CollegeBoard
connect to college success™

Research Report
No. 2006-7

The Effects of Essay Placement and Prompt Type on Performance on the New SAT®

Hyeon-Joo Oh and Michael E. Walker

www.collegeboard.com

College Board Research Report No. 2006-7
ETS RR-06-34

The Effects of Essay Placement and Prompt Type on Performance on the New SAT®

Hyeon-Joo Oh and Michael E. Walker

The College Board, New York, 2007

Hyeon-Joo Oh is a psychometrician at Educational Testing Service.

Michael E. Walker is a principal psychometrician at Educational Testing Service.

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 #060481999) 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.

Acknowledgments

The research reported in this paper is collaborative in every respect, and the order of authorship is alphabetical. We would like to thank Jill Allspach, Patricia Baron, Hunter Breland, Miriam Feigenbaum, Jinghua Liu, Rosemary Reshetar, Cathy Wendler, and David Wright for their helpful comments on this manuscript.

Contents

Abstract

The new SAT®, which was introduced in March 2005, includes a writing section with essay. This study evaluated (1) whether *essay placement* (either at the beginning or at the end of the test battery) impacts test-takers' performance on the critical reading, mathematics, and writing multiple choice (MC) measures; and (2) whether *essay prompt type* (either a simple one-line prompt or a prompt including a short passage) affects test-takers' essay performance. A total of 2,086 test-takers took the essay section first and 1,921 test-takers took the essay section last. Test-takers' performance on the critical reading, mathematics, writing multiple-choice, and essay sections was then evaluated and the results were compared. The results indicate that essay placement only affects test-takers' performance on the essay itself, not on the other measures. The group that took the essay first performed better on the essay section than the group that took the essay last. A total of 1,975 test-takers received a prompt with an accompanying passage to provide context, and 1,921 test-takers responded to a conventional one-line prompt. The data indicate that the one-line prompt and the contextual prompt have a similar impact on the test-takers' essay performance. Mean differences in performance by subgroup (i.e., gender, race/ethnicity, school type, school location, and language) were also examined.

Introduction

The SAT Reasoning Test™ (SAT) has been changed substantially to strengthen the alignment of the test to the curriculum and instructional practices in high schools and colleges. One of the changes is the addition of the SAT writing section. The writing section is similar to the SAT Subject Test in Writing, which was discontinued in January 2005. It contains multiple-choice grammar questions and an essay.

In the spring of 2003, Educational Testing Service (ETS) undertook a field trial on behalf of the College Board. As its primary focus, the field trial evaluated whether essay placement (either at the beginning or at the end of the test battery) affects test-takers' performance on the SAT. Some test-takers received the essay at the beginning of the test battery (Essay First), while others received the essay at the end (Essay Last). The study compared the average critical reading, mathematics, writing multiple-choice, and essay scores of students who wrote the essay first to the scores of those who wrote the essay last.

A number of discussions were held to determine what kind of prompt should be used for the SAT essay section. One possibility was to use the same type of writing prompt used for the now-defunct SAT Subject Test in Writing (formerly the SAT II: Writing Test; the prompt style will be referred to here as the SAT II prompt). This writing prompt consists of a single sentence stating a position, followed by instructions to write an essay arguing either for or against the position. Another possibility was to use what has been termed a "persuasive" prompt. The persuasive prompt encourages the test-taker to take a position on an issue and to support it with evidence from reading, observation, or experience. The prompt includes a short quotation or paragraph as background information (Appendix C lists an example of each type of prompt). A primary consideration in making a decision on the prompt type was whether changing to a new prompt type would have a negative impact on any racial/ethnic, gender, or English as a second language (ESL) subgroup.

Breland, Kubota, Nickerson, Trapani, and Walker (2003) studied the effects of essay prompt type (persuasive versus SAT II) on test-takers' essay performance by gender, racial/ethnic, and language group. The results from the study indicated that the persuasive prompt and the SAT II prompt have similar impact. Breland et al. found significant racial/ethnic differences in essay performance, with whites and Asian Americans obtaining higher mean scores than African Americans and Hispanics, regardless of prompt type. There was also a significant gender difference in performance (with females outperforming males) regardless of prompt type. Significant mean score differences were observed between EFL (English First Language) and ENFL (English Not First Language) groups. There were no interaction effects among these variables.

In addition to investigating essay placement effects, the current study also investigated the effects of essay prompt type on the test-takers' essay performance by gender, race/ethnicity, school type, school location, and EFL status using the field trial data. Whereas in the Breland et al. (2003) study, examinees only responded to an essay prompt, in the current study test-takers received the entire SAT test battery, including the critical reading, mathematics, and writing multiple-choice portions. Thus it was possible to examine the effects of prompt type in the context in which the essay is given.

