. The present study used more hours (105 and 66 ) and produced large statistically significant improvements from the pretest to the posttest. So those earlier studies suggest that the mere fact of taking the test twice, without adequate training during the
interim, will not produce the sizable improvements found in this study.

Practical Significance

Cole (1982) asserts that if such small, but detectable, gains as 10, 20, or 30 points were important in the admission process, then students should participate in a coaching program. In that light, the mean improvements for the 1994 courses of 40 points (verbal), 56 points (quantitative), and 96 points (combined) suggest that intensive, structured training can produce practically significant results, thus supporting the notion that training can have an impact on standardized test scores, which may ultimately affect student admission. Finally, it should be emphasized that no earlier GRE studies were found that have shown improvements near the magnitude of those obtained in the 1993 and 1994 studies in the current line of thematic research.

Future Research

The present study showed that a complex intervention program could produce statistically significant effects. But, because the present experiments used a cluster of study materials, the assessment of the specific components of the training materials was precluded. Such an assessment might be the basis of future research.

The attendance of the 1994 students was much lower than the 1993 students. Future research could also focus on various attendance policies, such as
the larger monetary incentive or a grade at the end of the course. It might also be of interest to study the effects of smaller monetary contingencies distributed throughout the course. Other factors to investigate might be the nature of the study activities, such as lectures versus self-study.

Finally, there is the question of predictive validity. Will large GRE improvements resulting from GRE training courses be correlated with proportional improvements in graduate-school success? This is an important area of future research.
Appendix A

Course Syllabi
Time: 9:00 a.m. to noon

Room: 352 & 353 Wood Hall

Instructor: R. W. Malott

Assistants: Jan M. Miller
            Brad Frieswyk

Rebate Policy: If you miss three or fewer classes and participate actively in all the others, you'll earn a $25 rebate after all special course materials have been returned.

Course Structure: This course will provide the structure for the hard work you need to do. The class will meet Monday through Friday from 9:00 a.m.-noon (May 2-June 20). You will not receive credit for this course; however, prompt attendance is essential. Attendance will be the key factor in putting in the requisite amount of work. Studying will take place in a computer lab as well as a classroom in Wood Hall.

Course Rationale: If you receive good scores on the post test, you should take the computerized GRE at the completion of this course. If your scores are good on the computerized GRE, you can get drunk. Otherwise, we recommend you take the summer GRE course. Upon completion of the summer session, check to see if Jan Miller and Dr. Richard Malott will be offering a free review. If so, you're advised to attend the review session and then register for the October GRE.

Class Assignments for Week 1

Monday, May 2          Pretest & begin reading the Princeton book
Tuesday, May 3          Reading
Wednesday, May 4        Reading
Thursday, May 5         Reading
Friday, May 6           Cliffs computer program (Computer Lab, rm. 352)

Standard Format for Future Weeks

Monday             Computer
Tuesday            Reading
Wednesday         Computer
Thursday          Reading
Friday             Computer
Time: 9:00 a.m. to noon

Room: 352 & 353 Wood Hall

Instructor: R. W. Malott

Assistants: Jan M. Miller
Jennifer Tairiol
Brad Frieswyk

Rebate Policy: If you miss three or fewer classes and participate actively in all the others, you'll earn a $25 rebate after all special course materials have been returned.

Course Structure: This course will provide the structure for the hard work you need to do. The class will meet Monday through Friday from 9:00 a.m.-noon (June 29-Aug. 17). You will not receive credit for this course; however, prompt attendance is essential. Attendance will be the key factor in putting in the requisite amount of work. Studying will take place in a computer lab as well as a classroom in Wood Hall.

Course Rationale: If you receive good scores on the post test, you should take the computerized GRE at the completion of this course. If your scores are good on the computerized GRE, you can get drunk. Otherwise, we recommend you check to see if Jan Miller and Dr. Richard Malott will be offering a free review. If so, you're advised to attend the review session and then register for the October GRE.

