Research Report
No. 2006-9

Investigating the Effects of Increased SAT Reasoning Test™ Length and Time on Performance of Regular SAT® Examinees

Xiang Bo Wang

www.collegeboard.com

College Board Research Report No. 2006-9

Investigating the Effects of Increased SAT Reasoning Test™ Length and Time on Performance of Regular SAT® Examinees

Xiang Bo Wang

The College Board, New York, 2007

Acknowledgments

The author wishes to sincerely thank Brian Patterson for the use of his unscrambling programs, Gerald Melican and Amy Schmidt for their valuable review and feedback, and Carrie Dirks for her careful copy editing.

Xiang Bo Wang is a senior psychometrician at the College Board.

Researchers are encouraged to freely express their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

The College Board: Connecting Students to College Success

The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 5,200 schools, colleges, universities, and other educational organizations. Each year, the College Board serves seven million students and their parents, 23,000 high schools, and 3,500 colleges through major programs and services in college admissions, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT®, the PSAT/NMSQT®, and the Advanced Placement Program® (AP®). The College Board is committed to the principles of excellence and equity, and that commitment is embodied in all of its programs, services, activities, and concerns.

For further information, visit www.collegeboard.com.

Additional copies of this report (item #060481980) may be obtained from College Board Publications, Box 886, New York, NY 10101-0886, 800 323-7155. The price is $15. Please include $4 for postage and handling.

© 2007 The College Board. All rights reserved. College Board, Advanced Placement Program, AP, SAT, and the acorn logo are registered trademarks of the College Board. connect to college success and SAT Reasoning Test are trademarks owned by the College Board. PSAT/NMSQT is a registered trademark of the College Board and National Merit Scholarship Corporation. All other products and services may be trademarks of their respective owners. Visit the College Board on the Web: www.collegeboard.com.

Printed in the United States of America.

Contents

Tables

Figures

Executive Summary

In March 2005, the new SAT Reasoning Test™ was launched with the addition of a writing section. Before that, the SAT® I: Reasoning Test had been administered without writing. With the additional section, the testing time of the SAT Reasoning Test has been increased from 3 hours to 3 hours and 45 minutes. Has the 45-minute time increase adversely affected examinee performance on the SAT?

The SAT has a total of eight operational multiple-choice (MC) sections: three critical reading (CR), three mathematics, and two writing (WR) sections, and one variable section that is used for item pretesting or test equating purposes. Two unique features of these nine sections and one feature of SAT examinees were utilized in this research. First, the three CR sections and three math sections are constructed to be as comparable as possible in terms of difficulty level, item number, and content specifications. Although their average difficulty levels are also similar, the two WR sections differ substantially in their numbers of items. Second, the nine MC sections are interspersed and arranged in different orders to make up different test-form variants. As a result, sections of similar content and difficulty are administered in different places during an SAT administration. For example, one math section can appear as the second section on one test form, and it can also be the seventh section on a different variation of a test form during the same SAT administration. Finally, test booklets of different test-form variants are randomly spiraled among examinees, resulting in randomly equivalent examinee subpopulations for all test-form variants.

Four performance-related indices were constructed for each examinee on each section: the right score ratio, wrong score ratio, the omit ratio, and the omit ratio on the last six items on a section. These indices, especially the last two, can directly reflect the possible impact on behavior and performance on a single section.

Three data sets were analyzed in this research. The first set consisted of 288,905 examinees from the first SAT Reasoning Test administration in March 2005, when the impact of an increased test length, if it existed, was assumed to be the most salient. The second data set was of 409,040 examinees from the October 2005 administration to cross-validate findings from the March 2005 SAT administration. In order to compare examinee performance trends between the current SAT Reasoning Test and its predecessor, the SAT I: Reasoning Test, the third data set consisting of data on 437,434 examinees from the May 2002 administration was also included.

This research examines the effect of increased testing time by comparing the four performance indices of randomly equivalent examinee subpopulations on sections of similar content and difficulty administered at different times on three SAT administrations. A variety of analyses were used in this study and found no evidence that the current SAT test length has affected examinee performance at the population level or differentially across gender, racial/ethnic, and best-language subgroups. On the contrary, this study produced consistent findings that examinees performed virtually identically on sections of similar content and difficulty, both marginally and conditionally, throughout the entire SAT. Furthermore, the findings from the March and October 2005 SAT administration data were replicated using the May 2002 SAT administration data, indicating no significant changes in performance trends between the two tests. Explanations of the findings are presented from a theoretical perspective, including a review of past research.

Introduction

The SAT Reasoning Test, launched in March 2005 with a new writing component, was created to meet the national call to assess the writing of high school students in America (College Board, 2004). Due to this addition and other improvements, the total testing time for the SAT increased to 3 hours and 45 minutes from its previous time of 3 hours. There has been concern that examinee performance may be impaired by the increased test length and time (Cloud, 2003; Mathews, 2006; Black Issues in Higher Education, 2005). The purpose of this study is to investigate the extent to which the new test length has affected the performance of regular SAT examinees who took the SAT without any accommodations.[1]

From a psychometric point of view, the optimal test length for an educational assessment depends mainly on two closely linked factors after the content, type, and quality of items are decided. The first factor is the ideal number of items that are required to produce a reliable test score (Hambleton and Swaminathan, 1985). The second is the total amount of time necessary to complete all items on a test so that an optimal balance is struck for the majority of examinees between performance and efficiency.

The current number of items for the critical reading, math, and writing sections of the SAT Reasoning Test has been selected through multiple stages of research, experimentation, and consultation with measurement experts and content specialists. Extensive research

[1]A separate study is under way to investigate the effect of increased time on SSD students with different accommodations. Note that the author of this study does not deny the possibility that examinees may feel more fatigued with the increased test length and time. The key point here is whether examinee performance has been differentially and adversely affected.

results have been reviewed by the SAT Psychometric Advisory Panel, which oversaw the development of the SAT Reasoning Test. Reports have also been presented at professional conferences (Liu and Feigenbaum, 2004; Reshestar, Liu, and Feigenbaum, 2004; Walker and Liu, 2004; Liu, Feigenbaum, and Dorans, 2004).

Time limits for the SAT and other testing programs have been set based on historical analysis, research, and review by experts. Additionally, time limits have been estimated for items of various types (Bridgeman, Cahalan, and Cline, 2003). Examples of such estimates include 1.0 to 1.2 minutes for one reading item, and 0.7 minutes per sentence completion item.

Based on various but relatively small sample sizes of students of different PSAT/NMSQT® score ranges, racial compositions, and school types, Bridgeman, Cahalan, and Cline (2003) estimated time requirements for the proposed new SAT items through three approaches: (1) via a computer-adaptive testing (CAT) version of the SAT system, which automatically recorded student response times under minimal time pressure; (2) by closely watching and recording how fast students responded to SAT items under strict time limits; and (3) by asking students to record their own response time on SAT sections of single item types in group test administration format. Recognizing that experimental test conditions could not fully represent those of real operational SAT administrations, these researchers concluded that although time estimates from Educational Testing Service (ETS) were accurate in rank ordering item types in terms of time requirements, they generally seemed to underestimate the time spent by students in the study.

Two other ETS writing studies also shed light on the time requirements for the new writing section of the SAT Reasoning Test. The first study by Livingston (1987) examined the differences in essay scores completed under three separate timing conditions: (1) 20 minutes; (2) 30 minutes; and (3) 30 minutes; divided into one 10-minute planning stage and another 20-minute writing stage. They found that an extra 10 minutes only increased the average score of high-ability students by about half a point, but did not seem to have benefited students of middle or low abilities. In addition, separating essay planning and writing clearly lowered students' scores when the essay preceded multiple-choice writing questions. Based on approximately 7,100 high school juniors and seniors, Crone, Wright, and Baron (1993) evaluated the effects of time on the SAT II: Writing Test and found that students clearly wrote better essays in 30 minutes than in 15 minutes across the board. This was also true of students of different racial/ethnic groups and language backgrounds. They concluded that providing additional time would result in better essays but would not advantage any group of students or have an impact on the score scale.

The question remains whether the increased testing time has caused sufficient examinee fatigue as to significantly affect the test performance of the entire examinee population both marginally across the board, and differentially across examinees of different gender, racial/ethnic, and language groups, especially as compared with its predecessor, the SAT I: Reasoning Test.

The extensive literature reviews by Burton and Kostin (2002) and Ackerman and Kanfer (2006) revealed a large body of literature on definitions of intellectual fatigue (Muscio, 1921; Bartley and Chute, 1947) and on models of cognitive fatigue (Schmidtke, 1976; Grandjean, 1970; Kahneman, 1973; Ackerman, 1988). Although generalities about the common factors that can attribute to, and alleviate, cognitive fatigue can be drawn from these studies (Thorndike, 1912; Myers, 1937; Cameron, 1973; Norman and Bobrow, 1975; Van der Linden, Frese, and Meijman, 2003), there has been little research on which conditions were sufficiently specific to those on the SAT Reasoning Test. The only research directly related to the SAT was conducted by Allspach, Feigenbaum, and Liu (2003). Allspach et al. surveyed 45 male and 52 female examinees[2] about their perceptions of fatigue, the adequacy of the number of breaks, and the amount of break time, among other issues. The examinees were randomly divided into three groups to take two versions of the SAT including one 3-hour SAT version and one 3-hour-and-35-minute simulated SAT. While it was difficult to achieve statistical significance due to the small sample sizes, three findings were clear. First, slightly more students who took the extended-time SAT indicated that they were "very tired" and "very hungry." Second, more students who took the extended-time SAT felt that two breaks of 10 minutes each were adequate. Third, the perceived negative effect of hunger on performance between the students who took the extended-time SAT versus the standard-time SAT was highly similar.

The objective of this research is to determine if student performance decreases due to fatigue on the SAT Reasoning Test that is now 3 hours and 45 minutes in length.

Research Approach

Seven general research issues have been addressed in this study:

1. To what extent did examinee scores change on sections of similar content and difficulty from the

[2]The sample size of 97 students was mainly determined by funding limits.

beginning to the end of the test?

2. To what extent did the numbers of items omitted and/or not reached by examinees change on sections of similar content and difficulty from the beginning to the end of the test?
3. To what extent did the findings to questions 1 and 2 vary across examinees of different gender?
4. To what extent did the findings to questions 1 and 2 vary across examinees of different racial/ethnic groups?
5. To what extent did the findings to questions 1 and 2 vary across examinees whose best language was either English or a different language?
6. To what extent could the findings to all the above questions be generalized across different SAT Reasoning Test administrations?
7. To what extent did the findings from the SAT Reasoning Test data hold true for the May 2002 SAT I: Reasoning Test administration?

Several features of both the SAT Reasoning Test and SAT I: Reasoning Test are related to this study. The first feature is the varied order of presentation of the different test sections on a particular SAT administration. It is well known that in order to reduce the opportunity for student copying, the nine operational sections of the SAT—three critical reading sections (R1, R2, and R3 hereafter), three math sections (M1, M2, and M3), two writing sections (W1 and W2), one essay section, and one variable section—are arranged in different orders in the test booklets to form different test spirals, although the essay section always appears first and a 10-minute writing section always appears last. For example, there was one primary test form used for the Saturday domestic March 2005 administration of the SAT, although different essays were used on the East and West Coast versions. Across the different test forms, the eight multiple-choice[3] (MC) item sections, namely R1 to R3, M1 to M3, and W1 to W2, are further presented in two orders, commonly referred to as the "main" versus the "scrambled" test spirals. For example, the main spiral of an SAT test presented R1 as its first MC section, while the scrambled spiral presented M3 as its first MC section.

Second, the overall difficulty and content coverage across different sections of the SAT are constructed to be as similar as possible. For example, the overall difficulty of the three critical reading sections (R1, R2, and R3) are highly parallel, though not exactly the same, as are the three math sections (M1, M2, and M3), and the two writing sections (W1 and W2).

Third, many SAT test forms are developed each year, and test booklets of different spirals within each test form are randomly distributed among examinees in order to maximize the likelihood that subpopulations of examinees taking both test spirals are as equal in ability as possible. More specifically, test booklets of the main and scrambled spirals for East Coast testing will be distributed to approximately equal numbers of examinees of different gender, racial/ethnic, and best-language groups. This random distribution of test booklets ensures that the overall abilities of examinees taking the two spirals of the test are as similar as possible across different gender, racial/ethnic, and best-language groups.

It is these three important features of the SAT—different spirals of test section presentation, parallel difficulties for sections of similar content and skill, and comparable examinee abilities—that will be utilized to investigate the effect of test length and position by comparing examinee performance on different sections of similar content and difficulty throughout the SAT test. In addition, all the analyses are replicated on a different new SAT Reasoning Test and SAT I: Reasoning Test administration in order to cross-validate the findings.

Research Data

This research was based on data from three different SAT administrations. The first set consisted of data from 288,905[4] examinees who took the Saturday domestic version of the March 2005 SAT administration. This administration was chosen because it was the first SAT Reasoning Test administration. If the increased test length did affect examinee performance, this effect most likely would be found during this administration, since later examinees could be more prepared for subsequent administrations. In order to assess the generalizability of the findings from the March 2005 administration, the analyses were replicated using data from the 409,040 examinees who took the October 2005 SAT Reasoning Test.

How much did examinee performance differ between the SAT Reasoning Test and the SAT I: Reasoning Test? To answer this question, a third data set consisting of 437,434 examinees who took the May 2002 SAT I: Reasoning Test was analyzed in order to ascertain any potential differences in the findings between the SAT Reasoning Test and the SAT I: Reasoning Test.

Due to the complexity and scope of the analyses and findings to be reported, this study is divided into three sections. The main body of this paper reports the analyses and findings on the March 2005 administration, while the results of the October 2005 and May 2002 administrations are summarized in Appendixes A and B, respectively.

[3]The term "multiple choice" is used to refer to both the objective and student-produced response items.

[4]4,695 SSD examinees and 85 examinees with missing test-form information were excluded from this study. Furthermore, 1,360 Sabbath examinees who took the Sunday test form were also excluded due to their self-selecting factor.

Although details differed slightly, the findings and conclusions from the March 2005 analysis were replicated in the October 2005 and the May 2002 administrations. While the reader is encouraged to refer to Appendixes A and B to verify replication results, the author has chosen only to briefly mention replication consistency.

SAT Reasoning Test™ Versus SAT® I: Reasoning Test Structure

Table 1 lists the number of examinees who took each of the four main test-form variants in the March 2005 SAT administration: (1) 108,789 regular examinees took the East Coast essay prompt on Saturday under the main spiral; (2) 105,807 examinees took the East Coast essay prompt on Saturday under the scrambled spiral; (3) 37,620 examinees took the West Coast essay prompt on Saturday under the main spiral; and (4) 36,689 examinees took the West Coast essay prompt on Saturday under the scrambled spiral. Note that administering two separate essay prompts for the East and West Coasts is a standard practice for test security reasons, but all four versions had the same set of operational multiple-choice items, albeit in different orders.

Two clear conclusions can be drawn from Table 1. First, the total East Coast examinee volume was about three times as large as the West Coast examinee volume for this administration. Second, between the East and West Coasts, the main and scrambled spirals received highly similar numbers of examinees, respectively, about 37 percent for the East Coast and 13 percent for the West Coast, indicating the success of random spiraling.

Table 1

Examinee Distributions Across the Four Test Forms and Two Spirals, East Versus West Coast Testing, March 2005 SAT Administration

System Form ID	*Frequency*	*Percent*	*Cumulative Frequency*	*Cumulative Percent*
East Coast Essay Prompt Saturday Main Spiral	108,789	37.66	108,789	37.66
East Coast Essay Prompt Saturday Scrambled Spiral	105,807	36.62	214,596	74.28
West Coast Essay Prompt Saturday Main Spiral	37,620	13.02	252,216	87.30
West Coast Essay Prompt Saturday Scrambled Spiral	36,689	12.70	288,905	100.00

Table 2 details the order in which the eight operational MC sections were presented for the March 2005 administration. (Here "Main" and "Scrb" refer to the main and scrambled spirals, respectively.) Two conclusions can be drawn. First, the MC section orders were the same within a particular spiral across different test forms. Specifically, the MC section order for the main spiral for the East Coast testing was the same as that of the main spiral for the West Coast testing. Second, between the two spirals, the MC sections of the same content were presented in different orders. In particular, under the main spiral, examinees encountered W1 as the second section and W2 as the eighth section, while under the scrambled spiral, W1 and W2 were presented as the third and the eighth sections, respectively. The configurations of sections both within and across different test spirals offer several comparison opportunities to assess test-length effect.

The details above demonstrate the complex structure of the March 2005 administration, and describe how this study employed the order and position of SAT sections for

Table 2

Order of Multiple-Choice Sections Across the Main and Scrambled Test Spirals, March 2005 SAT Administration*

System Form ID	*Spiral*	*Essay Prompt*	*Spiral Order of MC Sections*							
			1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
East Coast Essay Prompt Saturday Main Spiral	Main	1	M2	W1	R2	M1	R1	M3	R3	W2
East Coast Essay Prompt Saturday Scrambled Spiral	Scrb**	1	R1	M2	W1	R2	M1	R3	M3	W2
West Coast Essay Prompt Saturday Main Spiral	Main	2	M2	W1	R2	M1	R1	M3	R3	W2
West Coast Essay Prompt Saturday Scrambled Spiral	Scrb	2	R1	M2	W1	R2	M1	R3	M3	W2

*For test security reasons, this table omits the position of variable sections.
**Scrb is the abbreviation for Scrambled.

this research. Other SAT Reasoning Test administrations share these features, although their details vary. For example, as shown in Appendix A, the October 2005 administration had more test forms and spirals than did the March 2005 administration. While the SAT I: Reasoning Test differs from the current SAT Reasoning Test in that it did not have the essay and MC writing sections, it did follow similar rules in organizing test spirals, as exemplified in Appendix B.

Since the East versus West Coast distinction was not important to the current research purpose, this study combined all the examinees on the East and West Coasts who took the main spiral into one main spiral group, and all the examinees on the East and West Coasts who took the scrambled spiral were combined into the scrambled spiral group.

Research Analyses

This section will summarize eight sets of analyses conducted to examine the impact of fatigue on test performance.

1. Overall difficulty level and number of items for each of the eight operational sections
2. Examinee ability levels in reading, math, and writing by the two test spirals
3. Examinee performance accuracy (i.e., number correct) and omit tendency throughout the eight operational sections across the two test spirals based on summary statistics
4. Examinee performance accuracy and omit tendency throughout the eight operational sections across the two test spirals based on correlations
5. Examinee performance accuracy and omit tendency throughout the eight operational sections across the two test spirals based on summary statistics conditional on total test score
6. Examinee performance accuracy and omit tendency by gender (replication of relevant analyses)
7. Examinee performance accuracy and omit tendency by racial/ethnic group (replication of relevant analyses)
8. Examinee performance accuracy and omit tendency by different language groups (replication of relevant analyses)

Due to space limitations, only select results will be reported for analyses 6 and 8.

Analyses and Results

Analysis 1: Overall Difficulty Levels and Numbers of Items of the Eight Operational Sections

Research Questions: How many items were there in each section? How similar were sections of similar content in item difficulty?

As indicated above, similarity in section difficulty is one of the three premises of this study. Table 3 summarizes the numbers of items, and the means and standard deviations of the difficulty levels of the eight MC sections. Two conclusions can be made. First, different sections with the same content differed in their numbers of items. For example, the numbers of items in the three math sections ranged from 20 items for M1 to 16 items for M3. The number of items of W2 (14) was less than half of that of W1 (35).

Despite differences in items per section, the average difficulty (percent correct) and their associated standard deviations for sections with the same content were virtually identical. For example, the difficulty levels of the three reading sections were 0.56, 0.59, and 0.58 for R1, R2, and R3, respectively.[5] Note that the section difficulty levels were computed on the basis of pretest statistics only. Highly similar levels of item difficulty were found for sections of similar content for the October 2005 administration (SAT Reasoning Test) and the May 2002

Table 3

Average Section Difficulty, March 2005 SAT Administration

Section	# of Items	Percent Correct		Weight of One Item
		Mean	Std	
M1	20	0.58	0.22	0.05
M2	18	0.51	0.25	0.06
M3	16	0.57	0.20	0.06
R1	25	0.56	0.21	0.04
R2	23	0.59	0.22	0.04
R3	19	0.58	0.17	0.05
W1	35	0.64	0.20	0.03
W2	14	0.67	0.21	0.07

[5]Two points should be clarified here. First, the average percent correct values were based on the pretest statistics. Second, the percent correct values were not equated. Because not all items had equated deltas, a common index used by ETS, percent correct values were used. However, given the highly stable SAT examinee populations, percent correct values are relatively stable enough for this research. The percent correct values reported are based on the number of students answering items correctly divided by the number of students who reached the item. That is, students who did not respond to an item or any other item sequentially following the item are not included in the denominator.

administration (SAT I: Reasoning Test). The last column shows the percentage, or weight, of a particular item number in each of the eight sections that will be referred to later in this study.

Analysis 2: Examinee Ability Levels in Reading, Math, and Writing by the Two Test Spirals

Research Question: *How similarly did examinees perform on the critical reading, math, and writing sections between the main and scrambled spirals on the March 2005 administration?*

Table 4 summarizes the means and standard deviations for the total correct raw scores on the critical reading (CR), math, and writing (WR) sections between the main and scrambled spirals on the March 2005 SAT administration. The means and standard deviations were all virtually identical, differing only by one-tenth of a percent. Furthermore, Figures 1 to 3 show that the percentages of examinees conditional on total correct scores of critical reading, math, and writing sections were also virtually identical. Based on these findings, two conclusions can be made. First, the ability levels of examinees who took the main and scrambled test spirals of the March 2005 administration were virtually identical. This confirmed the premise for this study as discussed earlier. This conclusion was expected, as these two spirals were randomly distributed throughout examinees on both the East and West Coasts. Second, although the two spirals did differ in the order of presenting the three critical reading, three math, and two writing sections, such a difference in order did not seem to exert any significant effect on the three total section scores.

Virtually identical performances were also confirmed for the October 2005 and the May 2002 SAT administrations, as shown in Appendixes A and B.

Table 4

Descriptive Statistics on Critical Reading, Math, and Writing Sections, March 2005 SAT Administration

Spiral	*Frequency*	*CR Mean*	*CR Std*	*Math Mean*	*Math Std*	*WRMC Mean*	*WRMC Std*
Main	146,409	39.11	13.32	32.30	11.00	30.73	8.91
Scrb	142,496	39.24	13.22	32.35	11.01	30.67	8.92

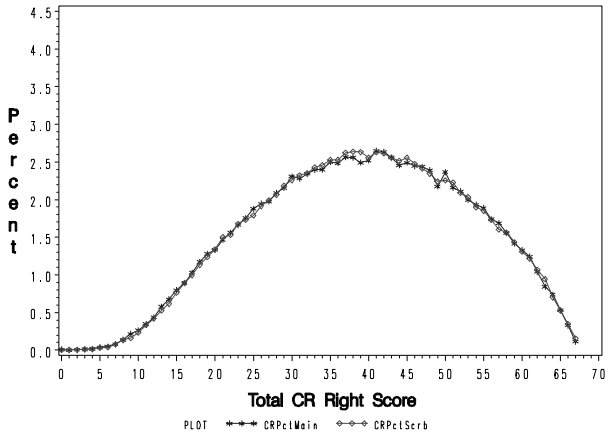

Figure 1. Total critical reading score distributions, March 2005 SAT administration.