A third focus of the research involved group differences in performance on the SAT critical reading, mathematics, and writing measures. Racial/ethnic and gender differences on the SAT have often been studied and are well documented. Another area of interest involves educational access. It stands to reason that those students with more opportunities to learn or with greater resources at their disposal would perform better on academic tests. Availability of resources is influenced to a great extent by type of school (public or private) and

school location (urban, suburban, or rural). This study explored the relationship of school variables (i.e., school type and school location) to the SAT critical reading, mathematics, writing MC, and essay scores.

The following research questions were investigated:

1. Were there differences in performance between the group that took an SAT with the essay section first and the group that took the essay section last? Both groups had identical testing time, so the critical issue was whether essay placement affected test-takers' performance. Those who wrote the essay last could show a decreased level of performance on the rest of the test because they were overly concerned about the essay. On the other hand, those who wrote the essay first could show decreased performance if they continued to worry about the essay during the rest of the test. Similarly, performance on the essay itself could differ because of essay placement (e.g., if examinees taking the essay last were more fatigued at the end of the test). Previous research (Liu and Oh, 2003) found no differential performance on verbal, mathematics, or writing sections based on essay placement. Although the previous study was limited in scope, it was expected that the negligible effect of essay placement on other test components would generalize (effect size estimates in the Liu and Oh study were near zero). Thus, it was hypothesized that test-takers' performance would not differ significantly by essay placement.
2. Were there differences in performance between the group that had a persuasive prompt and the group that had the SAT II prompt? It was hypothesized that there would be no significant differences in performance between essays written to a persuasive prompt and those written to the SAT II prompt. Statistically significant effects, however, were expected among gender, racial/ethnic, and language subgroups regardless of the essay prompt type.
3. Would particular school variables be related to students' performance on the SAT critical reading, mathematics, writing MC, and essay sections? This research question was motivated by a suggestion from the College Board Advisory Committee on Research. In discussing anticipated group differences on the SAT writing section, the committee acknowledged that access to educational opportunity played a major role in the performance of racial/ethnic groups on tests of academic ability (H. Everson, personal communication, April 18, 2002). Educational opportunity comprises a multitude of factors that may be related to socioeconomic status, parental education, parental occupation, primary caretaking during early years, and many more. Another contributing factor is the school environment itself. This research question examines the relationship of school variables (i.e., school type and school location) to test performance to gain some insight into this important issue.

Method

Participants

The data for this study were obtained from the spring 2003 field trial of the new SAT. Usable data were obtained from a total of 501 schools that volunteered for the study. Invitations were sent to 7,500 high schools that had at least 30 college-bound seniors in the previous year. A total of 750 schools responded, with 659 actually returning completed test materials. Of these 659 schools, 501 provided data relevant to the current study. The data were screened for lack of students' motivation and misgridding (Liu and Feigenbaum, 2003). A total of 5,982 test-takers participated in the study reported here; 5,121 of these students (86 percent) indicated that they had taken the PSAT/NMSQT® in either 2001 or 2002. Of the total of 5,982 test-takers, 5,970 test-takers reported their gender and 5,659 test-takers reported their ethnicity. The participants included 2,482 males and 3,488 females; 341 Asian American students including Pacific Islanders, 804 African American students, 719 Hispanic students, and 3,795 white students; 4,895 EFL students; and 1,034 ENFL students. A total of 2,086 students took the essay section first and 1,921 students took the essay section last. And 1,921 students received the SAT II essay prompt, while 1,975 students responded to a persuasive prompt. The number of participants for each condition was not the same because testing took place in intact classrooms, with test books being spiraled to obtain randomly equivalent groups. This procedure is not expected to yield groups of identical size.

Test Materials

The prototype of the SAT test used for the current study included three sections of multiple-choice items for each of the critical reading and mathematics sections: two 25-minute sections and one 20-minute section. One of the 25-minute mathematics sections contained 10 student-produced response (SPR) items in addition to multiple-choice items. For these items, the student must compute the solution to a problem and enter the answer in a grid-in section on the answer sheet. The test included two 25-minute sections for writing: one section of multiple-choice items and one essay section.

Total scaled scores for the critical reading, mathematics, and writing MC sections were obtained by equating the tests to forms of the SAT verbal and mathematics sections and the former Subject Test in Writing, respectively. The scores

were reported on a 20- to 80-point scale to distinguish them from actual SAT scores, which are reported on a 200- to 800-point scale. Essays were holistically scored by two readers on a scale of 1 to 6, using the same scoring guide formerly used for the Subject Test in Writing.[1] The total essay raw score was the sum of the two readers' scores.[2] Essays that were deemed to be off topic were given a score of zero. Thus, the essay scores could range from 0 to 12. No off-topic essays were used in the study. A writing section composite scaled score was also computed by adding the MC and essay raw scores and scaling the resulting composite raw score to the MC section. The writing section composite score was also reported on the 20- to 80-point scale.