Class Assignments for Weeks 1 & 2

<table>
<thead>
<tr>
<th>Day</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wednesday, June 29</td>
<td>Pretest &amp; begin reading the Princeton book</td>
</tr>
<tr>
<td>Thursday, June 30</td>
<td>Reading</td>
</tr>
<tr>
<td>Friday, July 1</td>
<td>Reading</td>
</tr>
<tr>
<td>Monday, July 4</td>
<td>No Class</td>
</tr>
<tr>
<td>Tuesday, July 5</td>
<td>Reading</td>
</tr>
<tr>
<td>Wednesday, July 6</td>
<td>Reading</td>
</tr>
<tr>
<td>Thursday, July 7</td>
<td>Reading</td>
</tr>
<tr>
<td>Friday, July 8</td>
<td>Cliffs computer program (Computer Lab, rm. 352)</td>
</tr>
</tbody>
</table>

Standard Format for Future Weeks

<table>
<thead>
<tr>
<th>Day</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>Computer</td>
</tr>
<tr>
<td>Tuesday</td>
<td>Reading</td>
</tr>
<tr>
<td>Wednesday</td>
<td>Computer</td>
</tr>
<tr>
<td>Thursday</td>
<td>Reading</td>
</tr>
<tr>
<td>Friday</td>
<td>Computer</td>
</tr>
</tbody>
</table>
GRE Preparation Course
Fall, 1994

Days: Monday-Thursday

Time: 7:00 p.m.-10:00 p.m.

Room: 352 & 353 Wood Hall

Instructor: R. W. Malott

Assistant: Jan M. Miller

Rebate Policy: If you earn 613 (92%) points or more through active participation, you'll receive a $50 rebate after all special course materials have been returned. For calculation of class points, see box below:

<table>
<thead>
<tr>
<th>Class Point System</th>
</tr>
</thead>
<tbody>
<tr>
<td>■ Earn 10 points per hour of active class participation</td>
</tr>
<tr>
<td>■ Earn 30 points a night</td>
</tr>
<tr>
<td>■ Potential to earn 660 points total</td>
</tr>
<tr>
<td>■ Lose 5 points per 1/2 hour missed (absent)</td>
</tr>
</tbody>
</table>

Course Structure: This course will provide the structure for the hard work you need to do. The class will meet Monday through Thursday (August 30-October 6). You will not receive credit for this course; however, prompt attendance is essential. Attendance will be the key factor in putting in the requisite amount of work. Studying will take place in a computer lab as well as a classroom in Wood Hall.

Course Rationale: This course concludes two days prior to the October administration of the GRE. Therefore, you should have already registered to take the exam on October 8. If you have not made arrangements and you receive good scores on the post test, you should take the computerized GRE at the completion of this course. If your scores are good on the computerized GRE, you can celebrate. If not, you're advised to attend the review session and then register for the December GRE.

Class Assignments for First Two Days

Tuesday, August 30  Pretest
Wednesday, August 31  Cliffs Exam #1 & Unit 1 Readings:
                       Cliffs-Part I, Princeton- Chps. 1,2,3, Barrons-Chp. 1

**No class on Monday, Sept. 5**
Standard Format for Non-Test Days

Monday-Thursday

7:00 p.m.-8:00 p.m.
- Readings from Daily Study Chart & other study activities

8:00 p.m.-10:00 p.m.
- Computer Programs: Cliffs, Thinkfast, & Basic Algebra
  (follow Daily Study Chart)

Monday, October 3
- Cliffs Exam # 2, resume activities noted on Daily Study Chart

Wednesday, October 5
- Re-Read Princeton-Chps. 1,2,3, then resume activities noted on Daily Study Chart

Thursday, October 6
- Posttest & Wrap-up
Appendix B

Informed Consent Form
Western Michigan University
Department of Psychology
Principal Investigator: Jan M. Miller
Advisor: Richard W. Malott, PhD

I give my permission to Jan M. Miller to use the data collected during this GRE Preparation course in her dissertation and in professional presentations and articles. She is collecting these data to evaluate the effectiveness of this course in helping students prepare for the GRE.

I understand that all the information collected from me is confidential. That means that my name will not appear on any papers on which this information is recorded. All forms will be coded, and the principal investigator will keep a separate master list with the names of the participants and the corresponding code numbers.

I understand that I may withdraw my permission at any time during this course without prejudice or penalty. If I have any questions or concerns about this study, I may contact Richard W. Malott, PhD at 387-4481. I may also contact the Chair of Human Subjects Institutional Review Board or the Vice President for Research with any concerns that I have (387-8293 and 387-8298, respectively). My signature below indicates that I understand the purpose and requirements of the study and that I agree to participate.