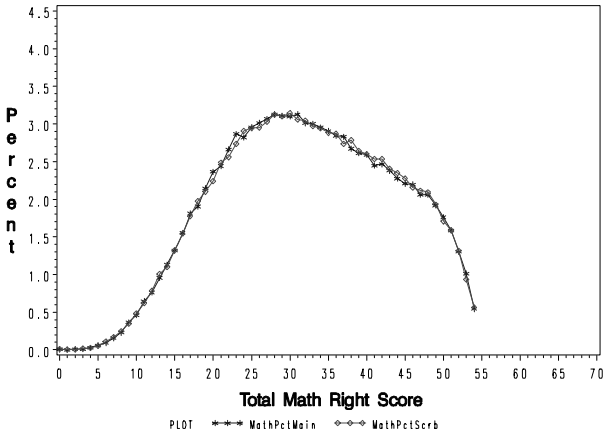

Figure 2. Total math score distributions, March 2005 SAT administration.

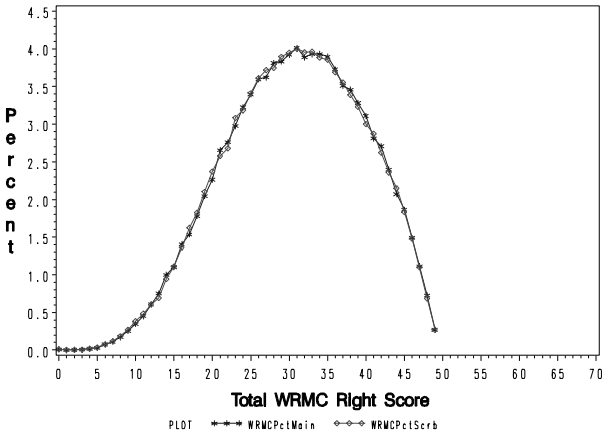

Figure 3. Total MC writing score distributions, March 2005 SAT administration.

Analysis 3: Examinee Performance Accuracy (i.e., Number Correct) and Omit Tendency Throughout the Eight Operational Sections Across the Two Test Spirals Based on Summary Statistics

Research Questions: *How similar was examinee performance on sections of similar content within each of the two spirals? How similar was examinee performance on sections of similar content when ordered differently between the two test spirals?*

As shown in Table 2, the numbers of items in the eight operational sections were different. This was especially true of the two writing sections. In order to compare performance on different sections, five ratio indices were created and computed for each examinee and for each of the eight operational multiple-choice sections:

- *A right score ratio:* A total number of right responses divided by the total number of items in a section. This index represents an examinee's correct performance level on a section. When applied to the eight operational multiple-choice sections later, this index is named "R1RRatio," "R2RRatio," "R3RRatio," "M1RRatio," "M2RRatio," "M3RRatio," "W1RRatio," or "W2RRatio," meaning the right ratios for R1, R2, R3, M1, M2, M3, W1, or W2 sections, respectively.
- *A wrong score ratio:* The total number of incorrect responses divided by the total number of items in a section. This index stands for an examinee's incorrect performance level on a section. Note that this index is not always the full complement of the first index, because examinees sometimes also omit or skip some items. When applied to the eight operational multiple-choice sections, this index is named "R1WRatio," "R2WRatio," "R3WRatio," "M1WRatio," "M2WRatio," "M3WRatio," "W1WRatio," or "W2WRatio," meaning the wrong ratios for R1, R2, R3, M1, M2, M3, W1, or W2 sections, respectively. Results using the wrong score ratios have largely been omitted in this study for reasons of limited space, but are available upon request.
- *An omit ratio:* The total number of omitted items divided by the total number of items in a section. This index reflects an examinee's levels of uncertainty on a section, and/or that the examinee was running out of time. When applied to the eight operational multiple-choice sections later, this index is named "R1ORatio," "R2ORatio," "R3ORatio," "M1ORatio," "M2ORatio," "M3ORatio," "W1ORatio," or "W2ORatio," meaning the omit ratios for R1, R2, R3, M1, M2, M3, W1, or W2 sections, respectively. Note that the current SAT scoring policy uses a correction-for-guessing scoring method to correct for random guessing, i.e., formula scoring. As a result, examinees tend not to guess randomly, and the extent of omits should be a relatively reliable indicator of knowledge uncertainty and/or running out of time.
- *A last-six-item omit ratio:* The total number of omitted responses on the last six items of a section divided by the total number of items in that section. This index is designed to resemble more closely the situation in which an examinee runs out of time toward the end of a test section. When applied to the eight operational multiple-choice sections later, this index is named "R1Last6OmitRatio," "R2Last6OmitRatio," "R3Last6OmitRatio," "M1Last6OmitRatio," "M2Last6OmitRatio," "M3Last6OmitRatio," "W1Last6OmitRatio," or "W2Last6OmitRatio," meaning the omit ratios for the last six items for R1, R2, R3, M1, M2, M3, W1, or W2 sections, respectively.
- *Total right scores:* The sum of all right answers from all the sections of the same content. This index represents the ability level of an examinee. When applied to the three content areas of the SAT, this index appears as a "Total CR Rights," "Total Math Rights," and "Total Writing Rights." Later in this study, these three total scores will be used as conditioning ability variables.

For most items in SAT math, critical reading sentence completions, and multiple-choice writing, it is a common practice to place relatively easier items at the beginning of a section and harder ones toward the end, resulting in an interaction of item difficulty and omits. In other words, when items were omitted toward the end of a section, it could be due to two confounding reasons: first, an examinee ran out of time; and second, an examinee found them too hard to solve.

Table 5 summarizes the means of the first four ratio indices across the eight operational sections between the two test spirals for the March 2005 SAT administration. Sharing the same headings and layout as Table 5, Table 6 indicates the differences in the mean ratios between the first and the later appearing sections of the same content. One feature in the two tables is worth noting. The two horizontal headings in the table list the abbreviated titles for the eight operational sections in their original test order with respect to their test spirals. Specifically, the section order for the main test spiral was M2, W1, R2, M1, R1, M3, R3, and W2, while the section order for the scrambled form was R1, M2, W1, R2, M1, R3, M3, and W2.

Table 5

Summary of Four Mean Ratios Across Eight Sections and Two Spirals, March 2005 SAT Administration

MC Section Order			*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Spiral	*Frequency*	*Ratio*	*M2*	*W1*	*R2*	*M1*	*R1*	*M3*	*R3*	*W2*
Main	146,409	Right	0.561	0.598	0.604	0.638	0.566	0.589	0.582	0.699
		Wrong	0.282	0.351	0.303	0.245	0.352	0.276	0.329	0.267
		Omit	0.156	0.050	0.092	0.116	0.082	0.135	0.089	0.034
		Last Six Omit	0.092	0.021	0.039	0.074	0.030	0.094	0.041	0.027
Spiral	*Frequency*	*Ratio*	*R1*	*M2*	*W1*	*R2*	*M1*	*R3*	*M3*	*W2*
Scrb	142,496	Right	0.567	0.564	0.599	0.606	0.638	0.586	0.589	0.694
		Wrong	0.358	0.277	0.355	0.303	0.249	0.324	0.276	0.271
		Omit	0.075	0.159	0.046	0.091	0.114	0.090	0.135	0.035
		Last Six Omit	0.028	0.095	0.019	0.037	0.072	0.041	0.094	0.028

Combining Tables 5 and 6 reveals three clear trends. First, within both the main and scrambled test spirals, the mean right, wrong, omit, and last-six-omit ratios were virtually identical, resulting in very small differences in mean ratios, mostly below 0.05 or 5 percent, except for three sections (M1, W2, and R2). Note that the positive difference on the right ratio and the negative differences on the wrong, omit, and last-six-omit ratios in Table 6 reflect that examinees performed better on later sections than on the earlier sections of the same content. In two of the three sections where differences exceeded 0.05 (or 5 percent), student performance was better on sections that appeared later in the test than earlier. To put these proportions in perspective, based on the last column in Table 3, "Weight of Item" indicates that one item constitutes about 0.03 (or 3 percent) to 0.07 (or 7 percent) of a section. The small mean ratio differences in Table 6 mean that examinees' average performance differences between the first and later appearing sections of the same content was by no more than one item.

Second, the four mean ratios of all corresponding sections between the main and scrambled test spirals were also highly similar, differing mostly by thousandths of a percent. For example, R1 was presented as the fifth section under the main spiral, and R1, as the first section under the scrambled spiral. In other words, these two sections of identical items were presented to examinees about 150 minutes apart across the two spirals. Yet, the four indices of R1 differed only by the third decimal place between the main and scrambled spirals, signifying little effect of fatigue.

Third, combining the information on the average section difficulty in Table 3, and the trends of mean ratios in Table 5, one can safely conclude that on average, examinees did not appear to have suffered

Table 6

Summary of Four Mean Ratios and Mean Ratio Differences Across Eight Sections and Two Spirals, March 2005 SAT Administration

Section Order			*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Spiral	*Frequency*	*Ratio*	*M2*	*W1*	*R2*	*M1*	*R1*	*M3*	*R3*	*W2*
Main	146,409	Right	0.561	0.598	0.604	0.077	-0.038	0.028	-0.022	0.101
		Wrong	0.282	0.351	0.303	-0.037	0.049	-0.006	0.026	-0.084
		Omit	0.156	0.050	0.092	-0.040	-0.010	-0.021	-0.003	-0.016
		Last Six Omit	0.098	0.024	0.045	-0.018	-0.009	0.002	0.002	0.006
Spiral	*Frequency*	*Ratio*	*R1*	*M2*	*W1*	*R2*	*M1*	*R3*	*M3*	*W2*
Scrb	142,496	Right	0.567	0.564	0.599	0.039	0.074	0.019	0.025	0.095
		Wrong	0.358	0.277	0.355	-0.055	-0.028	-0.034	-0.001	-0.084
		Omit	0.075	0.159	0.046	0.016	-0.045	0.015	-0.024	-0.011
		Last Six Omit	0.032	0.100	0.022	0.009	-0.023	0.013	-0.001	0.009

significantly on later sections of the same content from the viewpoint of mean right ratios. On the contrary, seeing mostly higher mean right ratios, lower mean wrong and omit ratios for most later appearing math, writing, and reading sections along both the main and scrambled test spirals, if any difference is evident it would suggest that examinees might have benefited from later sections as a result of what is commonly known as the warm-up effect. There are two reasons to reject this speculation.

The first reason to eliminate the warm-up effect as an explanation is that the pattern of the mean right ratios of the eight operational sections in Table 5 matched completely the pattern of their average section difficulty as shown in Table 3. In other words, the pattern of higher mean right ratios could have been caused by the pattern of average lower section difficulty levels.

Second, a close look at the presentation order of R1 and R2 across the two spirals also reinforces this rejection. As shown in Table 3, R1 section was slightly harder than R2 section. Under the main spiral, R1 was the second reading section and had a slightly lower mean right ratio than that of R2, the first reading section. However, under the scrambled test spiral, R1 was the first reading section and still maintained a virtually identical mean right ratio, which was lower than that of R2, the second reading section. If the order of presentation had any systematic effect on examinee performance, it could have changed the mean right ratio of R1.

Replications show that all findings in this section, including the tendencies of examinee response accuracy and omits, were true for the October 2005 and the May 2005 SAT administrations.

Analysis 4: Examinee Performance Accuracy and Omit Tendency Throughout the Eight Operational Sections Across the Two Test Spirals Based on Correlations

Research Question: *How consistently did examinees perform from one reading section to another in terms of the rates of their correct responses across the main and scrambled spirals?*

Table 7 summarizes the mean, standard deviations, and correlations of examinees' right score ratios among the three reading sections of the March 2005 SAT administration across the main and scrambled spirals. Note that two mean ratios in this table matched those reported in Table 5 as expected. Three findings are apparent. First, the six mean ratios were nearly identical to the corresponding R1, R2, and R3 mean proportions of right scores reported in Table 3, signifying that the overall reading ability level of the March 2003 administration was highly similar to that of the examinees in the sample used to calculate the pretest reading item statistics. Second, all the correlation coefficients were high and highly similar, ranging from 0.776 to 0.811, indicating similar correct test performance throughout the three reading sections. The small differences in the correlations can be explained by the small differences in the numbers of items across the three reading sections. Recall that R1, R2, and R3 had 25, 23, and 19 items, respectively, and the correlations among the three reading sections decreased accordingly. Third, between the main and scrambled spirals, the correlations among the three reading sections were nearly identical, signifying highly parallel accurate performance. Specifically, the accuracy correlations between R1 and R2 across the main and scrambled spirals were 0.811 and 0.803, respectively, signifying little or no order effect. Examinees' accuracy rates were highly consistent across the three reading sections and two test spirals.

Given the fact that examinees' correct performance correlated highly negatively with their incorrect performance, similar analyses and results on examinees' incorrect response ratios in reading, math, and writing were omitted in this report for brevity.

Table 7

Means, Standard Deviations, and Correlations of Examinees' Right Response Rates Across R1, R2, and R3 and Two Test Spirals, March 2005 SAT Administration

Spiral	*Index*	*Variable*	*R1RRatio*	*R2RRatio*	*R3RRatio*
Main	MEAN		0.566	0.604	0.582
	STD		0.219	0.207	0.215
	N		146,409	146,409	146,409
	CORR	R1RRatio*	1.000	0.811	0.793
	CORR	R2RRatio	0.811	1.000	0.776
	CORR	R3RRatio	0.793	0.776	1.000
Scrb	MEAN		0.567	0.606	0.586
	STD		0.214	0.208	0.216
	N		142,496	142,496	142,496
	CORR	R1RRatio	1.000	0.803	0.787
	CORR	R2RRatio	0.803	1.000	0.783
	CORR	R3RRatio	0.787	0.783	1.000

*"R1RRatio" means the "right ratio for critical reading section 1." Please refer to page 7 for the definitions of such abbreviated indices.

Research Question: *What was the trend of examinees' sectional omit rates from one reading section to another across the main and scrambled spirals?*

Sectional omit rates can reflect the extent that a student rushes through an entire section due to fatigue and/or lack of knowledge to answer questions. Table 8 summarizes the means, standard deviations, and correlations among examinees' omit ratios across the three reading sections and the two test spirals. The average omit rates were fairly low, ranging between 8 and 9 percent, equivalent to about two omitted items, and omit rate correlation coefficients are consistent (ranging from 0.709 to 0.776). On average, examinees' omit rates are similar across the three reading sections and two test spirals, and the order and position did not seem to have any noticeable effect on examinees' omit pattern at the group level.

Table 8

Means, Standard Deviations, and Correlations of Examinees' Omit Rates Across R1, R2, and R3 and Two Test Spirals, March 2005 SAT Administration

Spiral	*Index*	*Variable*	*R1ORatio*	*R2ORatio*	*R3ORatio*
Main	MEAN		0.082	0.092	0.089
	STD		0.122	0.134	0.131
	N		146,409	146,409	146,409
	CORR	R1ORatio	1.000	0.776	0.768
	CORR	R2ORatio	0.776	1.000	0.726
	CORR	R3ORatio	0.768	0.726	1.000
Scrb	MEAN		0.075	0.091	0.090
	STD		0.112	0.135	0.130
	N		142,496	142,496	142,496
	CORR	R1ORatio	1.000	0.746	0.709
	CORR	R2ORatio	0.746	1.000	0.756
	CORR	R3ORatio	0.709	0.756	1.000

Research Question: *What was the trend of examinees' omit rates on the last six items from one reading section to another across the main and scrambled spirals?*

As discussed earlier, to investigate more closely the extent of rushing, tiredness, and/or lack of knowledge, the omit rates on the last six items at the end of each section were examined, and Table 9 summarizes the means, standard deviations, and correlations of such omit ratios across the three reading sections and the two test spirals. Two findings were clear: The average omit rates on the last six items were under 5 percent across the three reading sections and the two test spirals. This 5 percent is equivalent to about one item. Based on the average 8 percent sectional omit rates reported in Table 8, it can be inferred that half of the omits on the reading section occurred on the last six items. Second, the correlations among the three reading sections across the two spirals were highly similar, ranging from 0.557 to 0.628, reflecting that at the group level, examinees tended to omit in a similar way on the last six items across the three reading sections. Based on these two findings, neither the section order nor the location seemed to have a significant impact on examinees' performance toward the end of each reading section.

Table 9

Means, Standard Deviations, and Correlations of Examinees' Omit Rates on the Last Six Items Across R1, R2, and R3 and Two Test Spirals, March 2005 SAT Administration

Spiral	*Index*	*Variable*	*R1Last6OmitRatio*	*R2Last6OmitRatio*	*R3Last6OmitRatio*
Main	MEAN		0.030	0.039	0.041
	STD		0.061	0.072	0.075
	N		146,409	146,409	146,409
	CORR	R1Last6OmitRatio	1.000	0.615	0.627
	CORR	R2Last6OmitRatio	0.615	1.000	0.567
	CORR	R3Last6OmitRatio	0.627	0.567	1.000
Scrb	MEAN		0.028	0.037	0.041
	STD		0.059	0.069	0.074
	N		142,496	142,496	142,496
	CORR	R1Last6OmitRatio	1.000	0.577	0.557
	CORR	R2Last6OmitRatio	0.577	1.000	0.596
	CORR	R3Last6OmitRatio	0.557	0.596	1.000

Research Question: How consistent were examinees' response accuracy rates across the three math sections and two test spirals?

Table 10 summarizes the means, standard deviations, and correlations of examinees' right score ratios among the three math sections of the March 2005 SAT administration across the main and scrambled spirals. Three findings stand out. First, the two pairs of mean right ratios between the two spirals were virtually identical, confirming little difference in examinee overall math ability levels between the two spirals. Second, all the correlation coefficients were highly similar, ranging from 0.793 to 0.839, generally indicating highly similar accurate performance throughout the three math sections. Again, the small differences in the correlations could have been explained by the small differences in the numbers of items across the three math sections. Recall that M1, M2, and M3 had 20, 18, and 16 items, respectively, and note that the correlations among the three math sections decreased accordingly. Third, between the main and scrambled spirals, the correlations among the three math sections were virtually identical, signifying highly parallel accurate performance. For example, the accuracy correlations between M1 and M2 between the main and scrambled spirals were 0.839 and 0.840, respectively, signifying little or no order effect. Examinees' accuracy rates were highly consistent across the three math sections, two test spirals, and the order of the three math sections. The location did not seem to affect examinees' performance on the math section.

Research Question: What was the trend of examinees' sectional omit rates across the three math sections and the main and scrambled spirals?

Table 11 summarizes the means, standard deviations, and correlations on examinees' omit ratios throughout the three math sections between the two test spirals. Again, the math sectional mean omit ratios were still highly similar (ranging from 0.114 to 0.159), and the math sectional mean omit ratios were about 2 to 5 percent higher than reading. Examinees tended to omit about one more item on each of the three math sections than they did on each of the three reading sections, even though the numbers of items in the three math sections were lower than those for the three reading sections. Such a small increase in the average math sectional omit rates seemed to reflect the fact that, overall, the math sections were slightly harder than the three reading sections as reflected by their corresponding percent correct values in Table 3. Correlations among the sectional omit ratios were also highly similar among the three math sections across the two test spirals, ranging from 0.617 to 0.743. In particular, the correlation between M2 and M1 was virtually identical to that between M2 and M3, both within each spiral and across the two spirals—all around 0.6. Recall that the presentation order of the three math sections was M2, M1, and M3. These findings point to the clear conclusion that the location did not seem to have affected the sectional omit rates on the three math sections.

Table 10

Means, Standard Deviations, and Correlations of Examinees' Right Response Rates Across M1, M2, and M3 and Two Test Spirals, March 2005 SAT Administration

Spiral	*Index*	*Variable*	*M1RRatio*	*M2RRatio*	*M3RRatio*
Main	MEAN		0.638	0.561	0.589
	STD		0.212	0.233	0.206
	N		146,409	146,409	146,409
	CORR	M1RRatio	1.000	0.839	0.795
	CORR	M2RRatio	0.839	1.000	0.802
	CORR	M3RRatio	0.795	0.802	1.000
Scrb	MEAN		0.638	0.564	0.589
	STD		0.213	0.233	0.206
	N		142,496	142,496	142,496
	CORR	M1RRatio	1.000	0.840	0.793
	CORR	M2RRatio	0.840	1.000	0.801
	CORR	M3RRatio	0.793	0.801	1.000

Table 11

Means, Standard Deviations, and Correlations of Examinees' Omit Rates Across M1, M2, and M3 and Two Test Spirals, March 2005 SAT Administration

Spiral	*Index*	*Variable*	*M1ORatio*	*M2ORatio*	*M3ORatio*
Main	MEAN		0.116	0.156	0.135
	STD		0.129	0.150	0.145
	N		146,409	146,409	146,409
	CORR	M1ORatio	1.000	0.673	0.736
	CORR	M2ORatio	0.673	1.000	0.622
	CORR	M3ORatio	0.736	0.622	1.000
Scrb	MEAN		0.114	0.159	0.135
	STD		0.129	0.149	0.144
	N		142,496	142,496	142,496
	CORR	M1ORatio	1.000	0.671	0.743
	CORR	M2ORatio	0.671	1.000	0.617
	CORR	M3ORatio	0.743	0.617	1.000

Table 12

Means, Standard Deviations, and Correlations of Examinees' Omit Rates on the Last Six Items Across M1, M2, and M3 and Two Test Spirals, March 2005 SAT Administration

Spiral	*Index*	*Variable*	*M1Last6OmitRatio*	*M2Last6OmitRatio*	*M3Last6OmitRatio*
Main	MEAN		0.078	0.098	0.100
	STD		0.085	0.101	0.104
	N		146,409	146,409	146,409
	CORR	M1Last6OmitRatio	1.000	0.510	0.654
	CORR	M2Last6OmitRatio	0.510	1.000	0.471
	CORR	M3Last6OmitRatio	0.654	0.471	1.000
Scrb	MEAN		0.076	0.100	0.101
	STD		0.084	0.100	0.104
	N		142,496	142,496	142,496
	CORR	M1Last6OmitRatio	1.000	0.502	0.666
	CORR	M2Last6OmitRatio	0.502	1.000	0.468
	CORR	M3Last6OmitRatio	0.666	0.468	1.000

Research Question: How consistently did examinees omit the last six items on the three math sections across the two test spirals?

Table 12 summarizes the means, standard deviations, and correlations of examinees' omit ratios on the last six items on the three math sections of the March 2005 SAT administration across the main and scrambled spirals. Based on the patterns of findings similar to those on sectional math omits, the position did not seem to have any significant effect on how examinees omitted the last six items on each of the three math sections. Finally, in comparing Tables 11 and 12, it is clear that more than half of the omits on a math section occurred on the last six items. Recall that this finding was true of the reading sections as well.

Research Question: How consistently did examinees perform throughout the two writing sections across the two test spirals?

Table 13 summarizes the means, standard deviations, and correlations of examinees' accuracy ratios on the two writing sections of the March 2005 SAT administration across the main and scrambled spirals. Of note is that the average right ratio on the second writing section (W2) was about 10 percent higher than that of the first writing section (W1). This is the largest increase in the average right ratio observed thus far, and is likely attributable to the fact that W2 had fewer than half the items of W1, assuming that shorter sections tended to be easier for students to solve, and given the comparable difficulty levels for both the short and long sections. W2 had 14 items while W1 had 35 items (Table 3).[6] This mean right ratio differential will most likely remain in subsequent administrations, given that W2 will always be substantially shorter than W1 according to current SAT test specifications.

Second, in spite of the small number of items (14) in W2, the correlations between W1 and W2 were as high as 0.752 for the main spiral, and 0.754 for the scrambled spiral, showing highly similar accuracy ratios across the two test spirals.

Research Question: How consistently did examinees omit between the two writing sections across the two test spirals?