Twenty-three test books were administered in the field trial, each containing a slightly different test configuration. Three of the 23 test books in the field trial were used for the current study. These were books 4, 5, and 8:

- Book 4 contained the essay section last
- Book 5 contained the essay section first
- Book 8 contained a persuasive prompt in the last position (books 4 and 5 used the SAT II prompt)

Table 1 shows the order of the test sections for each group. As shown, comparison of books 4 and 5 examined essay placement effects on the test-takers' performance.

Table 1

Section Order of the Test Books for Each Group[a]

Section	*Timing*	*Essay Last (Book 4)*	*Essay First (Book 5)*	*Persuasive Prompt (Book 8)*
1	25 Min.	Reading 1	SAT II Essay	P/N Reading 1
2	25 Min.	Math 1	Math 1	P/N Math 1
3	25 Min.	Writing	Reading 1	P/N Reading 2
4	25 Min.	Math 2[b]	Math 2[b]	P/N Math 2[b]
5	25 Min.	P/N Writing[c]	Writing	Pretest: Reading[c]
6	20 Min.	Reading 3	Reading 3	P/N Writing (37)
7	20 Min.	Math 3	Math 3	Survey[c]
8	25 Min.	Reading 2	Reading 2	Pretest: Math[c]
9	25 Min.	SAT II Essay	Pretest: Writing[c]	Persuasive Essay

[a] Reading 1, 2, and 3 are the new SAT critical reading multiple-choice sections. Math 1, 2, and 3 are the new SAT mathematics multiple-choice sections. Writing refers to the new SAT writing multiple-choice section. PSAT/NMSQT critical reading, mathematics, and writing multiple-choice sections are designated by "P/N" in the label. The SAT and P/N measures were considered comparable here because configurations and difficulty levels for the tests are similar. These similarities are intentional, as the PSAT/NMSQT was designed to be a practice test for the SAT Reasoning Test.

[b] This section contains 10 student-produced responses as well as 8 multiple-choice items.

[c] This variable section, which could contain pretest material, equating sets, or questionnaires, was not counted toward the students' scores.

Similarly, comparing books 4 and 8 revealed any effects of essay prompt type on the test-takers' essay performance. All three books used the same essay topic.

Results

To make sure that spiraling procedures were successful in producing groups that were equivalent in ability (i.e., Essay First versus Essay Last; persuasive prompt versus SAT II prompt), analysis of variance (ANOVA) was performed on test scores. For the essay placement and essay prompt type study, the previous PSAT/NMSQT (P/N) verbal, mathematics, and writing scores were analyzed. A total of 3,423 out of 4,007 participants in the essay placement study had taken the P/N previously in either 2001 or 2002; and 3,353 out of 3,896 participants in the essay prompt study had the P/N scores from either 2001 or 2002. The analysis revealed that there were no mean differences among groups on the P/N measures. Descriptive statistics and results of the analysis of variance for the P/N data by essay placement are presented in Appendix Tables A1 through A6 for the essay placement study and in Appendix Tables B1 and B2 for the essay prompt type study.

Essay Placement Effects

Because preliminary analysis of P/N verbal, mathematics, and writing scores revealed no significant differences between groups in initial ability, analyses of variance were conducted on SAT test scores to determine if essay placement might affect scores. The results of the analyses indicated that essay placement did not affect test-takers' performance on the SAT critical reading, mathematics, or writing MC scores. However, a statistically significant mean difference was found between Essay First and Essay Last groups on essay performance.

Table 2 displays the descriptive statistics for the critical reading, mathematics, writing MC, and essay scores for the two essay position groups (i.e., Essay First versus Essay Last). The mean SAT scores of the two groups were very close for critical reading, mathematics, and writing MC. However, as shown in Table 2, the Essay First group did slightly better than the Essay Last group on the essay.

Table 3 presents the analysis of variance for these differences. At an alpha level of .05, group differences for the Essay First and Essay Last groups were not statistically significant on the measures of critical reading, mathematics, and writing MC ($p > .20$ in all cases). For

[1] Details on the scoring guide for the essays may be found on the College Board Web site at http://www.collegeboard.com/prod_downloads/sat/satguide/SAT2_WR_LT_US_WH.pdf.

[2] When two readers' scores for an essay differed by more than 2 points, the essay was read by a third reader. The essay raw score was the sum of the third reader's score and the closer of the two other readers' scores.