__________________________________________  ______________________
Signature                                      Date
Appendix C

Self-Recording Sheets
<table>
<thead>
<tr>
<th>Date</th>
<th>C</th>
<th>R</th>
<th>Materials</th>
<th>Sections Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>Time Frame</td>
<td>Book(s) &amp; Pg. #'s</td>
<td>Computer Program(s) &amp; Activity</td>
<td>Other Activities (Be specific)</td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
<td>------------------</td>
<td>-------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td></td>
<td>9:00-10:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10:00-11:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11:00-12:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9:00-10:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10:00-11:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11:00-12:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9:00-10:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10:00-11:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11:00-12:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9:00-10:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10:00-11:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11:00-12:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9:00-10:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10:00-11:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11:00-12:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9:00-10:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10:00-11:00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11:00-12:00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Name __________________________

[Note: The table represents a time log with columns for date, time frame, book(s) & page numbers, computer program(s) & activity, and other activities. Each row corresponds to a different time frame and includes placeholders for the information.]
(1) Record date & activity performed.
(2) Record time spent (beginning & ending) below each activity.
(3) Remember, recording sheet must be turned in every Thursday evening after class.

<table>
<thead>
<tr>
<th>Date</th>
<th>Book(s) &amp; Pg. #'s</th>
<th>Computer Program(s) Name &amp; Activity</th>
<th>Other Activities (Be specific)</th>
<th>Time:</th>
<th>Time:</th>
<th>Time:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix D

Daily Study Chart
<table>
<thead>
<tr>
<th>Rank Order</th>
<th>Very Weak / Weak Areas</th>
<th>Cliffs Book</th>
<th>Cliffs Computer Drill</th>
<th>Princeton Book</th>
<th>Barrons Book</th>
<th>Thinktest</th>
<th>Basic Algebra</th>
<th>Completed?</th>
</tr>
</thead>
<tbody>
<tr>
<td>RdingComp-inf</td>
<td>All pgs. recommended</td>
<td>Drill # 2 (RdingComp-inf)</td>
<td>Chap. 5</td>
<td>Chap. 5</td>
<td>3 Decks</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RdingComp-Inf</td>
<td>All pgs. recommended</td>
<td>Drill # 2 (RdingComp-Inf)</td>
<td>Chap. 5</td>
<td>Chap. 5</td>
<td>3 Decks</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RdingComp-Main</td>
<td>All pgs. recommended</td>
<td>Drill # 2 (RdingComp-Main)</td>
<td>Chap. 5</td>
<td>Chap. 5</td>
<td>3 Decks</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RdingComp-Tone</td>
<td>All pgs. recommended</td>
<td>Drill # 2 (RdingComp-Tone)</td>
<td>Chap. 5</td>
<td>Chap. 5</td>
<td>3 Decks</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RdingComp-App</td>
<td>All pgs. recommended</td>
<td>Drill # 2 (RdingComp-App)</td>
<td>Chap. 5</td>
<td>Chap. 5</td>
<td>3 Decks</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RdingComp-Del</td>
<td>All pgs. recommended</td>
<td>Drill # 2 (RdingComp-Del)</td>
<td>Chap. 5</td>
<td>Chap. 5</td>
<td>3 Decks</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analyses</td>
<td>All pgs. recommended</td>
<td>Drill # 2 (Analyses)</td>
<td>Chap. 4</td>
<td>Chap. 4</td>
<td>3 Decks</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sent Completion</td>
<td>All pgs. recommended</td>
<td>Drill # 2 (SentCompletion)</td>
<td>Chap. 7</td>
<td>Chap. 7</td>
<td>3 Decks</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QuantComp-Alg</td>
<td>All pgs. recommended</td>
<td>Drill # 1 (QuantComp-Alg)</td>
<td>Chaps. 8, 9, &amp; 11</td>
<td>Chaps. 8, 9</td>
<td>Math Decks</td>
<td>Tutorial &amp; Posttest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QuantComp-Geom</td>
<td>All pgs. recommended</td>
<td>Drill # 1 (QuantComp-Geom)</td>
<td>Chaps. 8, 9, &amp; 11</td>
<td>Chaps. 8, 9</td>
<td>Math Decks</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QuantComp-Arth</td>
<td>All pgs. recommended</td>
<td>Drill # 1 (QuantComp-Arth)</td>
<td>Chaps. 8, 9</td>
<td>Chaps. 8, 9</td>
<td>Math Decks</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MathAbil-Arth</td>
<td>All pgs. recommended</td>
<td>Drill # 1 (Discrete Quant-Arth)</td>
<td>Chaps. 8, 9</td>
<td>Chaps. 9, 11</td>
<td>Math Decks</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MathAbil-Geom</td>
<td>All pgs. recommended</td>
<td>Drill # 1 (Discrete Quant-Geom)</td>
<td>Chaps. 8, 9</td>
<td>Chaps. 9, 11</td>
<td>Math Decks</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MathAbil-Alg</td>
<td>All pgs. recommended</td>
<td>Drill # 1 (Discrete Quant-Alg)</td>
<td>Chaps. 8, 9</td>
<td>Chaps. 9, 11</td>
<td>Math Decks</td>
<td>Tutorial &amp; Posttest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graph/Charts</td>
<td>All pgs. recommended</td>
<td>Drill # 1 (Graph/Table)</td>
<td>Chap. 10</td>
<td>Chaps. 10, 11</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note:
1) Rank areas in same order noted on your Individual Study Plan.
2) Pick up Daily Study Sheet for each section.
3) Complete sections in rank order (from very weak to weak).
4) Follow guidelines for each section & return Daily Study Sheet when completed.
Appendix E