Table 14 summarizes the means, standard deviations, and correlations of examinees' omit ratios on the two writing

Table 13

Means, Standard Deviations, and Correlations of Examinees' Right Response Rates Across W1 and W2 and Two Test Spirals, March 2005 SAT Administration

Spiral	*Index*	*Variable*	*W1RRatio*	*W2RRatio*
Main	MEAN		0.598	0.699
	STD		0.190	0.197
	N		146,409	146,409
	CORR	W1RRatio	1.000	0.752
	CORR	W2RRatio	0.752	1.000
Scrb	MEAN		0.599	0.694
	STD		0.190	0.197
	N		142,496	142,496
	CORR	W1RRatio	1.000	0.754
	CORR	W2RRatio	0.754	1.000

[6]W1 is a 25-minute section and W2 is a 10-minute section.

Table 14

Means, Standard Deviations, and Correlations of Examinees' Omit Rates Across W1 and W2 and Two Test Spirals, March 2005 SAT Administration

Spiral	*Index*	*Variable*	*W1ORatio*	*W2ORatio*
Main	MEAN		0.050	0.034
	STD		0.096	0.084
	N		146,409	146,409
	CORR	W1ORatio	1.000	0.590
	CORR	W2ORatio	0.590	1.000
Scrb	MEAN		0.046	0.035
	STD		0.091	0.087
	N		142,496	142,496
	CORR	W1ORatio	1.000	0.622
	CORR	W2ORatio	0.622	1.000

sections of the March 2005 SAT administration across the main and scrambled spirals. The mean omit ratios were very low and similar (3 percent for W2 and 5 percent for W1). These mean omit ratios were the lowest for writing. In addition, given the large difference in the item numbers between the two sections, the correlation of 0.59 was fairly substantial, indicating that examinees omitted more or less similarly on the two writing sections.

Research Question: How consistently did examinees omit the last six items on the two writing sections across the two test spirals?

Table 15 summarizes the means, standard deviations, and correlations of examinees' omit ratios on the last six items of the two writing sections of the March 2005 SAT administration across the main and scrambled spirals. Combining this information with that shown in Table 14, it can be inferred that on the average, half of the omits on the longer W1 occurred on the last six items (about 2.5 percent for the last six items versus about 5 percent mean sectional omit ratio for W1), while virtually all of the omits on the shorter W2 occurred on the last six items (about 3 percent on the last six items versus 3.4 percent on the entire section for W2).

The above findings on the consistency of response accuracy and omits between the main and scrambled test spirals were confirmed to be highly similar with the October 2005 and the May 2002 administrations.

Analysis 5: Examinee Performance Accuracy and Omit Tendency Throughout the Eight Operational Sections Across the Two Test Spirals Based on Summary Statistics Conditional on Total Scores

Most of the results shown in the two previous sections of this report were based on marginal analyses by computing means and standard deviations without any conditioning. These marginal means and standard deviations serve to demonstrate similarities or differences between groups and sections. This section of the report will describe, mostly through figures, the conditional mean right and omit ratios on examinees' ability estimates on math, reading, and writing sections. The purpose of these conditional analyses is to demonstrate more detailed similarities or differences in mean right and omit ratios along the entire examinee ability distributions.

Two explanatory comments about the conditional mean

Table 15

Means, Standard Deviations, and Correlations of Examinees' Omit Rates on the Last Six Items Across W1 and W2 and Two Test Spirals, March 2005 SAT Administration

Spiral	*Index*	*Variable*	*W1Last6OmitRatio*	*W2Last6OmitRatio*
Main	MEAN		0.024	0.030
	STD		0.049	0.074
	N		146,409	146,409
	CORR	W1Last6OmitRatio	1.000	0.504
	CORR	W2Last6OmitRatio	0.504	1.000
Scrb	MEAN		0.022	0.031
	STD		0.048	0.075
	N		142,496	142,496
	CORR	W1Last6OmitRatio	1.000	0.529
	CORR	W2Last6OmitRatio	0.529	1.000

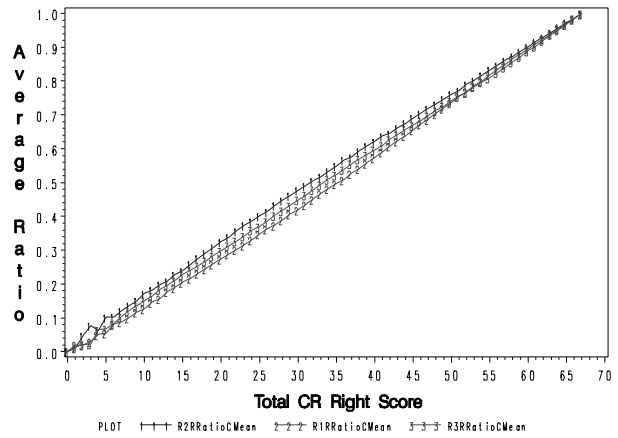

Figure 4. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores, main spiral, March 2005 SAT administration.

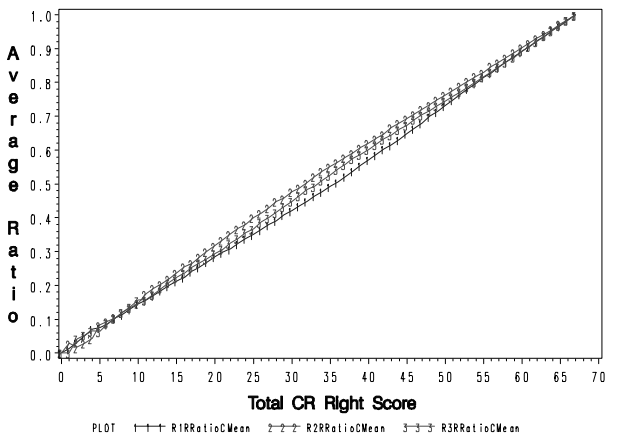

Figure 5. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores, scrambled spiral, March 2005 SAT administration.

ratio graphs are necessary. First, the mean right score ratios were computed using all the right score ratios for all examinees at particular total sectional right scores, as shown in all figures in this portion of this report. For example, the R1 mean right score ratios at the total reading score 40 in Figure 4 was computed on the right score ratios of 3,692 examinees whose total reading score was 40. Due to their size, all related numbers associated with the figures are omitted.

Second, the legends of "1," "2," and "3" for the distribution lines in each figure indicate the order in which their corresponding sections were presented to examinees. For example, section M2 was administered before section M1. On the figures, therefore, the mean right scores for section M2 are denoted by the line marked by 1s to indicate that this section was administered first.

The remainder of this report will address the conditional comparisons of examinee performance accuracy and omit tendency throughout the eight operational sections across the two test spirals.

Research Question: *What were the distributions of mean right ratios for the three reading sections conditional on examinee total right reading scores between the main and scrambled spirals?*

Figures 4 and 5 illustrate the distributions of mean right score ratios for the three SAT reading sections conditioned on examinees' total reading right scores between the main and scrambled spirals. ("RRatioCMean" stands for "conditional mean right ratio." When prefixed with a section name, such as "R1RRatioCMean," it stands for conditional mean right ratio for R1 section.) Four findings are clear. First, the conditional mean ratio distributions were virtually identical between the main and scrambled spirals. Second, examinees at 50 or higher total reading scores differed very little in their mean right score ratios across the three reading sections. This finding is expected since high-ability examinees scored high uniformly; however, the fact that the lines converge so early is notable. Third, the differences for examinees with a total right score of below 15 also gradually tapered off for the similar reason that low-ability examinees tend to score uniformly low. Fourth, the biggest difference of about 5 percent mean right score ratio between R1 and R2 occurred in the middle-ability range between 20 and 45, approximately. This finding was also anticipated since middle-ability examinees tend to vary more in their performance. Based on the above four findings, it can be concluded that the order and position of the three reading sections did not seem to have any significant effect on examinee performance along their entire ability range.

Research Question: *What were the distributions of mean omit ratios for the three reading sections conditional on examinee total right reading scores between the main and scrambled spirals?*

Figures 6 and 7 illustrate the distributions of mean omit ratios for the three SAT reading sections conditioned on examinees' total reading right scores between the main and scrambled spirals. ("ORatioCMean" stands for "conditional mean omit ratio." When prefixed with a section name, such as "R1ORatioCMean," it means conditional mean omit ratio for R1 section.) Four findings are clear. First, as expected, the higher the examinees' abilities were, the fewer items they tended to omit. As examinees' scores approached the highest possible scores, their mean omit ratios decreased to zero percent. Second, the opposite was true with lower ability examinees. The lower the examinees' abilities, the more items they tended to omit. Examinees (nine) at the 0 total reading score

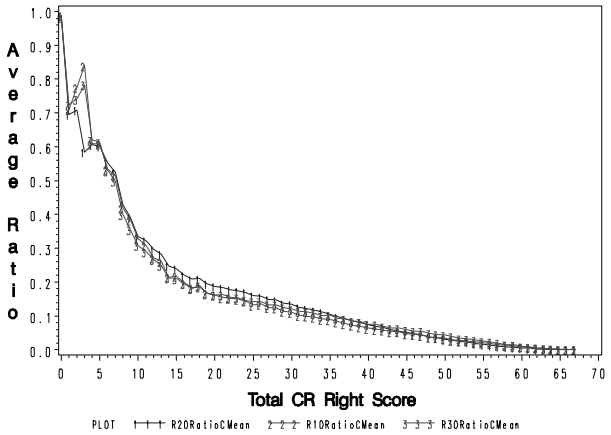

Figure 6. Distributions of mean omit ratios of R1, R2, and R3 conditional on total critical reading right scores, main spiral, March 2005 SAT administration.

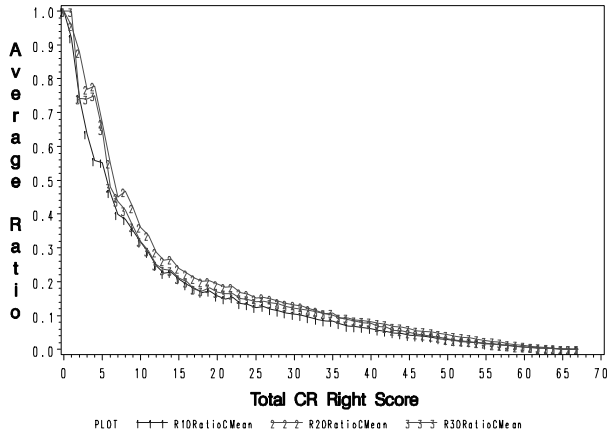

Figure 7. Distributions of mean omit ratios of R1, R2, and R3 conditional on total critical reading right scores, scrambled spiral, March 2005 SAT administration.

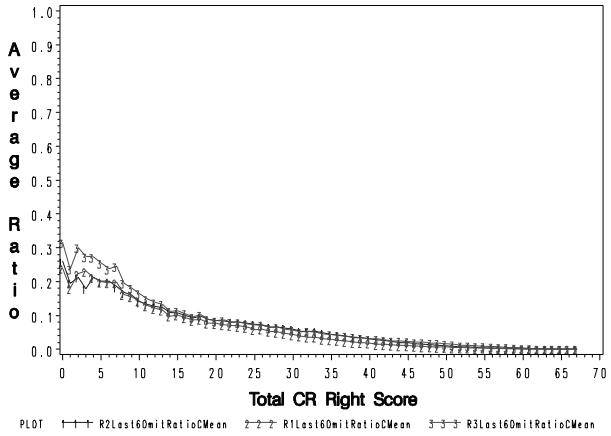

Figure 8. Distributions of mean omit ratios on the last six items of R1, R2, and R3 conditional on total critical reading right scores, main spiral, March 2005 SAT administration.

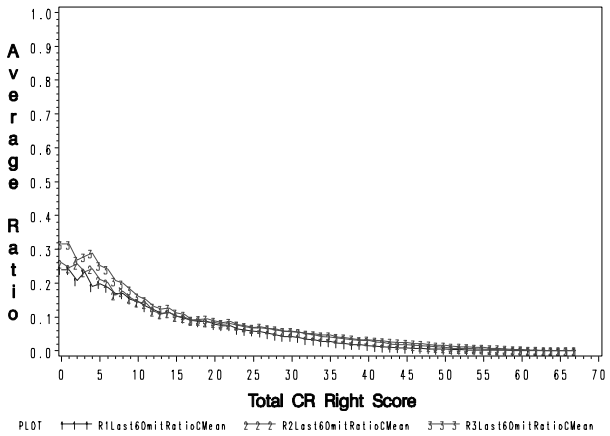

Figure 9. Distributions of mean omit ratios on the last six items of R1, R2, and R3 conditional on total critical reading right scores, scrambled spiral, March 2005 SAT administration.

point omitted 100 percent of the items across the three reading sections.[7] This finding also confirms the trend that examinees tended not to guess, since the SAT formula-scoring policy penalizes incorrect guessing. Third, the decreasing rate of omits appeared very steep between 0 to 10 total reading score range and became a more gradual descent from 15 score points on. Fourth, the mean omit ratio distributions between the main and scrambled test spirals virtually mirrored each other, with the exception of a few differences at the lower end of the total reading score range, caused mostly by the fluctuations of small sample sizes. The above findings confirm that the order and position of presenting the three reading sections did not have any significant effect on how examinees omitted along the entire reading ability range.

Research Question: *What were the distributions of mean omit ratios on the last six items for the three reading sections conditional on examinee total right reading scores between the main and scrambled spirals?*

Figures 8 and 9 show the distributions of mean omit ratios on the last six items for the three SAT reading sections conditioned on examinees' total reading right scores between the main and scrambled spirals. ("Last6OmitRatioCMean" stands for "conditional mean omit ratio on the last six items." When prefixed with a section name, such as "R1Last6OmitRatioCMean," it stands for conditional mean omit ratio on the last six items for R1.) Trends in these two figures were similar to those found in the previous two figures, except for the "spikes," or high percentages of omits at the extremely low-score ranges, which disappeared because only the last reading items were used for this analysis. Therefore, given the similarities of the conditional

[7]It is not clear why these examinees did not answer any item, and they were not excluded from this study for the sake of data entirety. Furthermore, the total number of such examinees was too small to impact the outcome of this study.

distributions, it can be concluded that the order and position of the three reading sections did not impact examinees' omit tendencies throughout the entire reading score range.

Research Question: *What were the distributions of mean right score ratios for the three math sections conditional on examinee total right math scores between the main and scrambled spirals?*

Figures 10 and 11 show the distributions of mean right score ratios for the three SAT math sections conditioned on examinees' total reading right scores between the main and scrambled spirals. Three findings can be seen. First, between total math scores of 5 and 47, examinees consistently correctly answered about 10 percent (or two more items) on M1 than on M2. As mentioned earlier, this difference could have been resulted from the difference in section difficulties but not by presentation order, since M2, the slightly more difficult math section, was presented before M1 in both test spirals. Second, M3, the third math section, appeared as easy as M1 for the lower one-third of the total score range and as hard as M2 for the upper one-third of the total math score range. Third, the two test spirals virtually mirrored each other in terms of the right score mean ratio distributions. It can be concluded that performance on the later math sections was not negatively impacted by location.

Research Question: *What were the distributions of mean omit ratios for the three math sections conditional on examinee total right math scores between the main and scrambled spirals?*

Figures 12 and 13 show the distributions of mean omit ratios for the three SAT math sections conditioned on examinees'

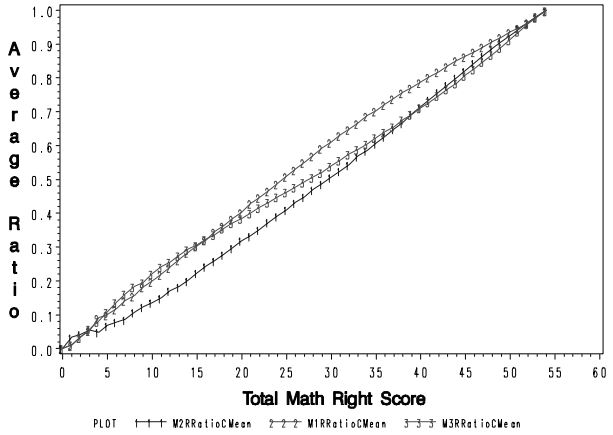

Figure 10. Distributions of mean right score ratios of M1, M2, and M3 conditional on total math right scores, main spiral, March 2005 SAT administration.

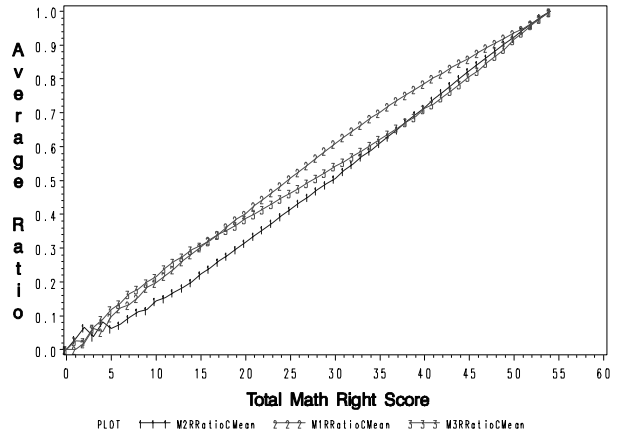

Figure 11. Distributions of mean right score ratios of M1, M2, and M3 conditional on total math right scores, scrambled spiral, March 2005 SAT administration.

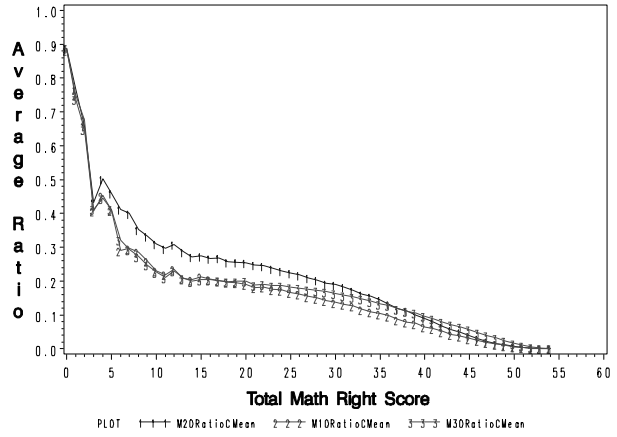

Figure 12. Distributions of mean omit ratios of M1, M2, and M3 conditional on total math right scores, main spiral, March 2005 SAT administration.

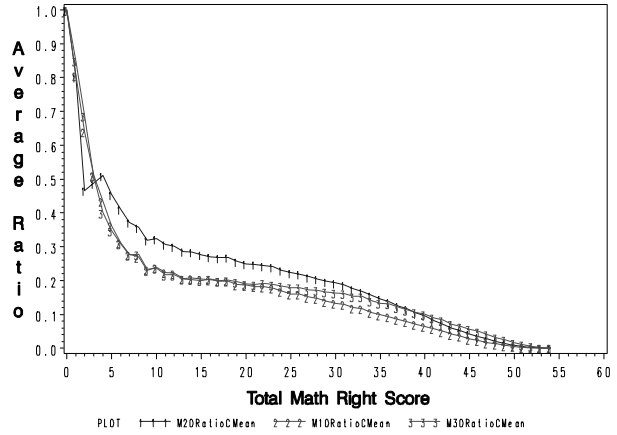

Figure 13. Distributions of mean omit ratios of M1, M2, and M3 conditional on total math right scores, scrambled spiral, March 2005 SAT administration.

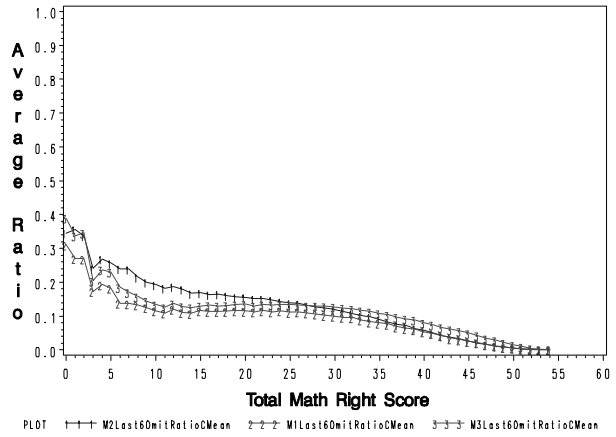

Figure 14. Distributions of mean omit ratios on the last six items of M1, M2, and M3 conditional on total math right scores, main spiral, March 2005 SAT administration.

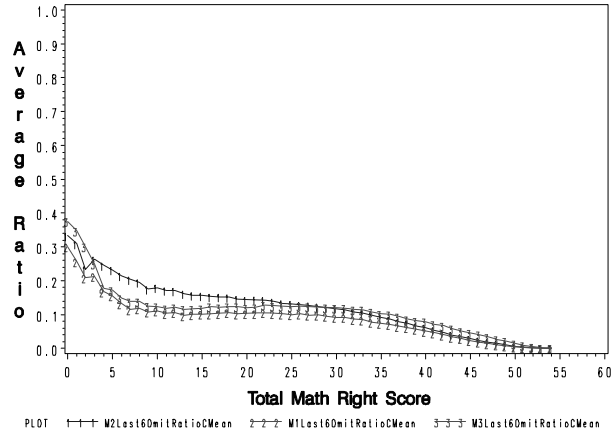

Figure 15. Distributions of mean omit ratios on the last six items of M1, M2, and M3 conditional on total math right scores, scrambled spiral, March 2005 SAT administration.

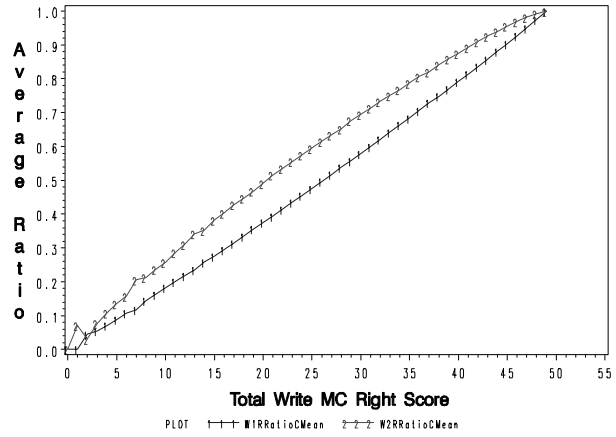

Figure 16. Distributions of mean right score ratios of W1 and W2 conditional on total MC writing right scores, main spiral, March 2005 SAT administration.

mean omit ratios between the main and scrambled test spirals, demonstrating the same trends as those of the two omit ratio figures for the three reading sections discussed earlier. The only unique thing here was that M2, the hardest section of the three and the section presented first, elicited higher omit rates for examinees below the total score of 35. It is likely that it was the difficulty of the section, not the position of presenting the three math sections, that caused the increase in the average omit ratios.

Research Question: *What were the distributions of mean omit ratios on the last six items for the three math sections conditional on examinee total right math scores between the main and scrambled spirals?*

Figures 14 and 15 show the distributions of mean omit ratios on the last six items for the three SAT math sections conditioned on examinees' mean omit ratios between the main and scrambled test spirals. These two figures demonstrated the same characteristics as their reading counterparts discussed earlier. Therefore, it can also be concluded that the way examinees omitted on the last six items did not seem to have been influenced by the position of the three math sections.