Table 2

Mean SAT Critical Reading, Mathematics, Writing MC, and Essay Scores by Essay Placement Group

	N	*Mean*	*SD*	*Min.*	*Max.*	*Effect Size*
Critical Reading						
Essay First	2,086	48.09	11.14	20	80	
Essay Last	1,921	47.69	10.70	20	80	0.04
Mathematics						
Essay First	2,086	48.93	10.75	20	80	
Essay Last	1,921	48.68	10.57	20	80	0.02
Writing MC						
Essay First	2,086	47.95	10.52	26	80	
Essay Last	1,921	47.76	10.33	26	80	0.02
Essay						
Essay First	2,086	6.49	1.98	2	12	
Essay Last	1,921	5.90	2.02	2	12	0.30

"Effect Size" here refers to the standardized mean difference: $(\overline{X}_{First} - \overline{X}_{Last}) / \sqrt{ms_{WG}}$, where ms_{WG} is a mean square within group.

Table 3

Analysis of Variance for Essay Placement

Source of Variation	*Sum of Squares*	df	*Mean Square*	F	p
Critical Reading					
Between Group	159.64	1	159.64	1.34	0.248
Within Group	478,611.17	4,005	119.50		
Total Group	478,770.81	4,006			
Mathematics					
Between Group	64.40	1	64.40	0.57	0.452
Within Group	455,599.17	4,005	113.76		
Total Group	455,663.57	4,006			
Writing MC					
Between Group	37.00	1	37.00	0.34	0.560
Within Group	435,550.95	4,005	108.75		
Total Group	435,587.96	4,006			
Essay					
Between Group	357.43	1	357.43	89.34	<0.001
Within Group	16,023.82	4,005	4.00		
Total Group	16,381.25	4,006			

the essay score, the group difference did reach statistical significance, with F (1, 4,005) = 89.34, $p < .001$.

The relatively large sample size undoubtedly contributed to the statistical significance of the essay score. Especially in this case, it is important to frame the result of the essay section in terms of effect magnitude as opposed to emphasizing statistical significance (Kirk, 1995). One useful measure is η^2, which is the between-group sum of squares divided by the total sum of squares from the ANOVA. This measure may be interpreted as the proportion of explained variation in the dependent variable. The computed η^2 of the essay was .02, indicating that only 2 percent of the variability in the essay can be explained by the essay placement. This value indicated that variability between groups was fairly small relative to the variability within each group. Nevertheless, the effect size as measured by the standardized mean difference between the groups was .30, or close to one-third of a standard deviation. Thus, the data indicate a small but possibly meaningful increase in scores when the essay is presented at the beginning of the test battery.

Gender analyses

Table 4 shows descriptive statistics for the SAT critical reading, mathematics, writing MC, and essay scores for the two gender groups by essay placement. As shown in

Table 4

Mean SAT Scores of Critical Reading, Mathematics, Writing MC, and Essay by Gender by Essay Placement

	Essay First			*Essay Last*			*Overall*		
	Male	*Female*	*Total*	*Male*	*Female*	*Total*	*Male*	*Female*	*Total*
Critical Reading									
N	851	1,230	2,081	815	1,102	1,917	1,666	2,332	3,998
Mean	48.14	48.04	48.08	47.51	47.84	47.70	47.83	47.94	47.89
SD	11.25	11.07	11.14	10.73	10.68	10.70	11.00	10.88	10.93
Effect Size			0.01			-0.03			-0.01
Mathematics									
N	851	1,230	2,081	815	1,102	1,917	1,666	2,332	3,998
Mean	50.61	47.75	48.92	50.28	47.50	48.68	50.45	47.63	48.81
SD	11.28	10.21	10.75	11.02	10.08	10.57	11.15	10.15	10.66
Effect Size			0.27			0.26			0.27
Writing MC									
N	851	1,230	2,081	815	1,102	1,917	1,666	2,332	3,998
Mean	46.97	48.62	47.95	46.94	48.38	47.77	46.95	48.51	47.86
SD	10.68	10.32	10.50	10.22	10.39	10.34	10.45	10.35	10.42
Effect Size			-0.16			-0.14			-0.15
Essay									
N	851	1,230	2,081	815	1,102	1,917	1,666	2,332	3,998
Mean	6.02	6.83	6.50	5.49	6.20	5.90	5.76	6.53	6.21
SD	2.08	1.85	1.98	2.11	1.89	2.02	2.11	1.89	2.02
Effect Size			-0.41			-0.36			-0.39

"Effect Size" here refers to the standardized mean difference: $(\overline{X}_{Male} - \overline{X}_{Female}) / \sqrt{ms_{WG}}$, where ms_{WG} is a mean square within group.

Table 4, male test-takers in this study performed better on the mathematics section while female test-takers performed better on both the writing MC and the essay sections, regardless of whether they took the essay first or last. Table 5 presents the analysis of variance for these differences. As in the previous analysis, there were no significant effects of essay position on critical reading, mathematics, and writing MC scores. There was, however, a statistically significant main effect of essay placement on essay score, with the Essay First group outperforming the Essay Last group.