Sample Daily Study Sheet
Daily Study Sheet

<table>
<thead>
<tr>
<th>Cliffs Book</th>
<th>Cliffs Computer Drill</th>
<th>Princeton Book</th>
<th>Barons Book</th>
<th>Thinkfast</th>
<th>Basic Algebra</th>
<th>Completion Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>All pgs. recommended</td>
<td>Drill #2 (RdgComp-Org)</td>
<td>Chap. 5</td>
<td>Chap. 5</td>
<td>3 Decks</td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

Note: 1. When all sections have been done, be sure to write date of completion before turning in.
2. Write down all activities on self-recording sheet.
Appendix F

Review Session Sheet
<table>
<thead>
<tr>
<th>Name/Date</th>
<th>Book &amp; Pg. #</th>
<th># Correct (total)</th>
<th># Correct After Review</th>
<th>Comments</th>
</tr>
</thead>
</table>

Review Session Sheet
Appendix G

Course Evaluation Forms
GRE Preparation Course Evaluation
Spring, 1994

1. How would you rate the Princeton book?  
   1  2  3  4  5  
   bad  good  
   comments/suggestions:

2. How would you rate the Cliff's program?  
   1  2  3  4  5  
   bad  good  
   comments/suggestions:

3. How would you rate the Thinkfast program?  
   1  2  3  4  5  
   bad  good  
   comments/suggestions:

4. How would you rate the Algebra program?  
   1  2  3  4  5  
   bad  good  
   comments/suggestions:

5. How valuable has this course been?  
   1  2  3  4  5  
   not at all  very  
   comments/suggestions:

6. How would you improve this course?
1. How would you rate the Princeton book?  
   1  2  3  4  5  
   bad  good  
   comments/suggestions:

2. How would you rate the Barrons book?  
   1  2  3  4  5  
   bad  good  
   comments/suggestions:

3. How would you rate the Cliff's program?  
   1  2  3  4  5  
   bad  good  
   comments/suggestions:

4. How would you rate the Thinkfast program?  
   1  2  3  4  5  
   bad  good  
   comments/suggestions?

5. How would you rate the Algebra program?  
   1  2  3  4  5  
   bad  good  
   comments/suggestions:

>>OVER>>
6. How valuable was the change in course structure from the initial format: M, W, & F–computer days/T & TH reading days to the revised format: 1 hr. of reading/studying & 2 hrs. on computer EACH DAY?

1  2  3  4  5
not at all  very

comments/suggestions:

7. How valuable has this course been?  

1  2  3  4  5
not at all  very

comments/suggestions:

8. How would you improve this course?
GRE Preparation Course Evaluation  
Fall, 1994

1. How would you rate the Princeton book? 1 2 3 4 5  
   bad  good 
   comments/suggestions:

2. How would you rate the Barrons book? 1 2 3 4 5  
   bad  good 
   comments/suggestions:

3. How would you rate the Cliff's program? 1 2 3 4 5  
   bad  good 
   comments/suggestions:

4. How would you rate the Thinkfast program? 1 2 3 4 5  
   bad  good 
   comments/suggestions?

5. How would you rate the Algebra program? 1 2 3 4 5  
   bad  good 
   comments/suggestions:

>>OVER>>
6. To what extent do you think the $50 rebate controlled your behavior?