Research Question: *What were the distributions of mean right score ratios for the two writing sections conditional on examinee total writing multiple-choice scores between the main and scrambled spirals?*

Figures 16 and 17 present the distributions of mean right score ratios on the two SAT writing sections

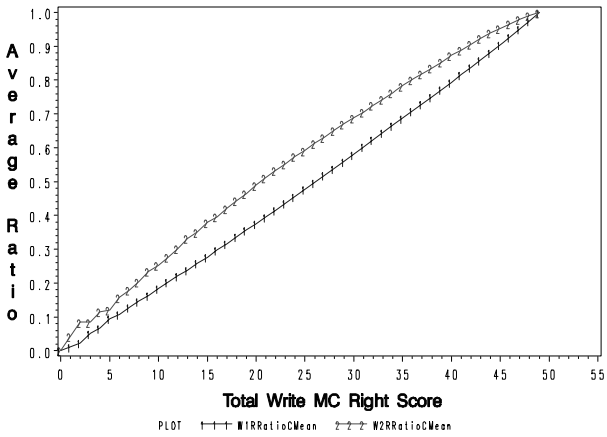

Figure 17. Distributions of mean right score ratios of W1 and W2 conditional on total MC writing right scores, scrambled spiral, March 2005 SAT administration.

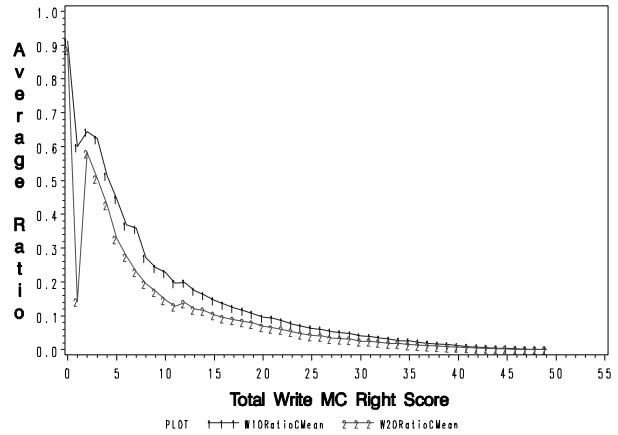

Figure 18. Distributions of mean omit ratios of W1 and W2 conditional on total MC writing right scores, main spiral, March 2005 SAT administration.

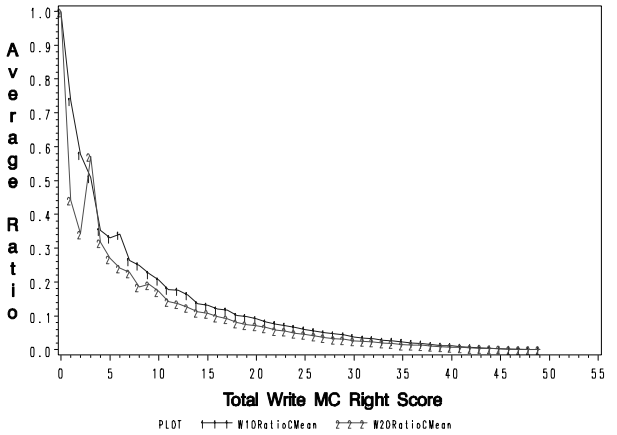

Figure 19. Distributions of mean omit ratios of W1 and W2 conditional on total MC writing right scores, scrambled spiral, March 2005 SAT administration.

conditioned on examinees' total MC writing scores between the main and scrambled test spirals. Recall two facts about W1 and W2: first, W1 was more than twice as long as W2, and second, they were very far away from each other in terms of position throughout the duration of the SAT administration. Given the first point, it should not be surprising to see that the mean right score ratios for W1 were consistently lower than those of W2, by as much as 10 percent, which supports the earlier marginal analyses. The same findings were true of both the main and scrambled spirals. Based on these findings, it can be concluded that there was not substantial evidence that examinees' performance on the second of the two writing sections was negatively influenced by the position.

Research Question: *What were the distributions of omit ratios for the two writing sections conditional on examinee total writing multiple-choice scores between the main and scrambled spirals?*

Figures 18 and 19 illustrate the distributions of mean omit ratios on the two SAT writing sections conditioned on examinees' total MC writing scores between the main and scrambled test spirals. Even though the differences of mean right scores were relatively large, as shown previously, the mean omit ratios were much smaller across the two spirals, except for a few in the score range below 10. Comparing the two figures, it can be seen that the relatively large gaps between W1 and W2 in the low-score range were largely caused by W2, and hence could have been attributed to the relatively small number of items on W2. These findings lend support to the conclusion that examinees' did not omit more on the last section of the test. That is, examinees did not seem to be negatively influenced by position of the sections.

Research Question: *What were the distributions of omit ratios on the last six items for the two writing sections conditional on examinee total writing multiple-choice scores between the main and scrambled spirals?*

Figures 20 and 21 show the distributions of mean omit ratios on the last six items on the two writing sections conditioned on examinees' total MC writing scores between the main and scrambled test spirals. Two findings stand out. First, the spikes along the lower writing ability range, caused by the small number of items on W2, were still obvious. Second, the curves in Figures 20 and 21

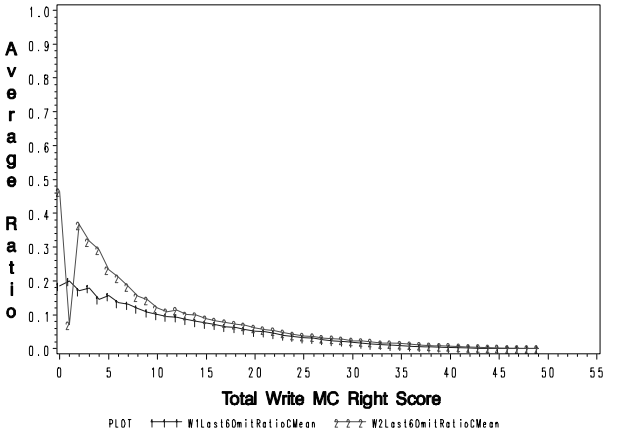

Figure 20. Distributions of mean omit ratios on last six items of W1 and W2 conditional on total MC writing right scores, main spiral, March 2005 SAT administration.

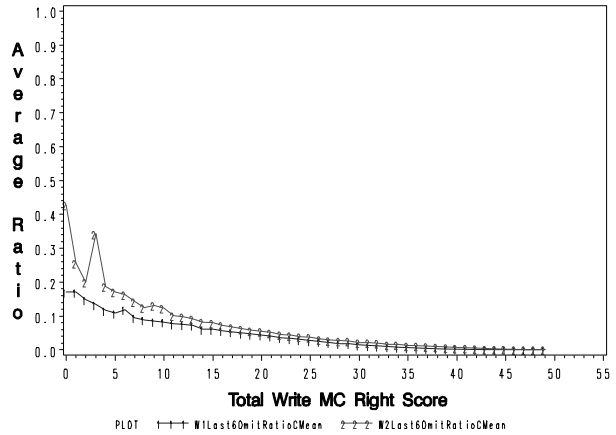

Figure 21. Distributions of mean omit ratios on last six items of W1 and W2 conditional on total MC writing right scores, scrambled spiral, March 2005 SAT administration.

converged after a score of 10, signifying that even though W1 and W2 were the furthest apart in the main spiral, examinees did not behave differently than when the sections were closer together. The second finding lends strong support for the conclusion that position did not have an effect on the majority of examinees.

The conditional analyses on examinee response accuracy and omits were cross-validated to be true with the October 2005 and the May 2002 SAT administrations.

Analysis 6: Examinee Performance Accuracy and Omit Tendency by Gender (Replication of Relevant Analyses)

Up to now, all the analyses have been based on all the examinees who took the March 2005 SAT administration. Will the relevant findings hold true for female and male examinees?

Historically, the mean scores of female examinees on the critical reading and math sections tended to be slightly lower than those of their male counterparts (College Board, 2005). For example, on the basis of the 1990 SAT College-Bound Seniors, the male and female CR means were 505 versus 496, while their math means were 521 versus 483, respectively. However, female examinees averaged higher than male examinees on writing; 507 versus 496. As a result of such systematic differences between the female and male examinees, only the conditional analyses are appropriate.[8]

Table 16

Distributions of Female and Male Examinees Between Main and Scrambled Spirals, March 2005 SAT Administration*

Spiral	*Gender*	*Frequency*	*Percent*	*Cumulative Percent*	*Cumulative Frequency*
Main	F	80,063	27.77	27.77	80,377
	M	66,032	22.90	50.67	146,409
Scrb	F	77,536	26.89	77.56	224,200
	M	64,705	22.44	100.00	288,905

*569 examinees did not report gender information.

Research Question: *Were the female and male examinees distributed equally across the two test spirals?*

Table 16 shows the distributions of female and male examinees between the main and scrambled spirals for the March 2005 SAT administration. It can be concluded that the percentages of female and male examinees were highly comparable across the two test spirals, constituting about 28 percent and 23 percent, respectively.

Research Question: *How differently did female and male examinees perform on the three reading, three math, and two writing sections conditional on their abilities?*

Figures 22 to 27 show the conditional distributions of the right score ratios of the female and male examinees on the three reading, three math, and two writing sections for the March 2005 SAT administration. The overall contours of all the distributions conditional on gender resembled closely those based on all examinees as previously shown and the female and male conditional distributions were virtually identical for each section of the test. These two findings show that female and male examinees performed identically, and their performance was not affected in any way by the order and position of the eight sections.

Research Question: *How differently did female and male examinees omit on the last six items on the three reading, three math, and two writing sections conditional on their abilities?*

Figures 28 to 33 depict the conditional distributions of the mean omit ratios on the last six items on the three

[8]Due to the fact that virtually all analyses based on gender yielded highly similar results as those based on the entire examinee population, this report will only present select conditional analyses and findings in a more condensed fashion. However, more detailed results are available upon request.

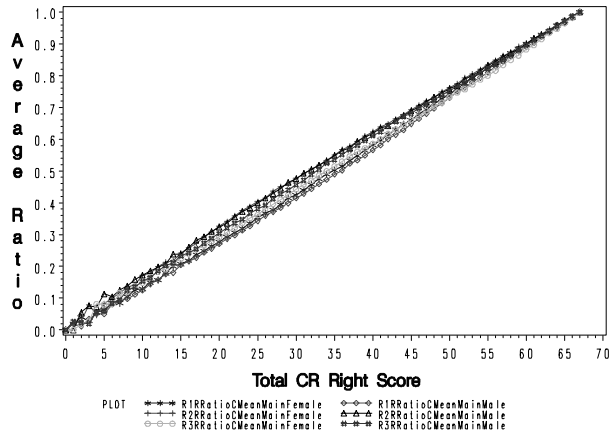

Figure 22. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by gender, main spiral, March 2005 SAT administration.

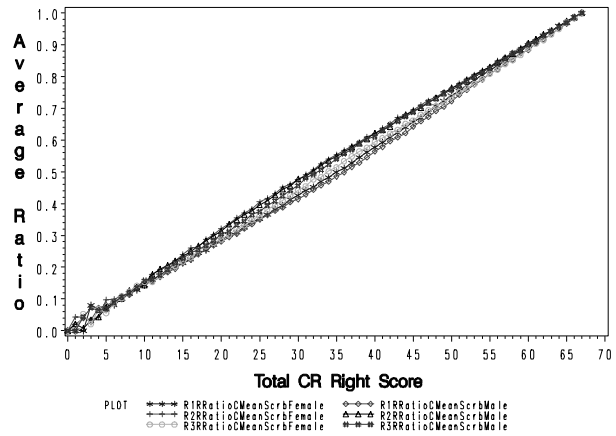

Figure 23. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by gender, scrambled spiral, March 2005 SAT administration.

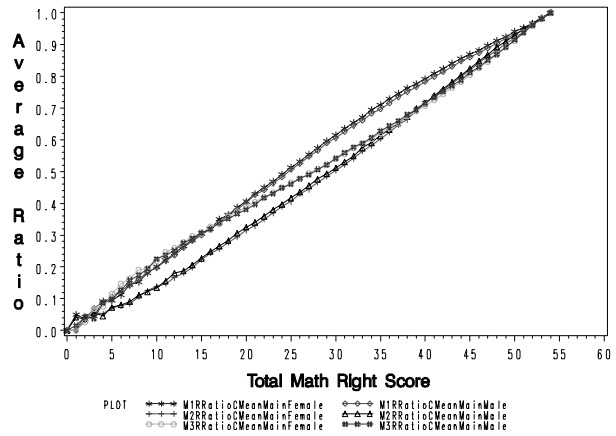

Figure 24. Distributions of mean right score ratios of M1, M2, and M3 conditional on total math right scores by gender, main spiral, March 2005 SAT administration.

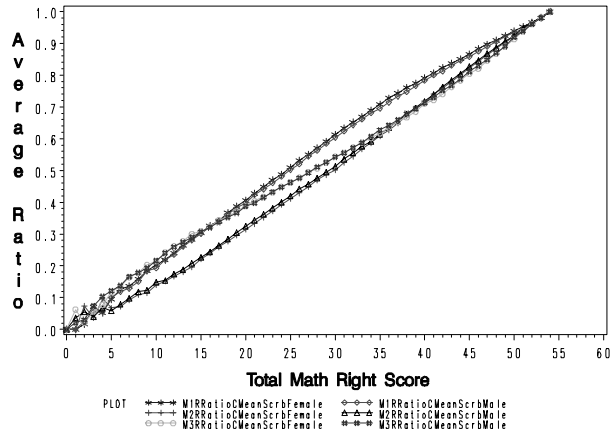

Figure 25. Distributions of mean right score ratios of M1, M2, and M3 conditional on total math right scores by gender, scrambled spiral, March 2005 SAT administration.

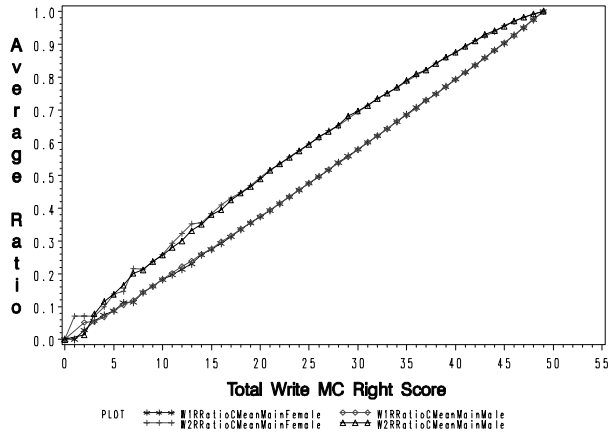

Figure 26. Distributions of mean right score ratios of W1 and W2 conditional on total MC writing right scores by gender, main spiral, March 2005 SAT administration.

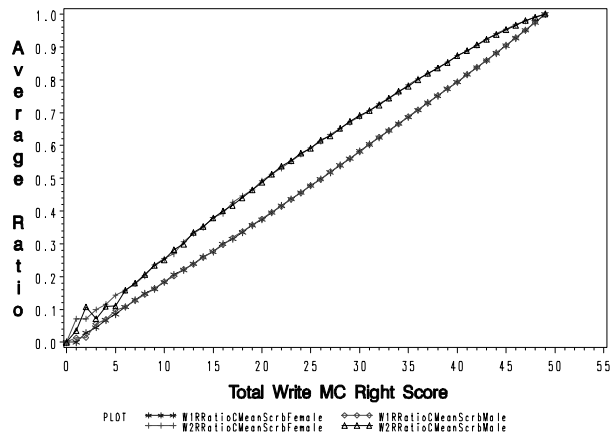

Figure 27. Distributions of mean right score ratios of W1 and W2 conditional on total MC writing right scores by gender, scrambled spiral, March 2005 SAT administration.

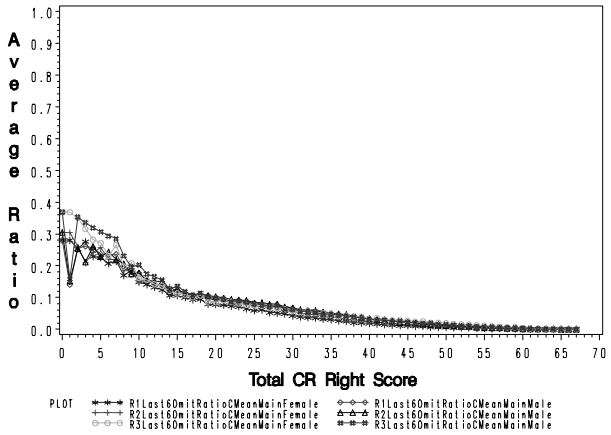

Figure 28. Distributions of mean omit ratios on the last six items of R1, R2, and R3 conditional on total critical reading right scores by gender, main spiral, March 2005 SAT administration.

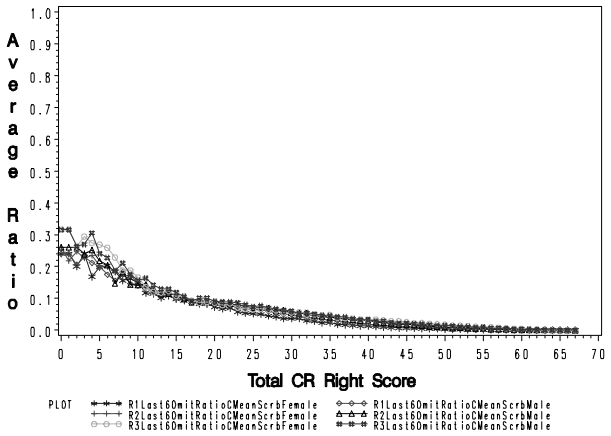

Figure 29. Distributions of mean omit ratios on the last six items of R1, R2, and R3 conditional on total critical reading right scores by gender, scrambled spiral, March 2005 SAT administration.

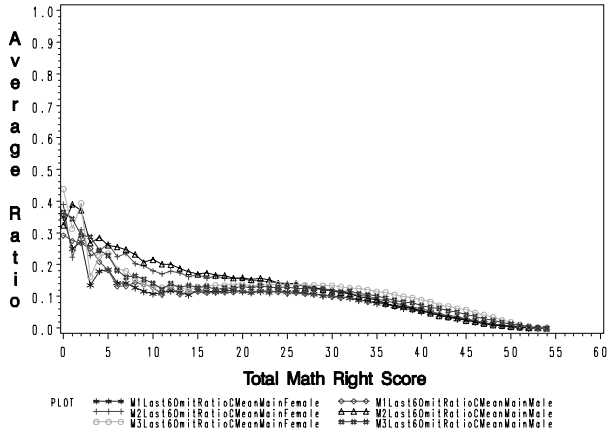

Figure 30. Distributions of mean omit ratios on the last six items of M1, M2, and M3 conditional on total math right scores by gender, main spiral, March 2005 SAT administration.

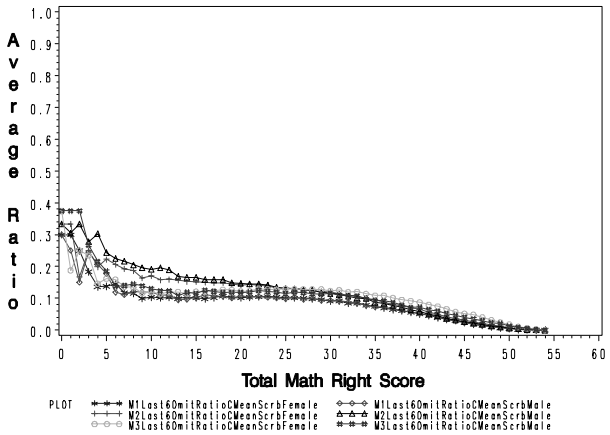

Figure 31. Distributions of mean omit ratios on the last six items of M1, M2, and M3 conditional on total math right scores by gender, scrambled spiral, March 2005 SAT administration.

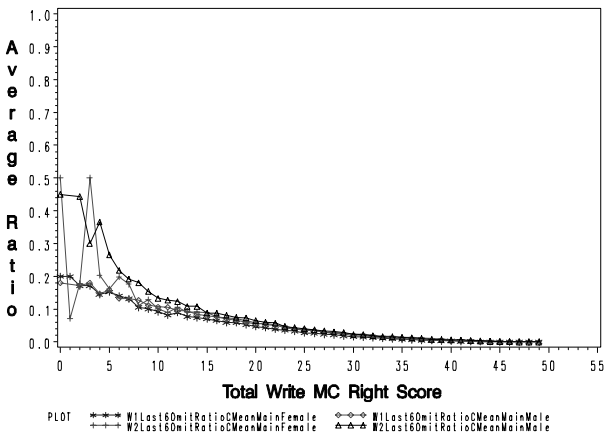

Figure 32. Distributions of mean omit ratios on the last six items of W1 and W2 conditional on total MC writing right scores by gender, main spiral, March 2005 SAT administration.

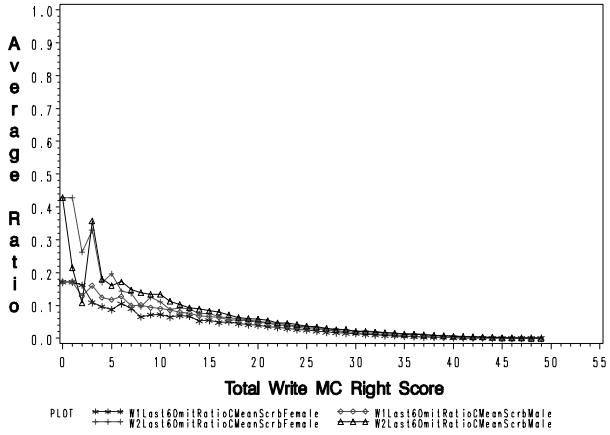

Figure 33. Distributions of mean omit ratios on the last six items of W1 and W2 conditional on total MC writing right scores by gender, scrambled spiral, March 2005 SAT administration.

reading, three math, and two writing sections between the female and male examinees for the March SAT 2005 administration. Examining these six figures reveals familiar findings. First, the overall contours of all the distributions conditional on gender resembled closely those based on all examinees, as shown previously. Second, the female and male conditional distributions were all virtually identical. These two findings show similar omit rates across genders and no differences due to section order or position.

All the above findings conditional on ability and gender were also true of the October 2005 and May 2002 administrations.

Analysis 7: Examinee Performance Accuracy and Omit Tendency by Racial/ Ethnic Group (Replication of Relevant Analyses)

The next analysis focuses on racial/ethnic groups. Currently, the SAT Questionnaire classifies examinees into eight racial/ethnic categories:

- American Indian (abbreviated as "AmInd" in later figures)
- Asian American (abbreviated as "AmAsian")
- Black
- Mexican American (abbreviated as "MexAm")
- Puerto Rican (abbreviated as "PRican")
- Latin American (abbreviated as "LatinAm")
- White
- Other

Like gender, historically, the mean scores of examinees of different racial/ethnic groups on SAT critical reading and math tend to be different (College Board, 1996–2005). As a result, only conditional analyses are applicable.[9]

Research Question: *Were examinees of different racial/ethnic groups distributed similarly across the two test spirals?*

Table 17 shows the distributions of examinees of different racial/ethnic groups between the main and scrambled spirals for the March 2005 SAT administration. It can be concluded that the percentages of female and male examinees were highly comparable across the two test spirals, differing only by one-tenth of a percent. For instance, as the smallest racial/ethnic group, American Indian examinees comprised about 0.3 percent in both the main and scrambled spirals.