At an alpha level of .05, the gender group differences were statistically significant for mathematics, writing MC, and essay scores, $p < .001$ in all cases. Statistically significant gender differences observed in the present study are consistent with the results of the regular SAT and Subject Test administrations. Typically, mean mathematics scores of male test-takers are higher than those of female test-takers, while females show better performance on the writing MC and essay sections. Interestingly, the standardized difference between males and females on the essay was more than twice as large as the standardized difference on the writing MC (–0.41 versus –0.16 for the Essay First condition, and –0.36 versus –0.14 for the Essay Last condition). There were no interaction effects between gender and essay placement on any of the measures.

Table 5

Analysis of Variance for Gender by Essay Placement

Source of Variation	*Sum of Squares*	df	*Mean Square*	F	p
Reading					
Gender	12.72	1	12.72	0.11	0.744
Essay Placement	166.41	1	166.41	1.39	0.238
Gender × Essay Placement	44.52	1	44.52	0.37	0.542
Within Group	477,339.84	3,994	119.51		
Total Group	477,563.49	3,997			
Mathematics					
Gender	7,711.86	1	7,711.86	68.94	< 0.001
Essay Position	81.30	1	81.30	0.73	0.394
Gender × Essay Placement	1.24	1	1.24	0.01	0.916
Within Group	446,798.08	3,994	111.87		
Total Group	454,592.48	3,997			
Writing MC					
Gender	2,324.65	1	2,324.65	21.52	< 0.001
Essay Position	18.23	1	18.23	0.17	0.681
Gender × Essay Placement	10.46	1	10.46	0.10	0.756
Within Group	431,527.03	3,994	108.04		
Total Group	433,880.37	3,997			
Essay					
Gender	556.00	1	556.00	144.10	< 0.001
Essay Position	324.42	1	324.42	84.08	< 0.001
Gender × Essay Placement	2.34	1	2.34	0.61	0.437
Within Group	15,410.59	3,994	3.86		
Total Group	16,293.35	3,997			

Race/ethnicity analyses

Table 6 displays the descriptive statistics for the SAT critical reading, mathematics, writing MC, and essay scores among four racial/ethnic groups. Table 6 also permits comparisons of the SAT critical reading, mathematics, writing MC, and essay scores by racial/ethnic groups. The comparisons showed that Asian American and white students performed better than African American and Hispanic students on all four measures. Table 7 presents the analysis of variance for these differences. At an alpha level of .05, the racial/ethnic group differences were statistically significant for all four measures, with $p < .001$ in all cases.

As with previous analyses, statistically significant differences were found for the essay scores by essay placement. Additionally, there was a statistically significant interaction effect between racial/ethnic group and essay placement on the mathematics section score, F (3, 3,780) = 2.88, $p = .035$. The computed η^2 of the interaction effect was .002, indicating that 0.2 percent of the variability in the mathematics can be explained by the interaction between essay placement and racial/ethnic groups. Table 6 shows that mean mathematics scores of Asian American and white students were slightly higher when these individuals took the essay section first, while mean mathematics scores of African American and Hispanic students were substantially lower when they took the essay section first. However, post hoc tests of these trends (Type I error rate was controlled using a variation of the Bonferroni inequality[3]) failed to detect any significant differences. Given the extremely small magnitude of the interaction effect and the nonsignificance of the post hoc analyses, it was concluded that the statistically significant interaction was no cause for great concern. Nonetheless, the presence of the significant interaction should be kept in mind when evaluating test-takers' mathematics performance by racial/ethnic groups.

[3] The procedure used to control the familywise Type I error rate was the Dunn-Šidák procedure, which is based on a modification of the Bonferroni inequality. Essentially, the procedure divides the overall error rate equally among all comparisons using the multiplicative inequality $\alpha_{PC} = 1 - (1-\alpha_{EN})^{1/k}$, where α_{PC} is the per-comparison alpha level, α_{EN} is the experiment-wise alpha, and k is the number of comparisons. See Kirk (1995, Chapter 4) for details.

Table 6

Mean SAT Scores of Critical Reading, Mathematics, Writing MC, and Essay by Racial/Ethnic Group