1 2 3 4 5
no extent at all great extent

comments/suggestions:

7. How valuable has this course been?

1 2 3 4 5
not at all very

comments/suggestions:

8. How would you improve this course?
Appendix H

Pretest and Posttest Scores of 1994 Enrolled Students
Table 9
Scores and Grade Point Averages of the 1994 Enrolled Students

<table>
<thead>
<tr>
<th>Student #</th>
<th>Pretest Score</th>
<th>Posttest Score</th>
<th>Grade Point Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>730</td>
<td>940</td>
<td>2.97</td>
</tr>
<tr>
<td>2</td>
<td>950</td>
<td>1,100</td>
<td>3.42</td>
</tr>
<tr>
<td>3</td>
<td>960</td>
<td>1,020</td>
<td>3.39</td>
</tr>
<tr>
<td>4</td>
<td>830</td>
<td>870</td>
<td>3.28</td>
</tr>
<tr>
<td>5</td>
<td>790</td>
<td>760</td>
<td>2.88</td>
</tr>
<tr>
<td>6</td>
<td>740</td>
<td>890</td>
<td>3.36</td>
</tr>
<tr>
<td>7</td>
<td>680</td>
<td>760</td>
<td>3.6</td>
</tr>
<tr>
<td>8</td>
<td>790</td>
<td>880</td>
<td>3.73</td>
</tr>
<tr>
<td>9</td>
<td>910</td>
<td>860</td>
<td>3.4</td>
</tr>
<tr>
<td>10</td>
<td>800</td>
<td>910</td>
<td>3.06</td>
</tr>
<tr>
<td>11</td>
<td>840</td>
<td>1,020</td>
<td>3.01</td>
</tr>
<tr>
<td>12</td>
<td>1,000</td>
<td>1,070</td>
<td>3.09</td>
</tr>
<tr>
<td>13</td>
<td>930</td>
<td>1,010</td>
<td>3.74</td>
</tr>
<tr>
<td>14</td>
<td>600</td>
<td>820</td>
<td>2.9</td>
</tr>
<tr>
<td>15</td>
<td>880</td>
<td>820</td>
<td>3.06</td>
</tr>
<tr>
<td>16</td>
<td>670</td>
<td>900</td>
<td>3.05</td>
</tr>
<tr>
<td>17</td>
<td>1,100</td>
<td>1,190</td>
<td>4</td>
</tr>
<tr>
<td>18</td>
<td>870</td>
<td>1,120</td>
<td>2.16</td>
</tr>
<tr>
<td>19</td>
<td>880</td>
<td>950</td>
<td>3.76</td>
</tr>
<tr>
<td>20</td>
<td>900</td>
<td>880</td>
<td>2.89</td>
</tr>
</tbody>
</table>
Table 9--Continued

<table>
<thead>
<tr>
<th>Student # (Drop Outs)</th>
<th>Pretest Scores Only</th>
<th>Grade Point Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>840</td>
<td>3.23</td>
</tr>
<tr>
<td>2</td>
<td>780</td>
<td>2.71</td>
</tr>
<tr>
<td>3</td>
<td>930</td>
<td>3.14</td>
</tr>
<tr>
<td>4</td>
<td>830</td>
<td>3.19</td>
</tr>
<tr>
<td>5</td>
<td>840</td>
<td>3.44</td>
</tr>
<tr>
<td>6</td>
<td>1,030</td>
<td>3.05</td>
</tr>
<tr>
<td>7</td>
<td>930</td>
<td>3.36</td>
</tr>
<tr>
<td>8</td>
<td>980</td>
<td>3.94</td>
</tr>
<tr>
<td>9</td>
<td>620</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>950</td>
<td>3.37</td>
</tr>
</tbody>
</table>

Subjects 1-5 were in the spring course; subjects 6-11 were in the summer course; and subjects 12-20 were in the fall course.
Appendix I

Human Subjects Institutional Review Board Approval
Date: April 19, 1995
To: Jan Miller
From: Richard Wright, Chair
Re: Old HSIRB Project Number 94-01-09
New HSIRB Project Number 95-04-19

This letter will serve as confirmation that an extension to your research project entitled "The effects of self-study on GRE Verbal and Quantitative scores" has been granted by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now continue to implement the research as described in the original application.

You must seek reapproval for any changes in this design. You must also seek reapproval if the project extends beyond the termination date. In addition if there are any unanticipated adverse or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

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

Approval Termination: April 19, 1996

xc: Richard Malott, PSY
BIBLIOGRAPHY


Cliffs StudyWare for the GRE (1992) (Version 4.3) [Computer software]. Lincoln, NE: Cliffs Notes.


Intelligent Computer Tutorial (ICT) for Basic Algebra. [Computer software]. (1993). Spring, TX.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