Table 17

Distributions of Examinees of Eight Racial/Ethnic Groups Between Main and Scrambled Spirals,* March 2005 SAT Administration

Spiral	*Racial/ Ethnic Group*	*Frequency*	*Percent*	*Cumulative Percent*	*Cumulative Frequency*
Main	Am Ind	801	0.30	0.30	13,430
	Am Asian	17,177	6.51	6.81	30,607
	Black	11,219	4.25	11.06	41,826
	Mex Am	5,537	2.10	13.16	47,363
	P Rican	1,312	0.50	13.66	48,675
	Latin Am	4,645	1.76	15.42	53,320
	White	88,237	33.44	48.86	141,557
	Other	4,852	1.84	50.70	146,409
Scrb	Am Ind	774	0.29	50.99	159,583
	Am Asian	16,773	6.36	57.35	176,356
	Black	10,909	4.13	61.48	187,265
	Mex Am	5,297	2.01	63.49	192,562
	P Rican	1,347	0.51	64.00	193,909
	Latin Am	4,556	1.73	65.73	198,465
	White	85,763	32.50	98.23	284,228
	Other	4,677	1.77	100.00	288,905

*A total of 25,029 examinees did not report their racial/ethnic group.

Research Question: *How differently did examinees of different racial/ethnic groups perform on the three reading, three math, and two writing sections conditional on their abilities?*

Due to the large number of possible combinations based on eight operational sections, eight racial/ethnic groups, and the two test spirals, only four representative figures are reported. In addition, to avoid overcrowding of too many conditional distribution lines in a figure, only three racial/ethnic groups are included in each graph. To provide a benchmark for comparison across figures, however, the data for whites are repeated in all four figures.

Figure 34 shows the conditional mean right ratio distributions for R1, R2, and R3 under the main test spiral for the Asian American, black, and white groups. As indicated in Table 17, these three groups were the three largest SAT examinee groups and should produce the smoothest and most stable distribution lines among all the

[9]Again, due to the fact that virtually all analyses based on racial/ethnic group yielded highly similar results as those based on the entire examinee population, this report will only present select conditional analyses and findings for brevity. More detailed results are available upon request.

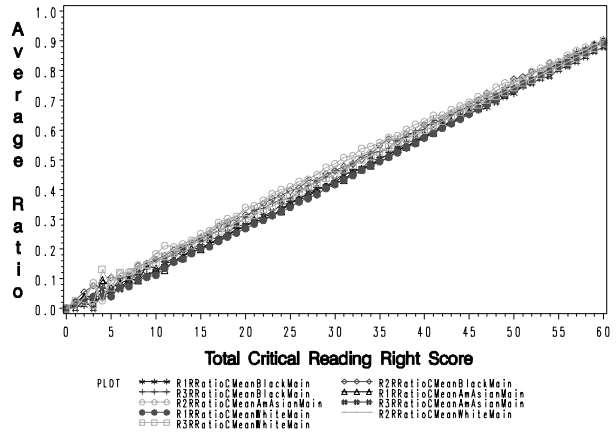

Figure 34. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three racial/ethnic groups: Asian American, black, and white, main spiral, March 2005 SAT administration.

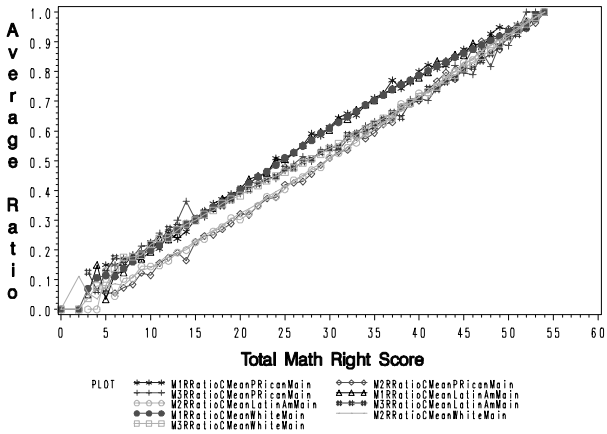

Figure 36. Distributions of mean right score ratios of M1, M2, and M3 conditional on total math right scores by three racial/ethnic groups: Latin American, Puerto Rican, and white, main spiral, March 2005 SAT administration.

racial/ethnic groups. The contours are highly similar to those exhibited in Figure 4, and distributions are similar, indicating that these ethnic groups performed more or less the same across the three critical reading sections.

Figure 35 shows the conditional mean right ratio distributions for R1, R2, and R3 under the scrambled test spiral for the American Indian, Mexican American, and white groups. As shown in Table 17, American Indian tends to be the smallest racial/ethnic group, and its distribution lines should therefore be the least stable. However, despite relatively small examinee sizes, the American Indian conditional right score ratio distributions for R1, R2, and R3 under the scrambled test spiral were consistent with those of the Mexican American and white groups. It can be concluded that the performance trends of the American Indian group were not significantly affected by the different order and position of the three reading sections.

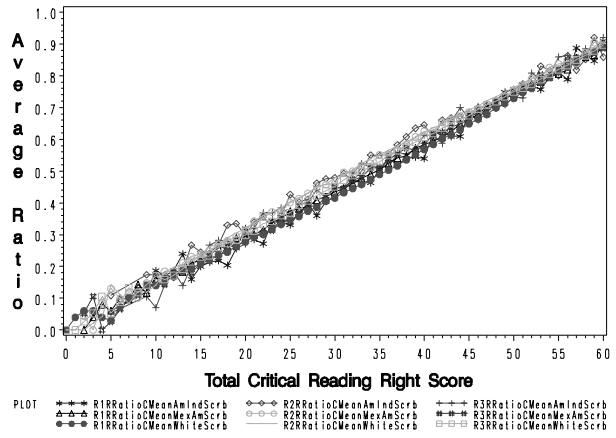

Figure 35. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three racial/ethnic groups: American Indian, Mexican American, and white, scrambled spiral, March 2005 SAT administration.

Figure 36 shows the conditional mean right ratio distributions for M1, M2, and M3 under the main test spiral for the Latin American, Puerto Rican, and white racial/ethnic groups. As shown in Table 17, the Latin American and Puerto Rican groups also tend to be very small. Two findings are evident. First, the general contours in Figure 36 were highly similar to those of Figure 24 based on the total regular SAT group. Second, all the distributions of the three racial/ethnic groups were virtually identical. These two findings support the conclusion that none of the three racial/ethnic groups' performance was affected by the different location in which they encountered the three math sections.

Figure 37 shows the conditional mean right ratio distributions for W1 and W2 for Latin American, Puerto Rican, and white groups. Again, its contour was highly similar to that of Figure 26, except for a few fluctuations primarily due to small sample sizes at lower total scores. It can be concluded that all three groups performed virtually identically on the two writing sections.

The above findings of virtually identical performance were also true of all the other SAT sections. Additionally, omit patterns for examinees of all the eight racial/ethnic groups were extremely similar. These results, therefore, have been omitted here.

All the findings conditional on ability and racial/ethnic group remained the same with the October 2005 and the May 2002 SAT administrations.

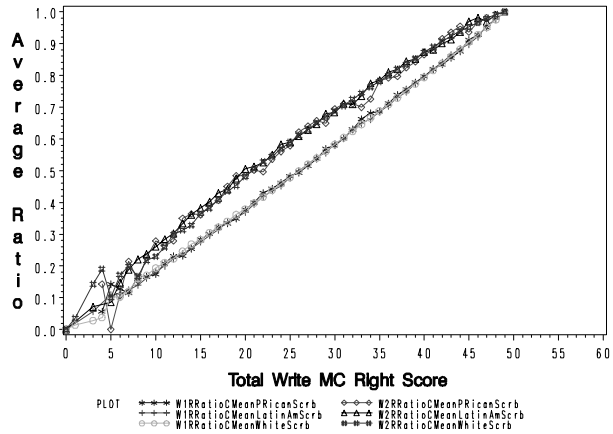

Figure 37. Distributions of mean right score ratios of W1 and W2 conditional on total MC writing right scores by three racial/ethnic groups: Latin American, Puerto Rican, and white, scrambled spiral, March 2005 SAT administration.

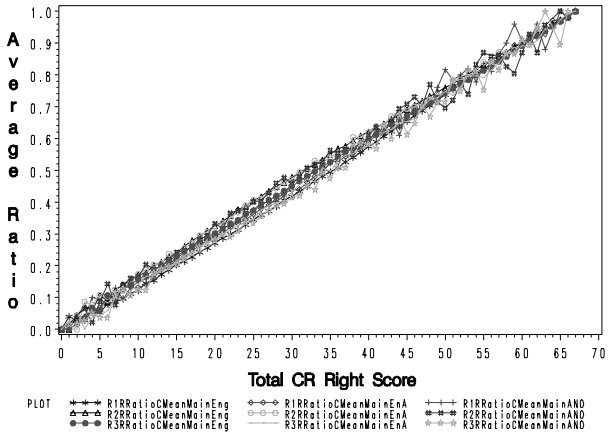

Figure 38. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three best-language groups, main spiral, March 2005 SAT administration.

Analysis 8: Examinee Performance Accuracy and Omit Tendency by Different Language Groups

How would the above findings hold for examinees of different languages? The College Board Student Questionnaire asks examinees about their best languages, classifying them into three groups:

- Those whose best language is English (abbreviated as "Eng");
- Those whose best language is English and who speak another language equally well (abbreviated as "EnA"); and
- Those whose best language is not English (abbreviated as "ANO").

Table 18

Distributions of Examinees Based on Their Best-Language Indications* Between Main and Scrambled Spirals, March 2005 SAT Administration

Spiral	*Best Language*	*Frequency*	*Percent*	*Cumulative Percent*	*Cumulative Frequency*
Main	Eng	109,347	46.78	46.78	109,347
	EnA	7,326	3.13	49.91	116,673
	ANO	1,815	0.78	50.69	118,488
Scrb	Eng	106,354	45.50	96.19	224,842
	EnA	7,097	3.04	99.23	231,939
	ANO	1,808	0.77	100.00	233,747

*12,653 examinees with unreported best language information and another 42,505 examinees that could not be properly matched were excluded from this analysis.

Three findings can be seen from Table 18. First, these three best-language groups were approximately equally distributed between the main and spiraled forms. Second, examinees whose best language was both English and another language constituted about 3 percent of the examinee population. Third, examinees whose best language was another language made up less than 1 percent of the examinee population.

Historically, the mean scores of examinees of language backgrounds on SAT critical reading and math tend to be different (College Board, 1996–2005). As a result, only conditional analyses would be applicable.[10]

All the findings conditional on ability and best-language background were consistently replicated with the October 2005 and the May 2002 SAT administrations.

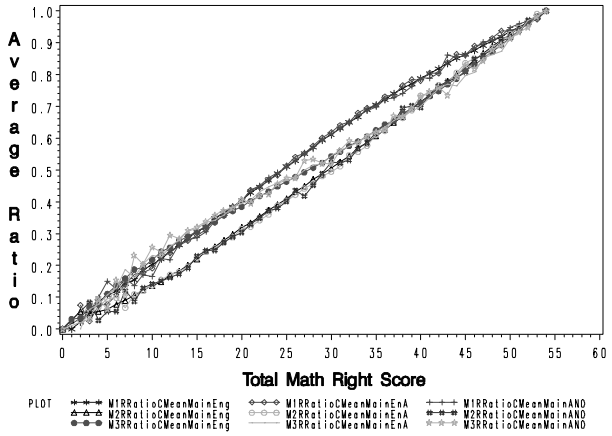

Figure 39. Distributions of mean right score ratios of M1, M2, and M3 conditional on total math right scores by three best-language groups, main spiral, March 2005 SAT administration.

[10]Again, due to the fact that virtually all analyses based on racial/ethnic group yielded highly similar results as those based on the entire examinee population, this report will only present select conditional analyses and findings for brevity. More detailed results are available upon request.

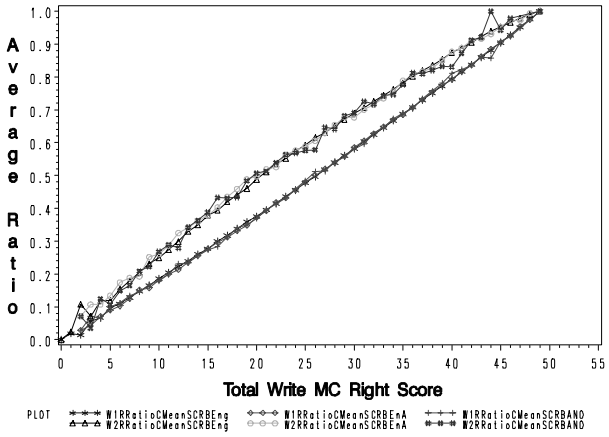

Figure 40. Distributions of mean right score ratios of W1 and W2 conditional on total MC writing right scores by three best-language groups, scrambled spiral, March 2005 SAT administration.

Conclusion

Using the data from the March 2005, October 2005, and May 2002 SAT administrations, this paper has systematically investigated the possible effects of the extended test time of the SAT Reasoning Test by studying examinee performance accuracy and omit tendency on operational sections presented to examinees in different orders and at different positions. Based on the large number of analyses and findings in this study, it can be concluded that, at a group level, the current length of the SAT Reasoning Test does not significantly impact examinee performance at the national level across different gender and racial/ethnic groups. On the contrary, examinee performance, regardless of gender and racial/ethnic group, was shown to be overwhelmingly parallel throughout the entire test. Furthermore, there were strong negating findings concerning decreased accuracy or increased omitting for sections that appeared later in the test. That is, the increased test length did not appear to have any negative effect on the performance of students on the later portion of the test at the population and subgroup levels. Furthermore, except for those on the writing section, all findings from the SAT Reasoning Test were completely replicated on the SAT I: Reasoning Test, indicating no significant changes in examinee performance trends.

The consistent findings in this research can be explained from both theoretical and practical perspectives. First, given the high-stakes nature of the SAT Reasoning Test, examinees are typically motivated to overcome the extra fatigue, if it exists. Research has shown that high-stakes testing and tasks tend to result in increased effort (Sarason, 1959; Weiner, 1986), enhanced self-monitoring, and regulation of energy for goal achievement (Kanfer and Ackerman, 1989; Hockey and Earle, 2006). Many motivational factors such as personal goals, perceptions of high levels of cognitive demands, or even unexpected difficulties in testing can influence examinee performance.

Second, although the SAT Reasoning Test is highly challenging, its cognitive demand or strain is substantially reduced by the alternating order in which different sections of different content are administered. Research by Soloman (1948), Hull (1943) and Zeaman and Kaufman (1955) has shown that prolonged homogeneous tasks tend to decrease performance, while changing task contents can cause improvement. In addition, students receive two 5-minute and one 1-minute breaks.

There is no doubt that fatigue is highly individualistic (Pearson, 1957) and that different examinees may experience different amounts of fatigue while working on the same task. In addition to examinee ability, the amount of fatigue may depend on many personal traits such as propensity of perseverance, motivation, and temperament. The main limitation of this study is that it is unable to systematically analyze what kinds of personal traits can significantly account for individual examinee performance differences.

There are at least two directions for future research. First, more analyses can be carried out on the distribution of individual differences in addition to the average performance levels. Second, and most important, new studies should be conducted to establish more validity evidence regarding the SAT Reasoning Test, especially on the newly added writing component.

References

Ackerman, P.L. (1988). Determinants of individual differences during skill acquisition: cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117*, 288–318.

Ackerman, P.L., & Kanfer, R. (2006). *Test length and cognitive fatigue: Phase 1. Survey of the theoretical and empirical literature*. An unpublished report submitted to the College Board.

Allspach, J., Feigenbaum, M., & Liu, J. (2003). *A survey of fatigue effects from the new SAT*. Princeton, NJ: Educational Testing Service.

Bartley, S.H., & Chute, E. (1947). *Fatigue and impairment in man*. New York: McGraw-Hill Book Company.

Black Issues in Higher Education. (2005). *Survey: length of new SAT is biggest complaint among first-round of test-takers*. Retrieved October 26, 2006, from http://www.findarticles.com/p/articles/mi_m0DXK/is_4_22/ai_n13720126.

Bridgeman, B., Cahalan, C., & Cline, F. (2003). *Time requirements for the different item types proposed for use in the revised SAT*. Report submitted to the College Board.

Burton, N., & Kostin, I. (2002). *Fatigue in test-taking: A brief report and literature review prepared for the College Board*. Internal Confidential Report: ETS.

Cameron, C. (1973). A theory of fatigue. *Ergonomics*, *16*, 633–648.

Cloud, J. (2003). *Inside the new SAT*. Time Archive. Retrieved October 26, 2006, from http://www.time.com/time/archive/preview/0,10987,1101031027-524393,00.html.

College Board. (2004). *Writing: A ticket to work…or a ticket out. A survey of business leaders*. New York: The College Board.

College Board. (2005). *2005 College-Bound Seniors: Total Group Profile Report*. Retrieved October 26, 2006, from http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2005/2005-college-bound-seniors.pdf.

College Board. (1996–2005). *National Reports*. Retrieved October 26, 2006, from http://www.collegeboard.com/about/news_info/cbsenior/yr2005/reports.html.

Crone, C., Wright, D., & Baron, P. (1993). *Performance of examinees for whom English is their second language on the spring 1992 SAT II: Writing Test*. Unpublished manuscript. Princeton, NJ: Educational Testing Service.

Grandjean, E. (1970). Fatigue: Yant Memorial Lecture—1970. A*merican Industrial Hygiene Association Journal*, *31*, 402–11.

Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory*. Boston: Kluwer-Nijhoff Publishing.

Hockey, G.R.J., & Earle, F. (2006). Control over the scheduling of simulated office work reduces the impact of workload on mental fatigue and task performance. *Journal of Experimental Psychology: Applied*, *12*, 50–65.

Hull, C.L. (1943). *Principles of behavior*. New York: Appleton-Century.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kanfer, R., & Ackerman, P.L. (1989). Motivation and cognitive abilities: an integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology—Monograph*, *74*, 657–90.

Liu, J., & Feigenbaum, M. (2004, April). *Data collection design and sample characteristics of the spring 2003 field trial for the new SAT and new PSAT/NMSQT*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Diego, CA.

Liu, J., Feigenbaum, M., & Dorans, N. (2004, April). *Equatability of the new SAT to the current SAT I*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Diego, CA.

Livingston, S.A. (1987). *The effects of time limits on the quality of student-written essays*. Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY.

Mathews, J. (2006). Scores on new SAT expected to decline. *Washington Post*. Retrieved October 26, 2006, from http://www.washingtonpost.com/wp-dyn/content/article/2006/05/10/AR2006051001435.html.

Muscio, B. (1921). Is a fatigue test possible? *British Journal of Psychology*, *12*, 31–46.

Myers, C.S. (1937). Conceptions of mental fatigue. *American Journal of Psychology*, *50*, 296–306.

Norman, D.A., & Bobrow, D.B. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *7*, 44–64.

Okogbaa, O.G. (1983). *An empirical model for mental work output and fatigue*. Unpublished doctoral dissertation, University of Cincinnati, OH.

Pearson, R.G. (1957). Scale analysis of a fatigue checklist. *Journal of Applied Psychology*, *41*(3), 186–91.

Reshestar, R., Liu, J., & Feigenbaum, M. (2004, April). *New SAT field trial prototype analysis*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Diego, CA.

Sarason, S.B. (1959). Test anxiety. *Journal of the National Education Association*, *48*(8), 26–27.

Schmidtke, H. (1976). Disturbance of processing of information. In E. Simonson & P.C. Weiser (Eds.), *Psychological aspects and physiological correlates of work and fatigue* (pp. 336–405). Springfield, IL: Charles C. Thomas.

Soloman, R.L. (1948). The influence of work on behavior. *Psychological Bulletin*, *45*, 1–40.

Thorndike, L.R. (1912). The curve of work. *Psychological Review*, *19*, 165–94.

Van der Linden, D., Frese, M., & Meijman, T.F. (2003). Mental fatigue and the control of cognitive processes: Effects on preservation and planning. *Acta Psychologica*, *113*, 45–65.

Walker, M., & Liu, J. (2004, April). *Scaling issues associated with the new SAT writing test*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Diego, CA.

Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag.

Zeaman, D., & Kaufman, H. (1955). Individual differences and theory in a motor learning task. *Psychological Monographs: General and Applied*, *69*(6, Whole No. 391).

Appendix A: Replication on the October 2005 Administration of the SAT Reasoning Test

In order to rule out the possibility that the previous findings from the March 2005 administration of the SAT Reasoning Test were specific only to this particular administration, the analyses reported previously were systematically replicated for the October 2005 administration.[11] Due to space limitations, only select results are presented here with full results available upon request.

Examinees and Test Spirals of the October 2005 SAT Administration

Based on the SC08 data feed, a total of 418,553 examinees participated in the October 2005 SAT administration. Of the 418,553 examinees, 410,338 tested on six unique test spirals on Saturday, and 409,040 of the 410,338 examinees were concentrated on two unique spirals. For stability, this replication study was carried out using the 409,040 examinees on the two spirals. To follow the convention used for the March 2005 study, these two spirals were referred to as main and scrambled ("Main" and "Scrb," respectively).

Table A1.1

Average Section Difficulty, October 2005 SAT Administration

Section	# of Items	Percent Correct*		Weight of One Item
		Mean	Std	
M1	20	0.627	0.226	0.050
M2	18	0.582	0.202	0.056
M3	16	0.616	0.209	0.063
R1	24	0.606	0.237	0.042
R2	24	0.596	0.193	0.042
R3	19	0.606	0.191	0.053
W1	35	0.637	0.235	0.029
W2	14	0.687	0.208	0.071

*Operational item difficulty statistics were used instead of pretest counterparts, due to the fact that the latter ones no longer existed.

Difficulty Levels of the Eight Operational Sections of the October 2005 SAT Administration

Table A1.1 shows the average percentages of the examinees who correctly answered the items of the eight operational sections of the October 2005 SAT administration. M1 and M3 were very close in difficulty, differing only by hundredths of a percent, while M2 was slightly harder than M1 and M3. The three reading sections had virtually identical difficulty levels, as did the two writing sections.

Analyses and Results

Research Question: *Did examinees of the October 2005 SAT administration perform highly similarly between the two spirals across the eight operational sections to their counterparts on the March 2005 administration?*

Table A1.2 summarizes the means and standard deviations for the total right scores on the three critical reading (CR), three math, and two writing (WR) sections between the two spirals on the October 2005 SAT administration. Concurring with their counterparts for the March 2005 administration, the means and standard deviations of the total right scores for the eight operational sections of the October 2005 administration were all virtually identical across the two spirals, differing only by tenths of a percent.

Figures A1.1 to A1.3 show that the percentages of examinees conditional on total right scores of critical reading, math, and writing sections were also virtually identical. It can be concluded that as in the March 2005 administration, the ability levels of the October 2005 examinees who took the main and scrambled test spirals were virtually identical. These findings confirmed the two premises for this replication. First, the examinees who took the two spirals were equivalent in their abilities. Second, although the two spirals differed in the order of presenting the three critical reading, three math, and two writing sections, such a difference in order did not

Table A1.2

Descriptive Statistics on Critical Reading, Math, and Writing Sections Across Main and Scrambled Spirals, October 2005 SAT Administration

Spiral	Frequency	CR Mean	CR Std	Math Mean	Math Std	WRMC Mean	WRMC Std
Main	207,011	38.49	12.42	31.78	10.69	30.52	8.45
Scrb	202,029	38.80	12.25	31.69	10.65	30.42	8.45

[11]The selection of October 2005 SAT administration data was made based on the availability of data, and the test structure that was amenable for the purpose of this study. More replications will be conducted on future administrations.

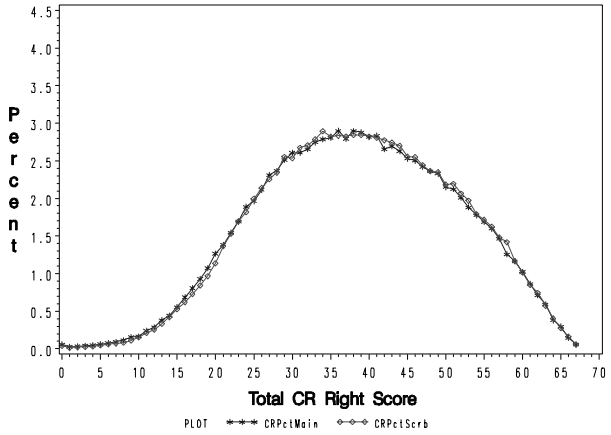

Figure A1.1. Total critical reading score distribution, October 2005 SAT administration.