	Essay First				*Essay Last*				*Overall*			
	N	***Mean***	***SD***	***ES***	***N***	***Mean***	***SD***	***ES***	***N***	***Mean***	***SD***	***ES***
Critical Reading												
Asian Am.	121	47.82	11.26	0.28	112	46.12	10.35	0.38	233	47.00	10.84	0.33
African Am.	272	40.69	9.49	0.97	257	42.27	10.03	0.75	529	41.46	9.78	0.87
Hispanic	249	42.88	10.40	0.76	231	42.19	9.42	0.76	480	42.55	9.94	0.76
White	1,338	50.69	10.46		1,208	50.01	10.30		2,546	50.37	10.39	
Mathematics												
Asian Am.	121	52.92	11.48	-0.15	112	51.46	10.79	-0.07	233	52.21	11.15	-0.11
African Am.	272	41.43	8.88	1.01	257	42.86	9.96	0.79	529	42.12	9.44	0.90
Hispanic	249	42.49	9.99	0.90	231	43.70	9.23	0.71	480	43.07	9.64	0.81
White	1,338	51.45	9.94		1,208	50.73	10.04		2,546	51.11	9.99	
Writing MC												
Asian Am.	121	47.58	11.27	0.28	112	46.09	9.39	0.38	233	46.86	10.41	0.13
African Am.	272	41.63	7.99	0.89	257	43.40	9.44	0.65	529	42.49	8.76	0.77
Hispanic	249	42.76	8.81	0.77	231	43.00	8.04	0.69	480	42.88	8.44	0.73
White	1,338	50.37	10.26		1,208	49.86	10.33		2,546	50.13	10.29	
Essay												
Asian Am.	121	6.80	2.04	-0.06	112	6.18	2.06	-0.03	233	6.50	2.07	-0.04
African Am.	272	5.73	1.90	0.49	257	5.22	1.98	0.46	529	5.48	1.96	0.48
Hispanic	249	6.11	1.97	0.30	231	5.37	1.71	0.38	480	5.75	1.89	0.34
White	1,338	6.69	1.95		1,208	6.12	2.00		2,546	6.42	1.99	

"Effect Size" (ES) here refers to the standardized mean difference: $(\overline{X}_{White} - \overline{X}_{Nonwhite}) / \sqrt{ms_{WG}}$, where ms_{WG} is a mean square within group.

Table 7

Analysis of Variance for Race/Ethnicity by Essay Placement

Source of Variation	***Sum of Squares***	**df**	***Mean Square***	**F**	**p**
Critical Reading					
Race/Ethnicity	51,022.31	3	17,007.44	161.22	<0.001
Essay Placement	64.56	1	64.56	0.61	0.434
Race/Ethnicity × Placement	670.84	3	223.61	2.12	0.096
Within Group	398,755.62	3,780	105.49		
Total Group	450,513.33	3,787			
Mathematics					
Race/Ethnicity	54,990.37	3	18,330.12	185.41	<0.001
Essay Placement	6.09	1	6.09	0.06	0.804
Race/Ethnicity × Placement	853.24	3	284.41	2.88	0.035
Within Group	373,710.35	3,780	98.87		
Total Group	427,002.93	3,787			
Writing MC					
Race/Ethnicity	40,078.56	3	13,359.52	136.99	<0.001
Essay Placement	0.01	1	0.01	0.00	0.991
Race/Ethnicity × Placement	697.39	3	232.46	2.38	0.067
Within Group	368,623.60	3,780	97.52		
Total Group	409,399.56	3,787			
Essay					
Race/Ethnicity	504.76	3	168.25	43.97	<0.001
Essay Placement	172.74	1	172.74	45.14	<0.001
Race/Ethnicity × Placement	3.78	3	1.26	0.33	0.804
Within Group	14,464.80	3,780	3.83		
Total Group	15,146.08	3,787			

Language analyses

Table 8 displays descriptive statistics for reading, mathematics, writing MC, and essay scores for each self-reported language group. As expected, mean scores of the English First Language (EFL) group were slightly higher than the scores of the English Not First Language (ENFL) group. Further analyses were performed to explore whether those differences were statistically significant. Table 9 shows the result of the analysis of variance comparing performance differences on the critical

Table 8

Mean SAT Scores of Critical Reading, Mathematics, Writing MC, and Essay by Language Group

	Essay First				*Essay Last*			
	N	***Mean***	***SD***	***ES***	***N***	***Mean***	***SD***	***ES***
Critical Reading								
EFL	1,706	49.08	10.99		1,580	48.67	10.55	
ENFL	363	43.67	10.83	0.50	325	43.04	10.18	0.52
Mathematics								
EFL	1,706	49.70	10.44		1,580	49.32	10.45	
ENFL	363	45.45	11.62	0.40	325	45.81	10.71	0.33
Writing MC								
EFL	1,706	48.93	10.41		1,580	48.65	10.32	
ENFL	363	43.57	9.93	0.52	325	43.55	9.32	0.50
Essay								
EFL	1,706	6.59	1.95		1,580	5.96	2.01	
ENFL	363	6.11	2.07	0.24	325	5.63	2.06	0.17

"Effect Size" (ES) here refers to the standardized mean difference: $(\overline{X}_{EFL} - \overline{X}_{ENFL}) / \sqrt{ms_{WG}}$, where ms_{WG} is a mean square within group.