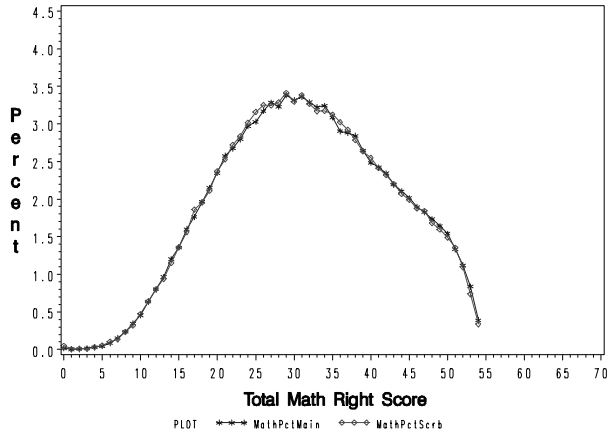

Figure A1.2. Total math score distribution, October 2005 SAT administration.

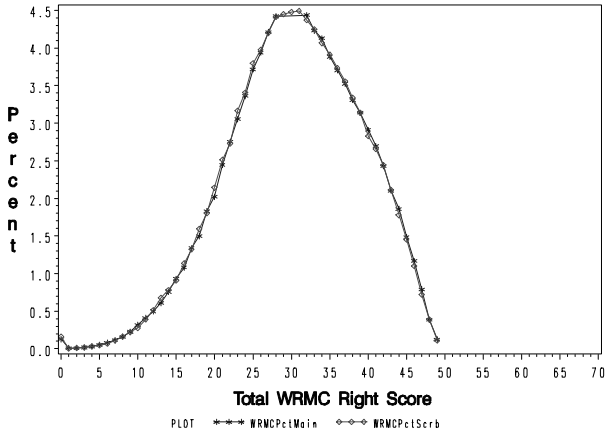

Figure A1.3. Total MC writing score distribution, October 2005 SAT administration.

seem to have exerted any significant effect on the examinees' three total section scores.

Research Questions: *Did the examinees of the October 2005 SAT administration perform similarly on sections of similar content within the two spirals to their March 2005 counterparts? Did they also perform similarly on sections of similar content but at different positions between the two test spirals?*

Table A1.3 shows virtually identical averages for the six indices for all the sections of similar contents both within and between the two test spirals, thus confirming the findings from the March 2005 SAT administration. Therefore, it can be concluded that, as with March 2005 data, the sectional performance of the October 2005 examinees did not seem to have been affected by the order of presentation of test sections.

Table A1.3

Summary of Four Mean Ratios Across Eight Sections and Two Spirals, October 2005 SAT Administration

Section Order			*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Spiral	Frequency	*Ratio*	M1	R2	W1	M2	R1	M3	R3	W2
Main	207,011	Right	0.607	0.565	0.604	0.562	0.582	0.595	0.577	0.671
		Wrong	0.264	0.329	0.323	0.278	0.326	0.240	0.313	0.299
		Omit	0.129	0.107	0.061	0.161	0.091	0.165	0.110	0.030
		Last Six Omit	0.064	0.063	0.026	0.098	0.053	0.099	0.056	0.023
Spiral	Frequency	Ratio	R1	M2	R2	M1	W1	R3	M3	W2
Scrb	202,029	Right	0.593	0.558	0.565	0.610	0.602	0.579	0.591	0.668
		Wrong	0.320	0.279	0.332	0.252	0.324	0.308	0.242	0.300
		Omit	0.087	0.164	0.103	0.139	0.062	0.113	0.167	0.032
		Last Six Omit	0.049	0.105	0.058	0.071	0.027	0.057	0.100	0.023

Research Question: *Were the correlations among the rates of their correct responses similar across the main and scrambled spirals?*

Table A1.4 shows that the correlations among the rates of math correct responses across the main and scrambled spirals differed only by hundredths of a percent. Such highly similar correlations were also found for the other performance indices and they are omitted here. All these findings confirmed those from the March 2005 SAT administration.

Research Question: *Were the October 2005 SAT administration distributions of mean right ratios for the three reading sections—conditional on examinee total right reading scores—highly similar between the main and scrambled spirals to their March 2005 counterparts?*

Figures A1.4 and A1.5 show the virtually identical conditional distributions of the average right ratios of the three critical reading sections between the main and scrambled spirals. Note that the crossing patterns in these two figures are highly similar. For example, R1, as the second CR section in the main spiral, had the highest average right ratios for most of the middle-ability examinees. The same was true for R1 in the scrambled spiral in which it was presented as the first CR section. Examinee performance on the three critical reading sections did not seem to be affected by the order in which the sections were presented, either within the same or across different test spirals. Such findings concur with those from the March 2005 SAT administration.

Table A1.4

Means, Standard Deviations, and Correlations of Examinee Right Response Rates Across R1, R2, and R3 and Two Test Spirals, October 2005 SAT Administration

Spiral	*Index*	*Variable*	*M1RRatio*	*M2RRatio*	*M3RRatio*
Main	MEAN		0.607	0.562	0.595
	STD		0.201	0.223	0.216
	N		207,011	207,011	207,011
	CORR	M1RRatio	1.000	0.814	0.779
	CORR	M2RRatio	0.814	1.000	0.798
	CORR	M3RRatio	0.779	0.798	1.000
Scrb	MEAN		0.610	0.558	0.591
	STD		0.201	0.220	0.217
	N		202,029	202,029	202,029
	CORR	M1RRatio	1.000	0.810	0.784
	CORR	M2RRatio	0.810	1.000	0.789
	CORR	M3RRatio	0.784	0.789	1.000

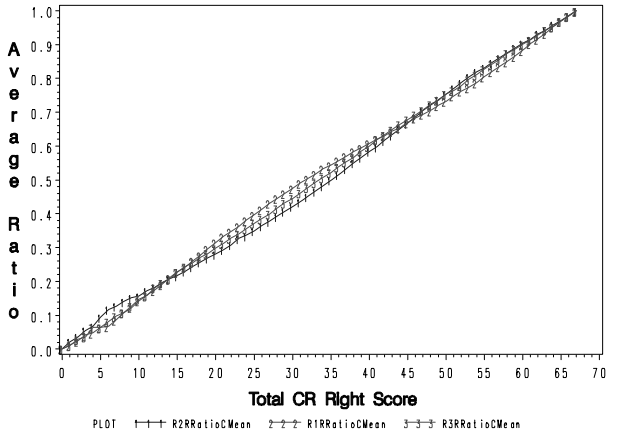

Figure A1.4. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores, main spiral, October 2005 SAT administration.

Research Question: *Were the October 2005 SAT administration distributions of mean omit ratios on the last six items for the three math sections—conditional on examinee total right math scores between the main and scrambled spirals—similar to their March 2005 counterparts?*

Figures A1.6 and A1.7 show that the October 2005 SAT administration conditional distributions of mean omit ratios on the last six items for the three math sections were virtually identical between the main and scrambled spirals. These findings were highly similar to those from their March 2005 counterparts. The conditional distributions for the other indices were also similar and omitted here.

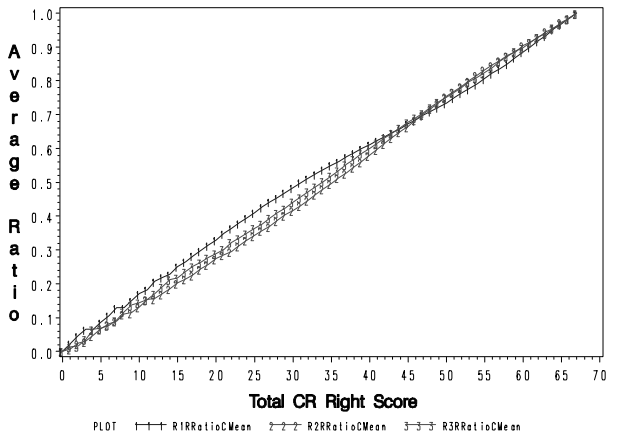

Figure A1.5. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores, scrambled spiral, October 2005 SAT administration.

Table A1.5

Distributions of Female and Male Examinees Between Main and Scrambled Spirals, October 2005 SAT Administration*

Spiral	*Gender*	*Frequency*	*Percent*	*Cumulative Percent*	*Cumulative Frequency*
Main	F	115,492	28.44	28.44	116,936
Main	M	90,075	22.18	50.62	207,011
Scrb	F	112,483	27.70	78.32	320,975
Scrb	M	88,065	21.68	100.00	409,040

*2,925 examinees had missing gender information, and were thus omitted from this analysis.

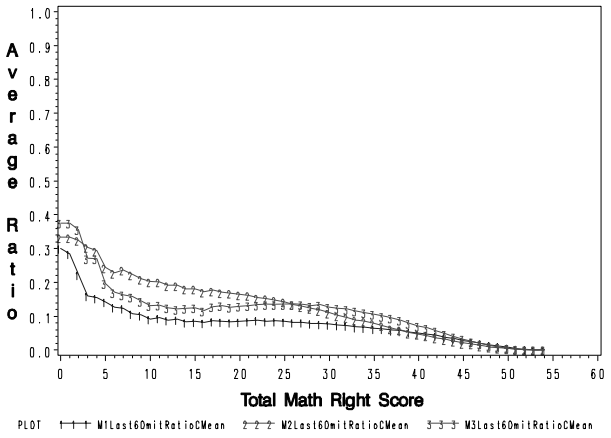

Figure A1.6. Distributions of mean omit ratios on the last six items of M1, M2, and M3 conditional on total math right scores, main spiral, October 2005 SAT administration.

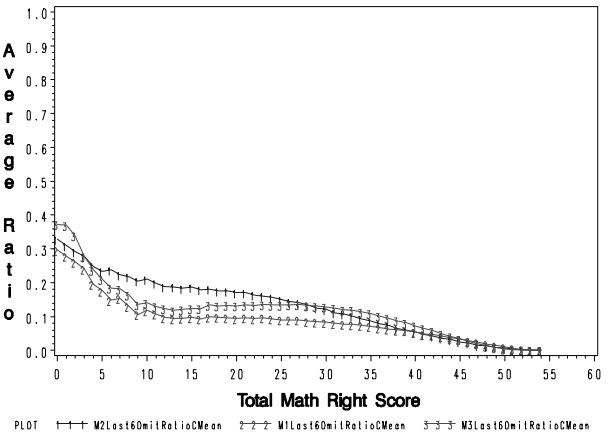

Figure A1.7. Distributions of mean omit ratios on the last six items of M1, M2, and M3 conditional on total math right scores, scrambled spiral, October 2005 SAT administration.

Research Question: Did female and male examinees on the October 2005 SAT administration perform similarly to their March 2005 counterparts on the critical reading, math, and writing sections?

Table A1.5 shows similar numbers and percentages of female and male examinees in the October 2005 SAT administration across the two test spirals. Furthermore, the corresponding percentages of female and male examinees between the March and October 2005 administrations differed by about 1 percent.

Figures A1.8 to A1.13 confirmed that, in terms of right score ratios, both the female and male examinees of the October 2005 administration performed virtually identically on the three critical reading sections, three math sections, and two writing sections, as did their counterparts on the March 2005 administration. Furthermore, the order in which the eight operational sections were presented did not seem to have any systematic effect on examinee performance. These findings held true for the other indices such as mean omit and incorrect response ratios, which were omitted here due to space limitations.

Research Question: Were the October 2005 SAT administration examinees of different racial/ethnic groups distributed similarly across the two test spirals to their March 2005 counterparts?

Table A1.6 shows that between the main and scrambled spirals for the October 2005 SAT administration, the distributions of the eight different racial/ethnic groups were highly similar, just as with the March 2005 administration. For instance, as the smallest racial/ethnic group, American Indian examinees made up about 0.3 percent in both the main and scrambled spirals in the October 2005 administration and the March 2005 administration.

Research Question: Did the examinees of different racial/ethnic groups of the October 2005 SAT administration perform similarly to their counterparts in the March 2005 administration on the three reading, three math, and two writing sections conditional on their abilities?

The proximity of the mean right ratio lines conditional on total critical reading right scores in Figures A1.14 to A1.19 indicates that the October 2005 SAT

Table A1.6

Distributions of Examinees by Racial/Ethnic Group Between Main and Scrambled Spirals,* October 2005 SAT Administration

Spiral	*Racial/ Ethnic Group*	*Frequency*	*Percent*	*Cumulative Percent*	*Cumulative Frequency*
Main	Am Ind	1,066	0.26	4.30	17,969
Main	Am Asian	26,268	6.43	10.73	44,237
Main	Black	16,290	3.99	14.73	60,527
Main	Mex Am	7,320	1.79	16.52	67,847
Main	P Rican	2,083	0.51	17.03	69,930
Main	Latin Am	8,538	2.09	19.12	78,468
Main	White	121,127	29.67	48.79	199,595
Main	Other	7,416	1.82	50.61	207,011
Scrb	Am Ind	1,096	0.27	54.79	224,493
Scrb	Am Asian	25,669	6.29	61.08	250,162
Scrb	Black	15,891	3.89	64.97	266,053
Scrb	Mex Am	7,105	1.74	66.71	273,158
Scrb	P Rican	2,038	0.50	67.21	275,196
Scrb	Latin Am.	8,332	2.04	69.25	283,528
Scrb	White	118,236	28.96	98.22	401,764
Scrb	Other	7,276	1.78	100.00	409,040

*The October 2005 administration had a total of 33,289 examinees missing their ethnicity information, and these examinees were excluded from analyses. This figure, 33,289, was 8,260 larger than its counterpart of 25,029 for the March 2005 administration.

administration examinees of the seven racial/ethnic groups performed highly similarly on the three reading sections across both the main and scrambled spirals, as did the examinees of these different groups for the March 2005 administration. Such performance similarity also held true on the three math and two writing sections across both spirals.

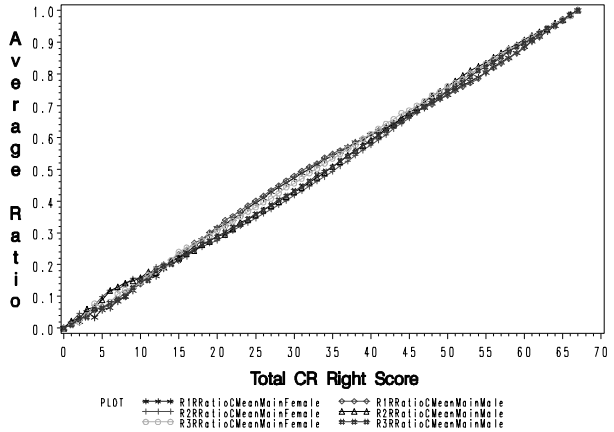

Figure A1.8. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by gender, main spiral, October 2005 SAT administration.

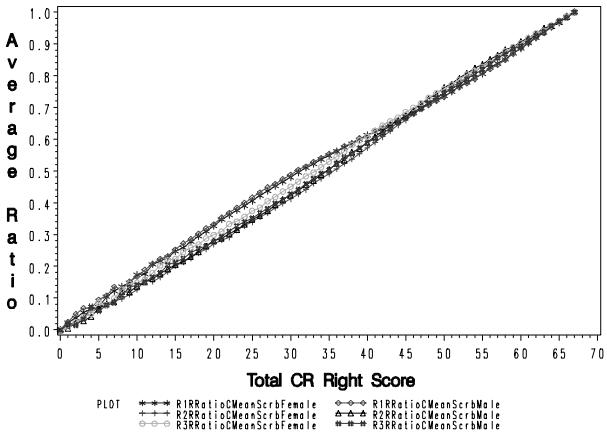

Figure A1.9. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by gender, scrambled spiral, October 2005 SAT administration.

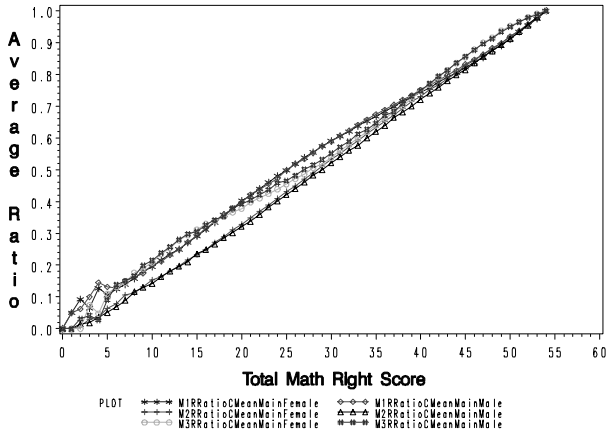

Figure A1.10. Distributions of mean right score ratios of M1, M2, and M3 conditional on total math right scores by gender, main spiral, October 2005 SAT administration.

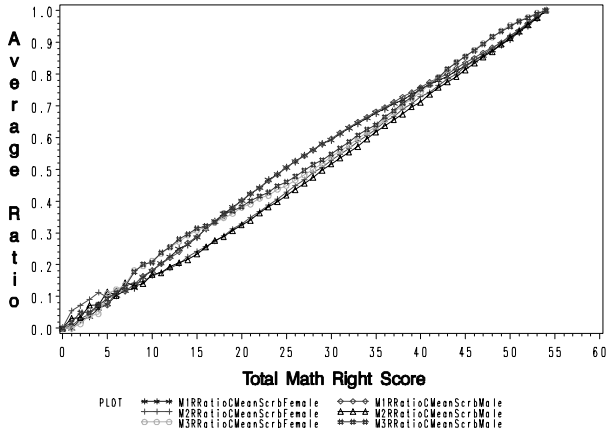

Figure A1.11. Distributions of mean right score ratios of M1, M2, and M3 conditional on total math right scores by gender, scrambled spiral, October 2005 SAT administration.

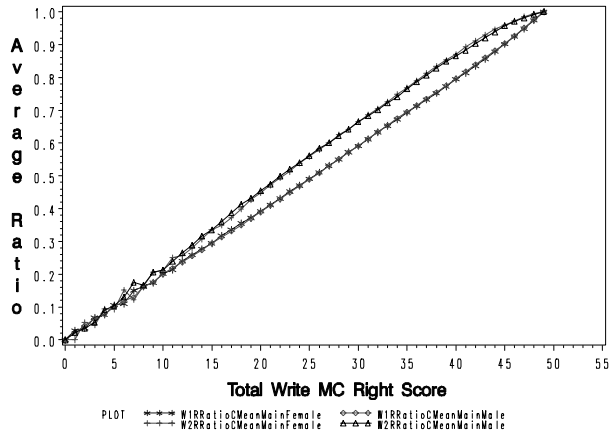

Figure A1.12. Distributions of mean right score ratios of W1 and W2 conditional on total MC writing right scores by gender, main spiral, October 2005 SAT administration.

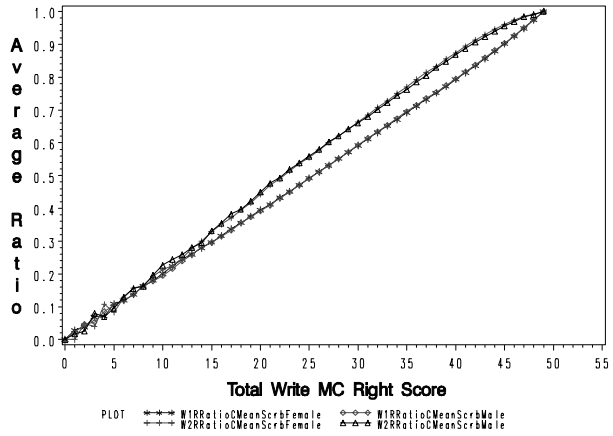

Figure A1.13. Distributions of mean right score ratios of W1 and W2 conditional on total MC writing right scores by gender, scrambled spiral, October 2005 SAT administration.

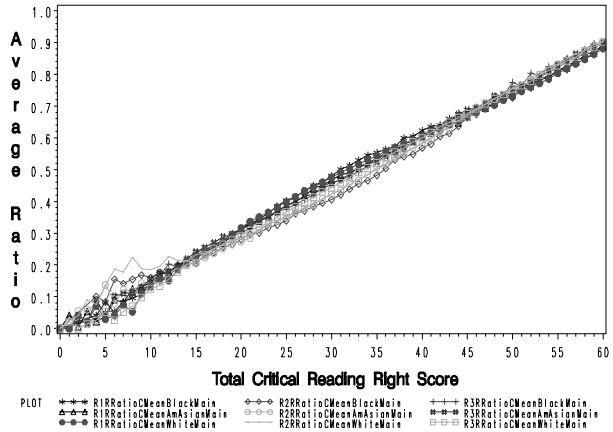

Figure A1.14. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three racial/ethnic groups: Asian American, black, and white, main spiral, October 2005 SAT administration.

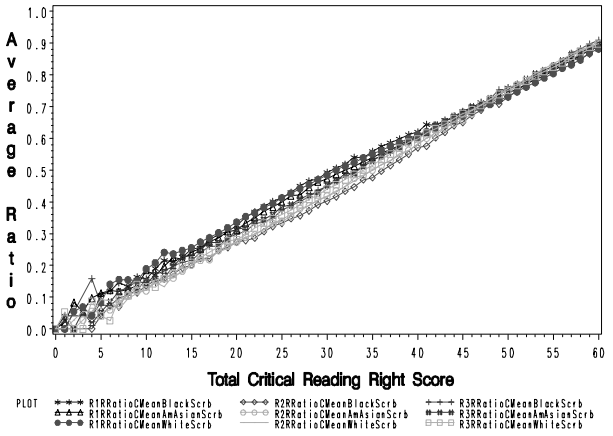

Figure A1.15. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three racial/ethnic groups: Asian American, black, and white, scrambled spiral, October 2005 SAT administration.

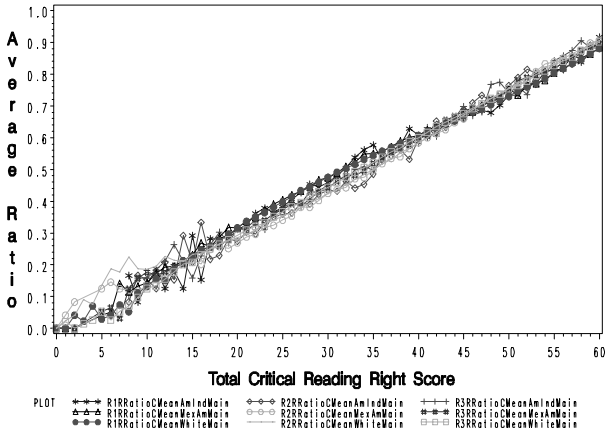

Figure A1.16. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three racial/ethnic groups: American Indian, Mexican American, and white, main spiral, October 2005 SAT administration.

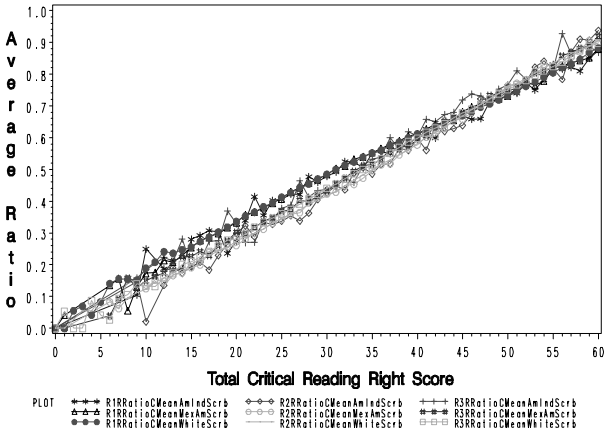

Figure A1.17. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three racial/ethnic groups: American Indian, Mexican American, and white, scrambled spiral, October 2005 SAT administration.

Research Question: Did the October 2005 SAT administration examinees of different best languages perform similarly within and across the two test spirals to their March 2005 counterparts?

Based on the information from the SAT Student Descriptive Questionnaire (SDQ), three best-language classifications were used: English; English and another language; and another language. Table A1.7 confirms that the two test spirals of the October 2005 SAT administration had highly similar numbers and proportions of examinees whose best language was English, English and another language, or another language, as did the March 2005 administration.

Figures A1.20 to A1.31 support the earlier conclusions that examinees of different best-language groups performed virtually identically throughout the entire test in terms of their accuracy (Figures A1.20 to A1.25) as well as their omit patterns (Figures A1.26 to A1.31).

Conclusion

All the analyses and results from the October 2005 SAT administration completely replicated those from the March 2005 administration.

Table A1.7

Distributions of Examinees by Three Best-Language* Groups Between Main and Scrambled Spirals, October 2005 SAT Administration

Spiral	*Best Language*	*Frequency*	*Percent*	*Cumulative Percent*	*Cumulative Frequency*
Main	English	175,048	45.04	45.04	175,048
Main	Eng. & Another Lg	15,842	4.08	49.12	190,890
Main	Another Lg	5,810	1.50	50.61	196,700
Scrb	English	170,982	44.00	94.61	367,682
Scrb	Eng. & Another Lg	15,324	3.94	98.56	383,006
Scrb	Another Lg	5,614	1.44	100.00	388,620

*A total of 20,420 examinees had missing information on their best language and were excluded from this analysis.

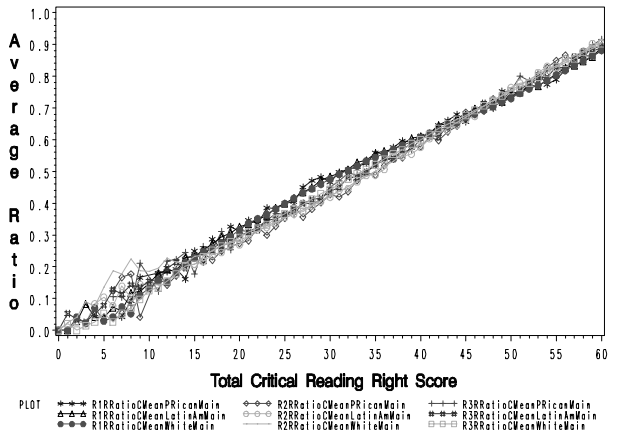

Figure A1.18. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three racial/ethnic groups: Latin American, Puerto Rican, and white, main spiral, October 2005 SAT administration.

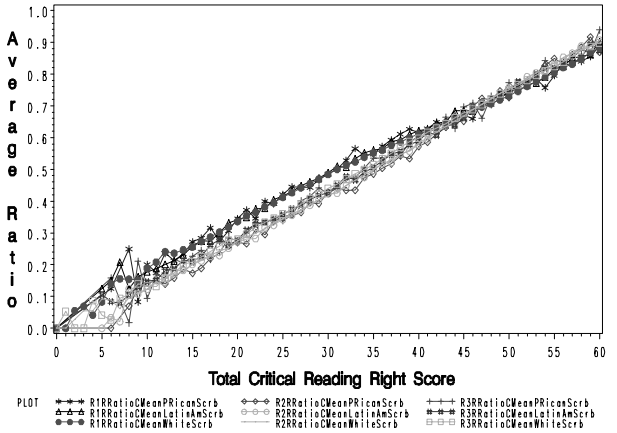

Figure A1.19. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three racial/ethnic groups: Latin American, Puerto Rican, and white, scrambled spiral, October 2005 SAT administration.

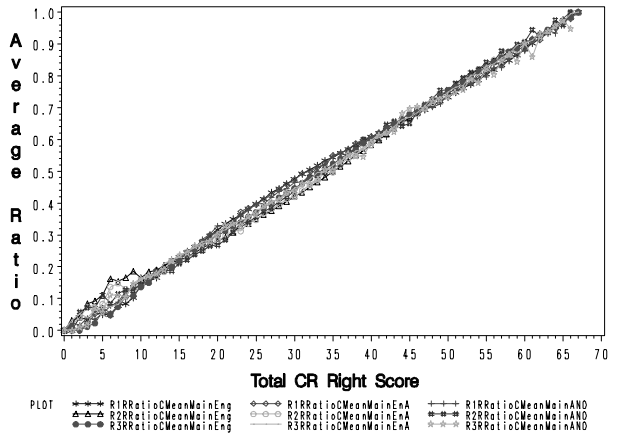

Figure A1.20. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three best-language groups, main spiral, October 2005 SAT administration.

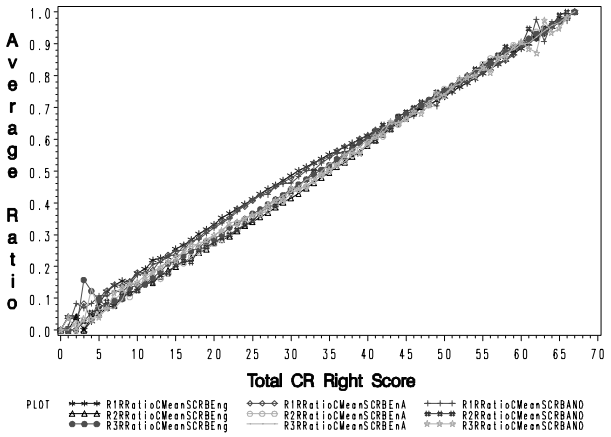

Figure A1.21. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three best-language groups, scrambled spiral, October 2005 SAT administration.

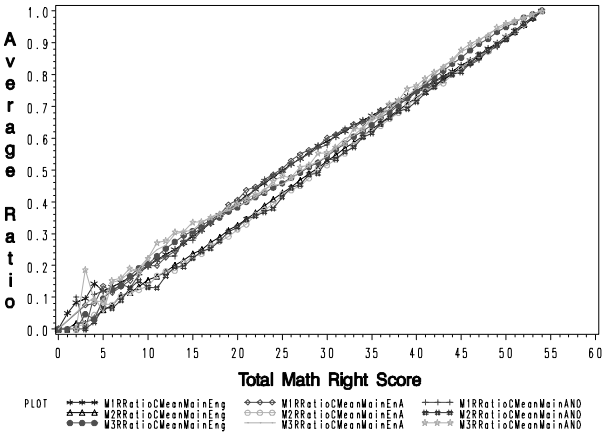

Figure A1.22. Distributions of mean right score ratios of M1, M2, and M3 conditional on total math right scores by three best-language groups, main spiral, October 2005 SAT administration.

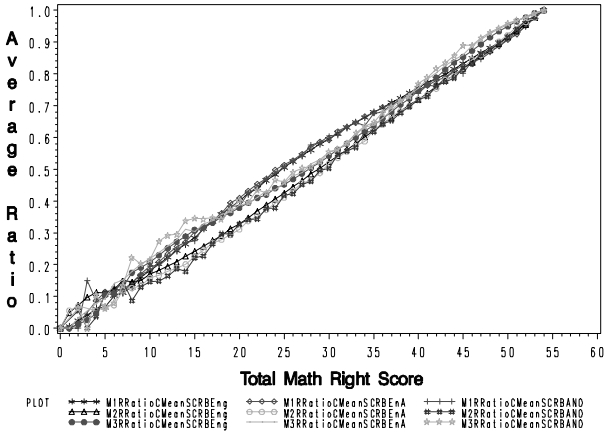

Figure A1.23. Distributions of mean right score ratios of M1, M2, and M3 conditional on total math right scores by three best-language groups, scrambled spiral, October 2005 SAT administration.

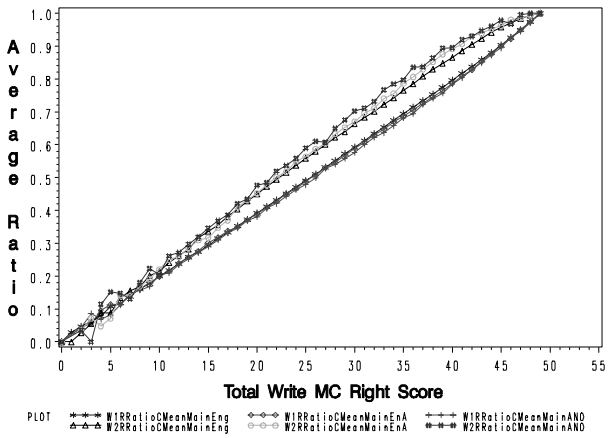

Figure A1.24. Distributions of mean right score ratios of W1 and W2 conditional on total MC writing right scores by three best-language groups, main spiral, October 2005 SAT administration.

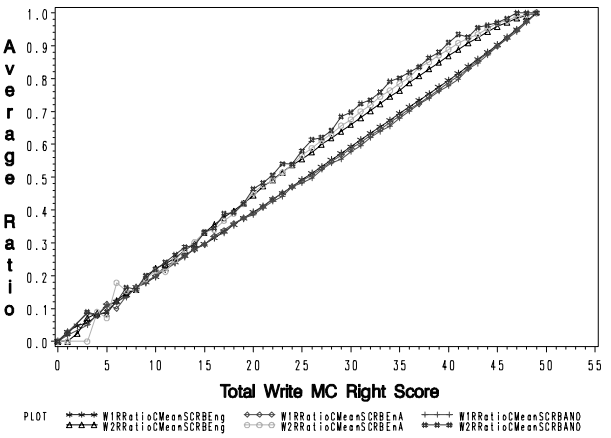

Figure A1.25. Distributions of mean right score ratios of W1 and W2 conditional on total MC writing right scores by three best-language groups, scrambled spiral, October 2005 SAT administration.

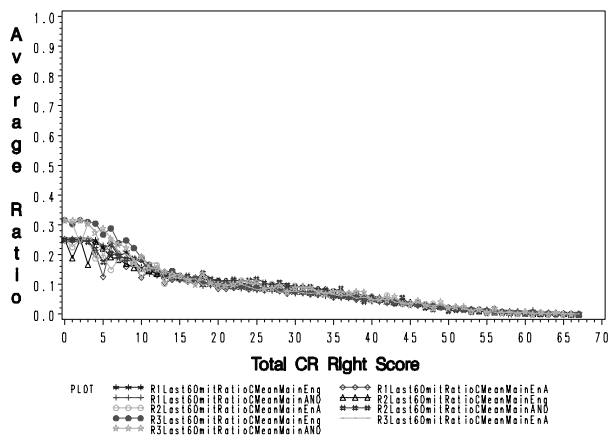

Figure A1.26. Distributions of mean omit ratios on last six items of R1, R2, and R3 conditional on critical reading right scores by three best-language groups, main spiral, October 2005 SAT administration.

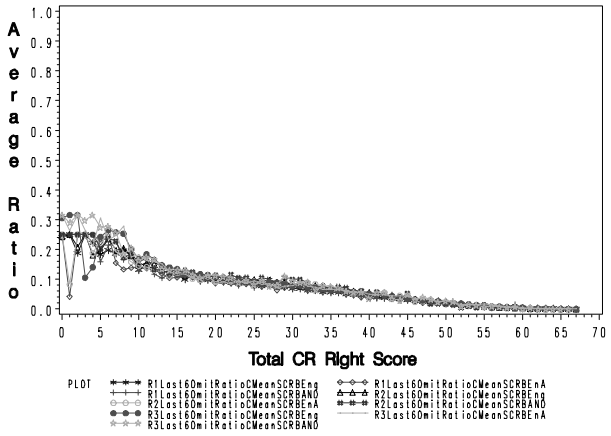

Figure A1.27. Distributions of mean omit ratios on last six items of R1, R2, and R3 conditional on critical reading right scores by three best-language groups, scrambled spiral, October 2005 SAT administration.

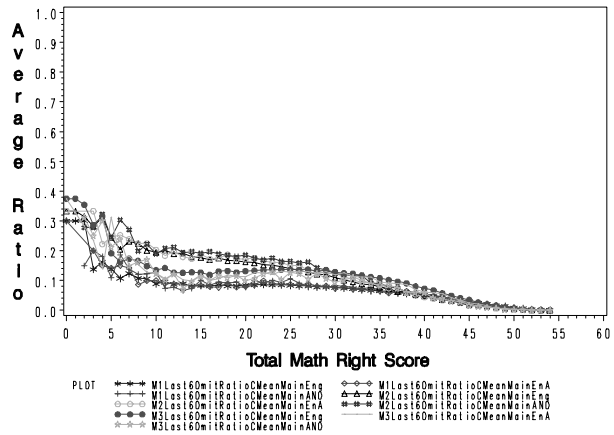

Figure A1.28. Distributions of mean omit ratios on last six items of M1, M2, and M3 conditional on math right scores by three best-language groups, main spiral, October 2005 SAT administration.

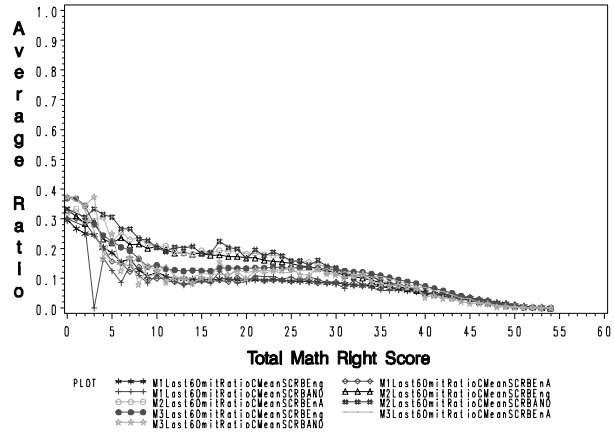

Figure A1.29. Distributions of mean omit ratios on last six items of M1, M2, and M3 conditional on math right scores by three best-language groups, scrambled spiral, October 2005 SAT administration.

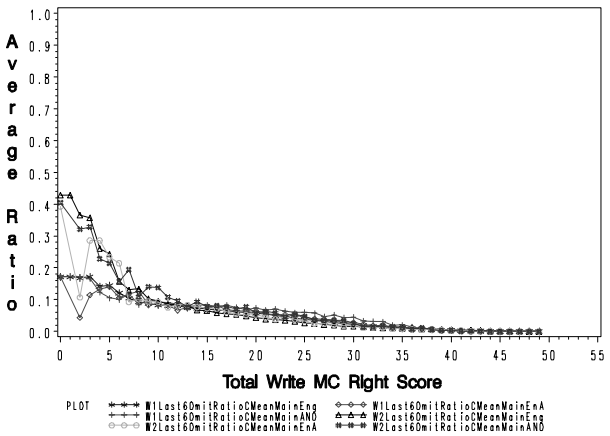

Figure A1.30. Distributions of mean omit ratios on last six items of W1 and W2 conditional on MC writing right scores by three best-language groups, main spiral, October 2005 SAT administration.

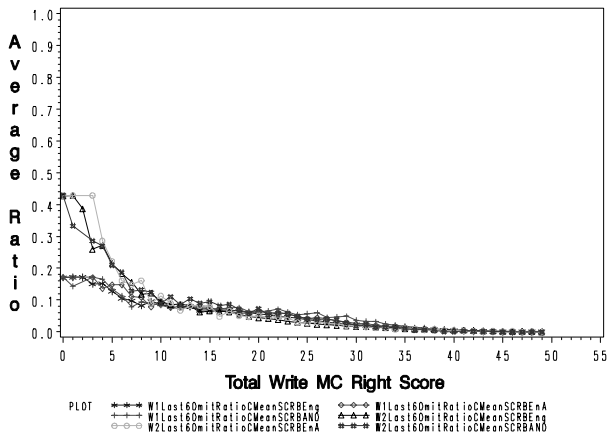

Figure A1.31. Distributions of mean omit ratios on the last six items of W1 and W2 conditional on MC writing right scores by three best-language groups, scrambled spiral, October 2005 SAT administration.

Appendix B: Replications on the May 2002 SAT I: Reasoning Test

After confirming virtually identical results between the March and October 2005 SAT administrations, it is beneficial to investigate to what extent the findings from the two SAT Reasoning Test administrations would hold true for any SAT I: Reasoning Test[12] administration. Data from the May 2002 administration[13] was used. Due to space limitations, only select results are presented here; full results can be provided upon request.

Examinees and Test Spirals of the May 2002 SAT Administration

The replication on the May 2002 SAT administration was carried out on the 437,434 examinees. As indicated in Table A2.1, of the 437,434 examinees, 221,687 took the main spiral, while 215,795 took the scrambled spiral.

Difficulty Levels of the Six Operational Sections of the May 2002 SAT Administration

Table A2.2 shows the average percentages of the examinees who correctly answered the items of the eight operational sections of the May 2002 SAT administration. M1 and M3 were very close in difficulty, differing only by hundredths of a percent, while M2 was slightly harder than M1 and M3. The three reading sections had virtually identical difficulty levels, as did the two writing sections.

Table A2.1

Descriptive Statistics on Critical Reading and Math Sections, May 2002 SAT Administration

Spiral	*Frequency*	*CR Mean*	*CR Std*	*Math Mean*	*Math Std*
Main	221,687	45.17	14.32	35.62	11.75
Scrb	215,795	45.13	14.37	35.53	11.79

Table A2.2

Average Section Difficulty, May 2002 SAT Administration

Section	*# of Items*	*Percent* Correct*		*Weight of One Item*
		Mean	*Std*	
M1	25	0.634	0.208	0.040
M2	25	0.566	0.215	0.040
M3	10	0.613	0.230	0.100
R1	35	0.577	0.207	0.029
R2	32	0.547	0.248	0.031
R3	13	0.529	0.148	0.077

*Operational item difficulty statistics were used instead of pretest counterparts, due to the fact that the latter ones no longer existed.

Analyses and Results

Research Question: *Did the examinees who took the main and scrambled spirals on the May 2002 SAT administration perform similarly to their counterparts on the critical reading and math sections of the March 2005 administration?*

The mean scores and standard deviations of the two examinee groups on the critical reading (CR) and math were virtually identical. Specifically, their CR mean scores were 45.17 versus 45.13, while their math mean scores were 35.62 versus 35.53. Note that the numbers of items on the CR and math sections differed between the SAT I: Reasoning Test and the SAT Reasoning Test.

Figures A2.1 and A2.2 also demonstrate that total score distributions on CR and math of the two examinee groups were virtually identical. Based on these findings, it can be concluded that the examinees who took the main and scrambled spirals on the May 2002 SAT administration performed virtually identically on the CR and math sections, as did their counterparts on the March 2005 administration.

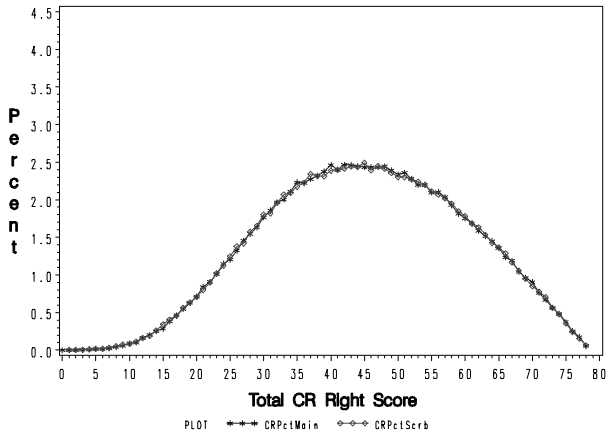

Figure A2.1. Total critical reading score distribution, May 2002 SAT administration.

[12]The SAT I: Reasoning Test did not have a writing component.

[13]The selection of the May 2002 and the October 2005 SAT administration data was made based on the availability of data, and the test structure that was amenable for the purpose of this study. More replications will be conducted on future administrations.

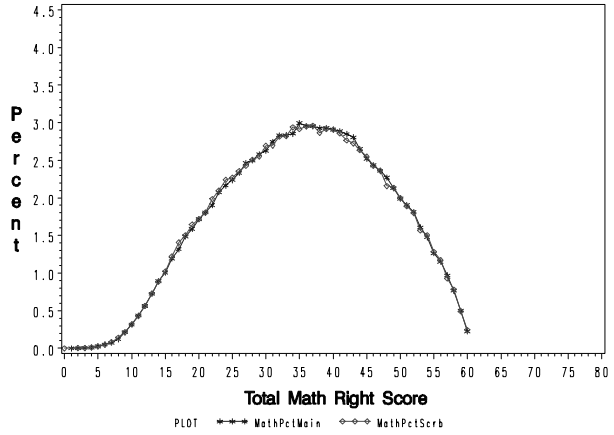

Figure A2.2. Total math score distribution, May 2002 SAT administration.

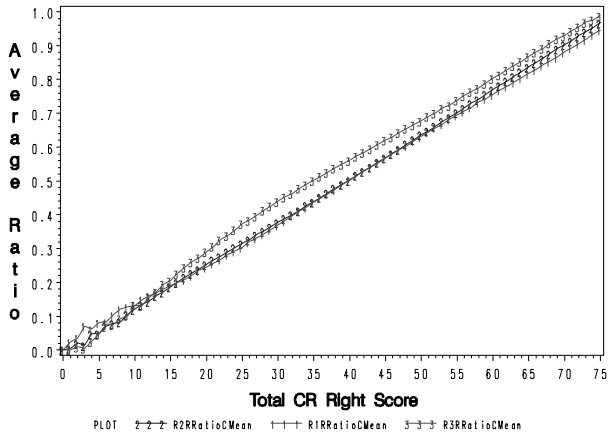

Figure A2.3. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores, main spiral, May 2002 SAT administration.

Research Question: *Did examinees of the May 2002 SAT administration perform similarly on sections of similar content but of different positions across the two spirals, as did the examinees of the March 2005 administration?*

Table A2.3 shows that the average right, wrong, omit, and last-six-omit ratios were also virtually identical on sections of similar content but of different positions across the two spirals for the May 2002 SAT administration. These findings exactly matched those of the March 2005 administration. It can be concluded that examinees of the May 2002 administration seemed to have performed equally well on sections of similar content but of different presentation order, and they did not seem to have rushed toward the end.

Comparing Table A2.3 and Table 5 reveals that the corresponding values in these two tables were highly comparable, even though a number of content and item differences exist between the two administrations. This strongly suggests that the changes implemented in the creation of the SAT Reasoning Test did not seem to have significantly impacted examinee performance.