Table 9

Analysis of Variance for Language by Essay Placement

Source of Variation	*Sum of Squares*	df	*Mean Square*	F	p
Critical Reading					
Language	17,269.36	1	17,269.36	149.88	< 0.001
Essay Placement	151.97	1	151.97	1.32	0.251
Language × Placement	7.41	1	7.41	0.06	0.800
Within Group	457,440.46	3,970	115.22		
Total Group	474,869.20	3,973			
Mathematics					
Language	8,536.17	1	8,536.17	76.32	< 0.001
Essay Placement	0.15	1	0.15	0.00	0.970
Language × Placement	76.86	1	76.86	0.69	0.407
Within Group	444,044.06	3,970	111.85		
Total Group	452,657.24	3,973			
Writing MC					
Language	15,518.90	1	15,518.90	147.76	< 0.001
Essay Placement	12.74	1	12.74	0.12	0.728
Language × Placement	9.64	1	9.64	0.09	0.762
Within Group	416,958.60	3,970	105.03		
Total Group	432,499.88	3,973			
Essay					
Language	91.58	1	91.58	23.07	< 0.001
Essay Placement	177.30	1	177.30	44.67	< 0.001
Language × Placement	3.09	1	3.09	0.78	0.378
Within Group	15,756.35	3,970	3.97		
Total Group	16,028.32	3,973			

reading, mathematics, writing MC, and essay sections between language groups. At an alpha level of .05, the group differences between EFL and ENFL for the SAT measures were statistically significant, with $p < .001$ in all cases. There were no statistically significant differences in performance by essay placement except on the essay score. There were no interaction effects between language groups and essay placement on any of the measures.

School type analyses

Table 10 displays performance on the SAT measures by school type (i.e., public versus private). Mean scores of private school students on the four measures of the SAT were slightly higher than the mean scores of public school students. Table 11 presents the results of the analysis of variance for these comparisons. At an alpha level of .05, statistically significant effects of school type were found on all four SAT measures, with $p < .001$ in all cases. The results indicated that average performance of private school students on the SAT was better than that of public school students.

Table 10

Mean SAT Scores of Critical Reading, Mathematics, Writing MC, and Essay by School Type

	Essay First			*Essay Last*		
	N	*Mean*	*SD*	*N*	*Mean*	*SD*
Critical Reading						
Public	1,377	47.12	11.10	1,281	46.59	10.40
Private	707	50.00	11.00	637	49.97	10.90
Mathematics						
Public	1,377	48.53	10.67	1,281	48.14	10.48
Private	707	49.75	10.86	637	49.81	10.66
Writing MC						
Public	1,377	47.07	10.30	1,281	46.58	9.88
Private	707	49.72	10.70	637	50.19	10.80
Essay						
Public	1,377	6.40	1.98	1,281	5.73	2.00
Private	707	6.67	1.97	637	6.25	2.01

There were no statistically significant differences in performance by essay placement except on the essay score. There were no interaction effects between school type and essay placement on any of the measures.

Table 11

Analysis of Variance for School Type by Essay Placement

Source of Variation	*Sum of Squares*	df	*Mean Square*	F	p
Critical Reading					
School Type	8,744.46	1	8,744.46	74.57	< 0.001
Essay Placement	68.28	1	68.28	0.58	0.445
School × Placement	58.37	1	58.37	0.50	0.481
Within Group	468,827.18	3,998	117.27		
Total Group	477,698.29	4,001			
Mathematics					
School Type	1,852.07	1	1,852.07	16.36	< 0.001
Essay Placement	24.36	1	24.36	0.22	0.643
School × Placement	42.98	1	42.98	0.38	0.538
Within Group	425,688.74	3,998	113.23		
Total Group	427,608.15	4,001			
Writing MC					
School Type	8,750.33	1	8,750.33	82.14	< 0.001
Essay Placement	0.02	1	0.02	0.00	0.988
School × Placement	208.73	1	208.73	1.96	0.162
Within Group	425,920.66	3,998	106.53		
Total Group	434,879.74	4,001			
Essay					
School Type	140.86	1	140.86	35.56	< 0.001
Essay Placement	270.41	1	270.41	68.26	< 0.001
School × Placement	14.48	1	14.48	3.66	0.056
Within Group	15,837.51	3,998	3.96		
Total Group	16,263.26	4,001			

Table 12

Mean SAT Scores of Critical Reading, Mathematics, Writing MC, and Essay by School Location