Research Question: *What were the May 2002 SAT administration distributions of mean right ratios for the three reading sections—conditional on examinee total right reading scores—between the main and scrambled spirals?*

Figures A2.3 and A2.4 confirm the same findings as those for the March 2005 SAT administration. First, the conditional mean right ratio distributions were virtually identical between the main and scrambled spirals. Second, the order of presentation of the sections did not matter. What does matter is the general difficulty level of the sections. Note that section R3 was the last reading section for both the main and scrambled spiral,

Table A2.3

Summary of Four Mean Ratios Across Six Sections and Two Spirals, May 2002 SAT Administration

Section Order			*1*	*2*	*3*	*4*	*5*	*6*
Spiral	Frequency	Ratio	R1	M1	M2	R2	M3	R3
Main	221,687	Right	0.567	0.625	0.558	0.575	0.606	0.621
		Wrong	0.313	0.249	0.271	0.337	0.313	0.301
		Omit	0.120	0.126	0.171	0.089	0.082	0.078
		Last Six Omit	0.024	0.065	0.101	0.021	0.070	0.059
Spiral	Frequency	Ratio	M1	R1	M2	R2	R3	M3
Scrb	215,795	Right	0.618	0.570	0.561	0.571	0.604	0.618
		Wrong	0.255	0.304	0.268	0.343	0.306	0.307
		Omit	0.127	0.125	0.171	0.086	0.090	0.075
		Last Six Omit	0.064	0.027	0.101	0.019	0.057	0.077

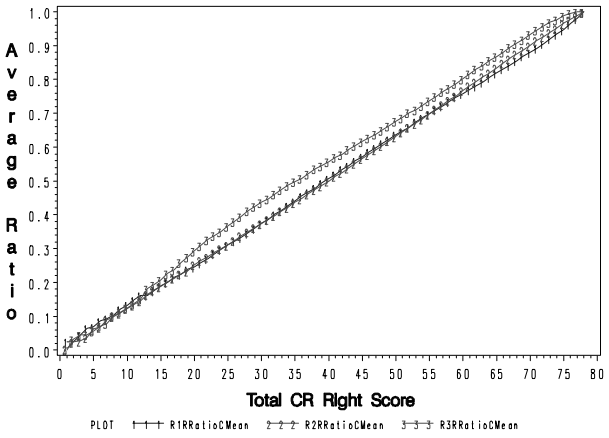

Figure A2.4. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores, scrambled spiral, May 2002 SAT administration.

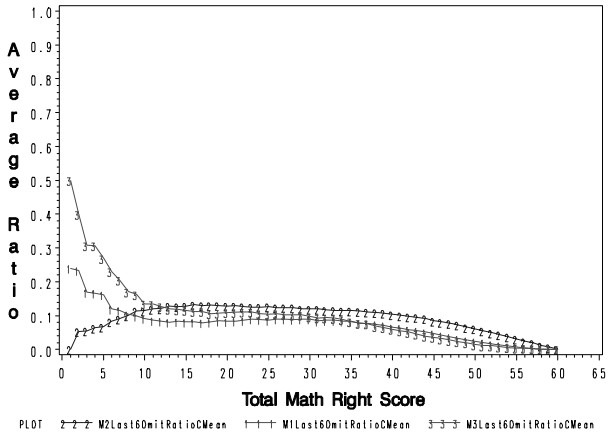

Figure A2.5. Distributions of mean omit ratios on the last six items of M1, M2, and M3 conditional on total math right scores, main spiral, May 2002 SAT administration.

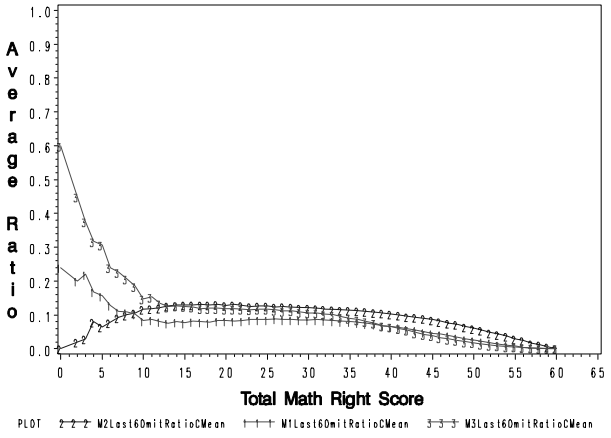

Figure A2.6. Distributions of mean omit ratios on the last six items of M1, M2, and M3 conditional on total math right scores, scrambled spiral, May 2002 SAT administration.

but its conditional average right ratios were the highest. All these findings held true for the three math sections as well, whose results are omitted here for brevity.

Research Question: *What were the May 2002 SAT administration distributions of mean omit ratios on the last six items for the three math sections—conditional on examinee total right math scores—between the main and scrambled spirals?*

Once again, as Figures A2.5 and A2.6 confirm, on the basis of the May 2002 SAT administration, examinees did not necessarily omit more in later math sections, and the tendency of omits seems to be correlated mostly with section difficulty. This finding was also true of the critical reading section, whose results were omitted here. Note again that the high percentages of omits for examinees below the 5-point math score were largely due to a very small sample size, similar to the March 2005 administration.

Research Question: *Did female and male examinees on the May 2002 SAT administration perform similarly to their March 2005 counterparts on the critical reading and math sections?*

Table A2.4 shows that both the main and scrambled spirals of the May 2002 SAT administration had virtually identical female and male examinee distributions. Figures A2.7 to A2.10 confirmed that, in terms of the right score ratios, both the female and male examinees of the May 2002 administration performed virtually identically on the three critical reading and three math sections, as did their counterparts on the March 2005 administration. Furthermore, the order in which the three CR and three math sections were presented did not seem to have any systematic effect on examinee performance. These findings held true of the other indices, such as mean omit and incorrect response ratios.

Table A2.4

Distributions of Female and Male Examinees Between Main and Scrambled Spirals, May 2002 SAT Administration*

Spiral	*Gender*	*Frequency*	*Percent*	*Cumulative Percent*	*Cumulative Frequency*
Main	Female	104,005	23.77	23.77	104,006
Main	Male	117,681	26.90	50.67	221,687
Scrb	Female	101,038	23.10	73.77	322,725
Scrb	Male	114,757	26.23	100.00	437,482

*One examinee had missing gender information and was omitted in this analysis.

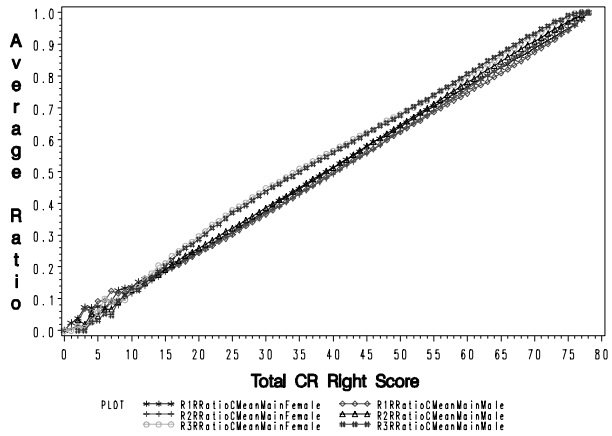

Figure A2.7. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by gender, main spiral, May 2002 SAT administration.

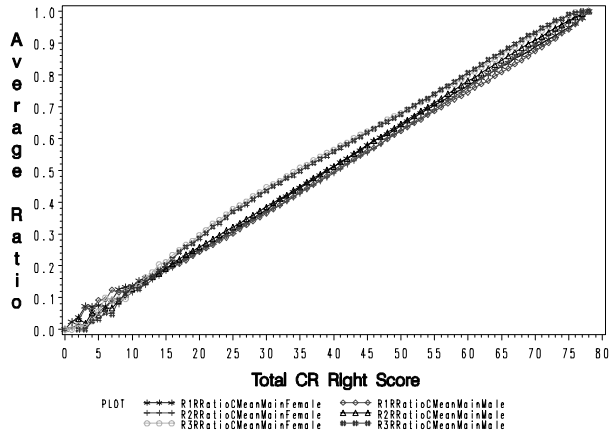

Figure A2.8. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by gender, scrambled spiral, May 2002 SAT administration.

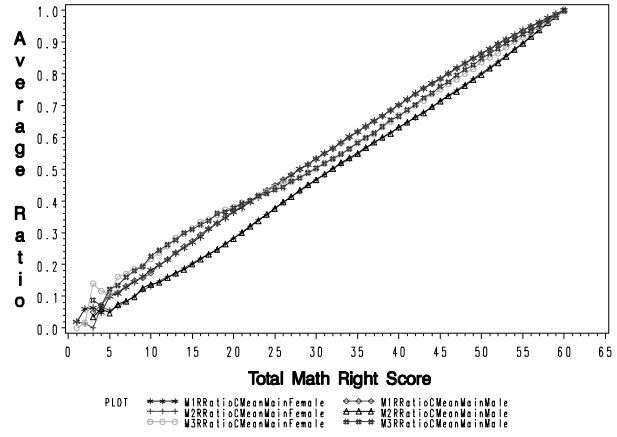

Figure A2.9. Distributions of mean right score ratios of M1, M2, and M3 conditional on total math right scores by gender, main spiral, May 2002 SAT administration.

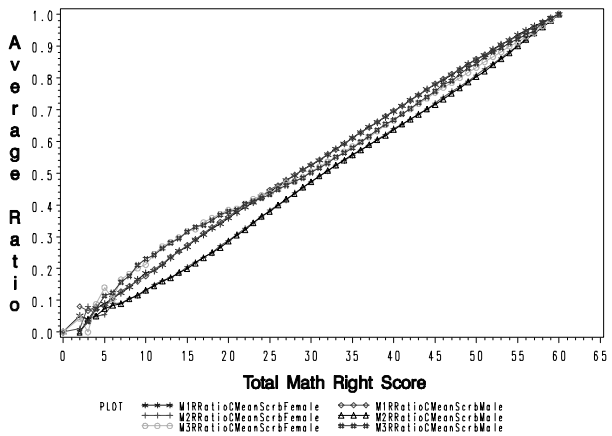

Figure A2.10. Distributions of mean right score ratios of M1, M2, and M3 conditional on total math right scores by gender, scrambled spiral, May 2002 SAT administration.

Research Question: *Did examinees of different racial/ethnic groups perform similarly on the May 2002 SAT administration to their counterparts on the March 2005 administration on the critical reading and math sections?*

Table A2.5 shows that both examinees of the eight racial/ethnic groups were highly similarly distributed between the main and scrambled spirals on the May 2002 SAT administration as they were on the March 2005 administration. Figures A2.11 to A2.13 clearly show that, except for the small fluctuations caused by relatively small sample sizes, examinees of various racial/ethnic groups performed virtually the same on the three main spiral CR sections. These findings were true of the three scrambled CR sections, and of the three math sections under both the main and scrambled spirals, whose results are omitted here.

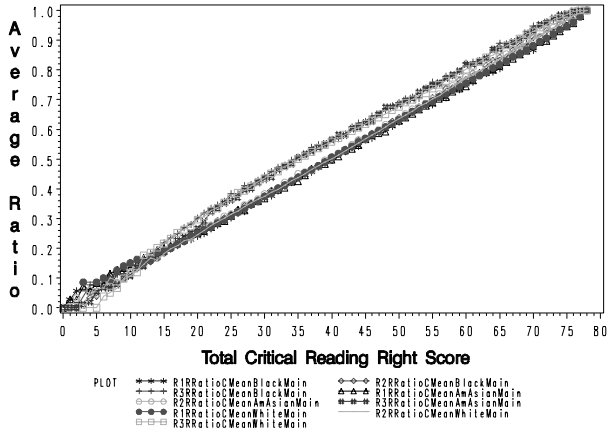

Figure A2.11. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three racial/ethnic groups: Asian American, black, and white, main spiral, May 2002 SAT administration.

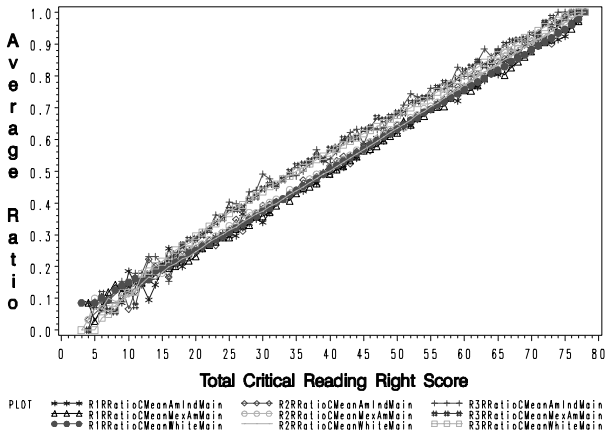

Figure A2.12. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three racial/ethnic groups: American Indian, Mexican American, and white, main spiral, May 2002 SAT administration.

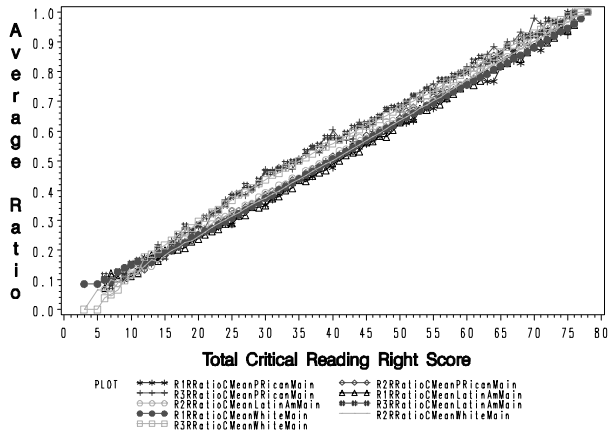

Figure A2.13. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three racial/ethnic groups: Latin American, Puerto Rican, and white, main spiral, May 2002 SAT administration.

Research Question: *Did examinees from different best-language groups perform similarly on the May 2002 SAT administration to their counterparts on the March 2005 administration on critical reading and math sections?*

Table A2.5

Distributions of Examinees of Eight Racial/Ethnic Groups Between Main and Scrambled Spirals, May 2002 SAT Administration

Spiral	*EthnSDQ*	*Frequency*	*Percent*	*Cumulative Percent*	*Cumulative Frequency*
Main	Am Ind	987	0.31	0.31	58,709
	Am Asian	13,509	4.18	4.48	72,218
	Black	14,345	4.44	8.92	86,563
	Mex Am	6,100	1.89	10.81	92,663
	P Rican	1,866	0.58	11.38	94,529
	Latin Am	5,611	1.74	13.12	100,140
	White	116,312	35.97	49.09	216,452
	Other	5,235	1.62	50.71	221,687
Scrb	Am Ind	945	0.29	51.00	279,047
	Am Asian	13,089	4.05	55.05	292,136
	Black	14,056	4.35	59.40	306,192
	Mex Am	5,949	1.84	61.24	312,141
	P Rican	1,876	0.58	61.82	314,017
	Latin Am	5,488	1.70	63.51	319,505
	White	112,972	34.94	98.45	432,477
	Other	5,005	1.55	100.00	437,482

Table A2.6 shows that 305,157 examinees listed English as their best language, followed by 23,348 examinees with English and another language, and only 5,087 examinees with another language as their best language.

Figures A2.14 to A2.21 illustrate that, except for a few total score points due to small sample sizes for the "another language" group, examinees of each best-language group scored highly similarly on the three critical reading and three math item sections as did their March 2005 SAT administration counterparts. Furthermore, the omit rates were similar on the last six items. Such findings were also true of the other indices. These findings lead to the conclusion that the conditional performance of examinees of different best-language groups did not differ significantly between the SAT I: Reasoning Test and the SAT Reasoning Test.

Table A2.6

Distributions of Examinees by Three Best-Language Groups, May 2002 SAT Administration*

BSLANSDQ	*Frequency*	*Percent*	*Cumulative Frequency*	*Cumulative Percent*
English	305,157	91.48	305,157	91.48
English & Another Lg	23,348	7.00	328,505	98.48
Another Lg	5,087	1.52	333,592	100.00

*103,890 examinees had missing information for their best language.

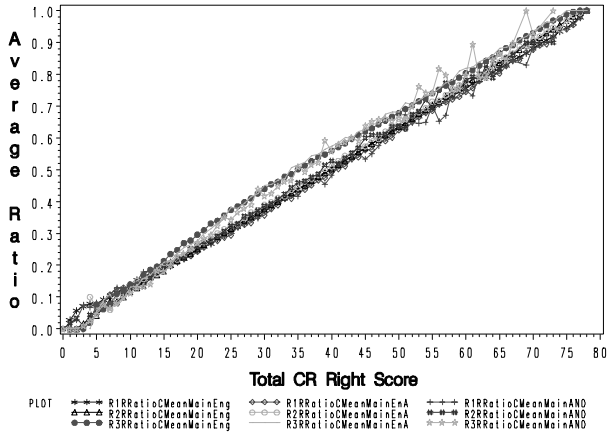

Figure A2.14. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three best-language groups, main spiral, May 2002 SAT administration.

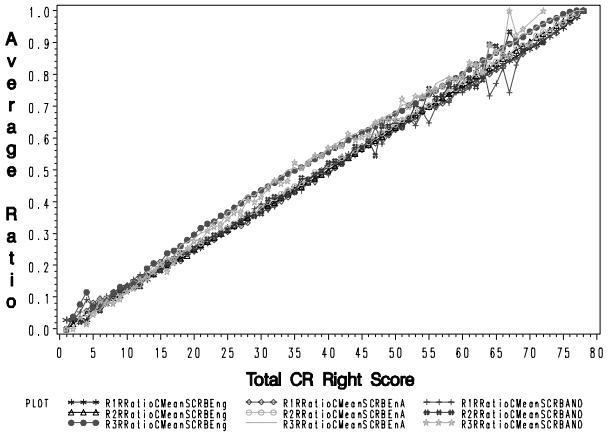

Figure A2.15. Distributions of mean right score ratios of R1, R2, and R3 conditional on total critical reading right scores by three best-language groups, scrambled spiral, May 2002 SAT administration.

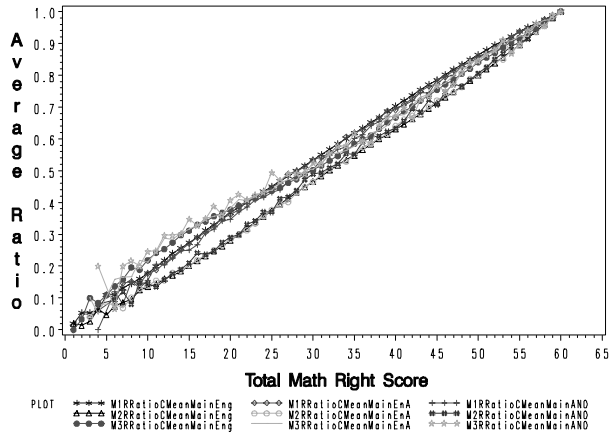

Figure A2.16. Distributions of mean right score ratios of M1, M2, and M3 conditional on total math right scores by three best-language groups, main spiral, May 2002 SAT administration.

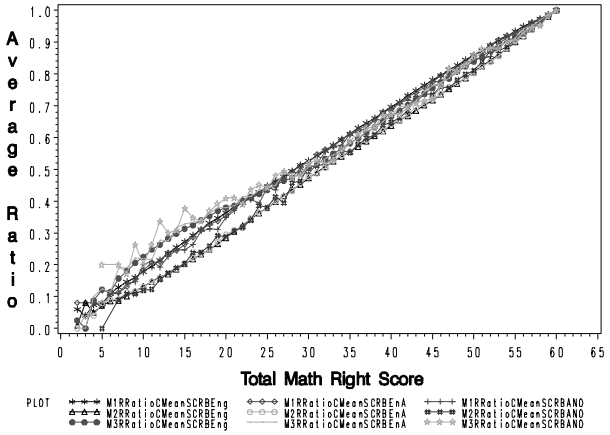

Figure A2.17. Distributions of mean right score ratios of M1, M2, and M3 conditional on total math right scores by three best-language groups, scrambled spiral, May 2002 SAT administration.

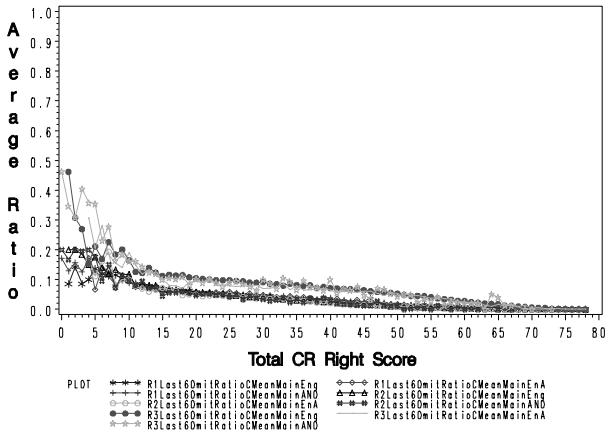

Figure A2.18. Distributions of mean reading omit ratios of R1, R2, and R3 conditional on total critical reading scores by three best-language groups, main spiral, May 2002 SAT administration.

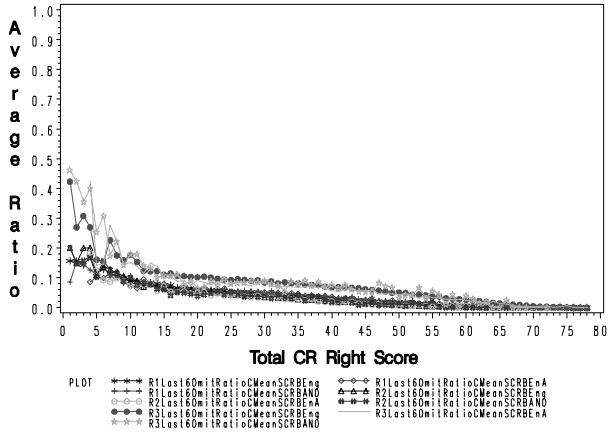

Figure A2.19. Distributions of mean reading omit ratios of R1, R2, and R3 conditional on total critical reading right scores by three best-language groups, scrambled spiral, May 2002 SAT administration.

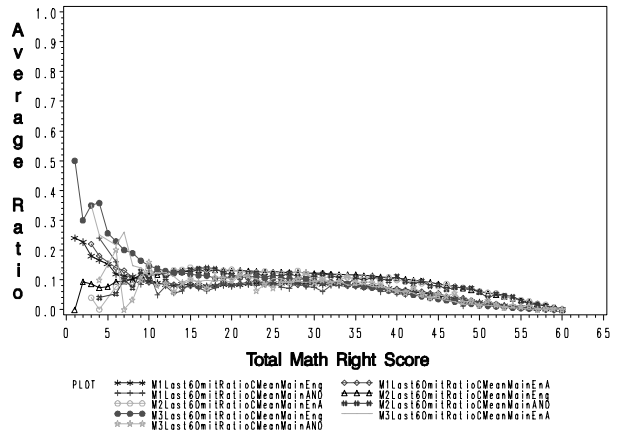

Figure A2.20. Distributions of mean omit ratios of M1, M2, and M3 conditional on total math right scores by three best-language groups, main spiral, May 2002 SAT administration.

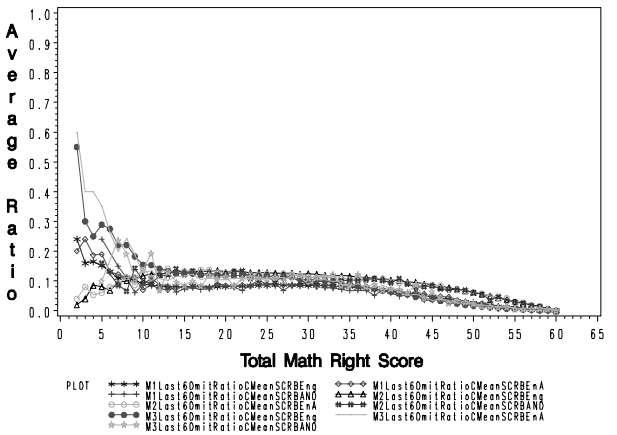

Figure A2.21. Distributions of mean omit ratios of M1, M2, and M3 conditional on total math right scores by three best-language groups, scrambled spiral, May 2002 SAT administration.

Conclusion

Based on the large number of analyses and findings in this study, it can be concluded that, at a group level, the current length of the SAT Reasoning Test does not significantly impact examinee performance at the national level across different gender, racial/ethnic, and best-language groups.


CollegeBoard
connect to college success™