	Essay First			*Essay Last*		
	N	***Mean***	***SD***	***N***	***Mean***	***SD***
Critical Reading						
Large City	573	44.71	12.10	519	47.61	11.69
Medium City	264	47.08	11.34	243	46.06	10.38
Small City	439	47.26	10.63	417	47.53	10.41
Suburban	573	50.14	10.86	521	49.34	10.33
Rural	171	45.53	9.37	160	46.15	9.45
Mathematics						
Large City	573	47.97	11.37	519	48.71	11.11
Medium City	264	47.46	11.18	243	47.74	10.84
Small City	439	48.43	10.30	417	48.41	10.17
Suburban	573	51.36	10.50	521	49.93	10.58
Rural	171	46.85	8.68	160	46.89	9.21
Writing MC						
Large City	573	47.63	11.33	519	47.83	11.03
Medium City	264	47.00	10.27	243	46.64	10.15
Small City	439	47.68	10.12	417	47.78	10.00
Suburban	573	49.63	10.57	521	48.76	10.57
Rural	171	45.60	8.67	160	46.46	8.59
Essay						
Large City	573	6.62	2.02	519	5.96	2.08
Medium City	264	6.44	2.07	243	5.85	2.04
Small City	439	6.28	1.97	417	5.70	2.05
Suburban	573	6.62	1.98	521	6.19	1.94
Rural	171	6.14	1.93	160	5.90	1.87

Table 13

Analysis of Variance for School Location by Essay Placement

Source of Variation	***Sum of Squares***	**df**	***Mean Square***	**F**	**p**
Critical Reading					
School Location	6,273.78	4	1,568.45	13.16	< 0.001
Essay Placement	33.85	1	33.85	0.28	0.594
Location × Placement	281.77	4	70.44	0.59	0.669
Within Group	461,092.26	3,870	119.15		
Total Group	467,681.66	3,879			
Mathematics					
School Location	6,055.82	4	1,513.95	13.43	< 0.001
Essay Placement	4.80	1	4.80	0.43	0.836
Location × Placement	689.75	4	172.44	1.53	0.191
Within Group	436,304.35	3,870	112.74		
Total Group	443,054.72	3,879			
Writing MC					
School Location	3,623.19	4	905.80	8.31	< 0.001
Essay Placement	0.13	1	0.13	0.00	0.972
Location × Placement	277.30	4	69.33	0.64	0.637
Within Group	422,081.38	3,870	109.07		
Total Group	425,982.00	3,879			
Essay					
School Location	153.43	4	38.36	9.55	< 0.001
Essay Placement	280.31	1	280.31	69.78	< 0.001
Location × Placement	9.45	4	2.36	0.59	0.671
Within Group	15,546.99	3,870	4.02		
Total Group	15,990.18	3,879			

School location analyses

Table 12 shows the descriptive statistics for the critical reading, mathematics, writing MC, and essay sections by test-takers' school location. The mean scores of five school locations (i.e., large city, medium city, small city, suburban, and rural) were compared and analyzed. In general, it appeared that the mean scores of the four SAT measures in the suburban area were the highest among the five locations. Table 13 presents the analysis of variance to examine statistical significance for these differences. At an alpha level of .05, the group differences among the five school locations were statistically significant for all four measures, $p < .001$ in all cases. The Bonferroni post hoc test revealed that students in a suburban area did better than the students in any other area on the four measures of the SAT.

Statistically significant effects of essay placement were found only on the essay scores. There were no interaction effects between school location and essay placement on any of the four SAT measures.

Essay Prompt Type Effects

The analyses of essay scores indicated that prompt type does not affect test-takers' essay performance. The current results were consistent with the previous study done by Breland et al. (2003). Analyses of covariance (controlling for critical reading, mathematics, and writing MC scores) were not conducted because, as mentioned earlier, the preliminary analyses of the SAT critical reading, mathematics, and writing MC data verified that there were no mean differences between the groups (i.e., those receiving the SAT II prompt versus the persuasive prompt) on these tests.

Table 14 displays descriptive statistics of the essay scores between the two groups (i.e., SAT II versus persuasive). The mean essay scores for the two groups were very close, as indicated by the standardized difference of -0.04 by prompt type. Table 15 presents the analysis of variance for the two different types of prompt on essay scores. At an alpha level of .05, the group differences were not statistically significant. Thus, the data offered no evidence that essay prompt type affected the essay scores.

Gender analyses

Table 16 shows the mean essay scores of the two gender groups by essay prompt type. Female test-takers outperformed male test-takers on the essay, for both

Table 14

Mean Essay Scores by Essay Prompt Type

Prompt Type	***N***	***Mean***	***SD***	***Min.***	***Max.***	***Effect Size***
SAT II	1,921	5.90	2.01	2	12	-0.04
Persuasive	1,975	5.98	1.94	2	12	

"Effect Size" here refers to the standardized mean difference: $(\overline{X}_{SATII} - \overline{X}_{Persuasive}) / \sqrt{ms_{WG}}$, where ms_{WG} is a mean square within group.

Table 15

Analysis of Variance for Essay Prompt Type

Source of Variation	*Sum of Squares*	df	*Mean Square*	F	p
Essay					
Between Group	7.35	1	7.35	1.88	0.171
Within Group	15,256.83	3,894	3.92		

Table 16

Mean Essay Scores by Gender and Essay Prompt Type

	SAT II Prompt				*Persuasive Prompt*			
	N	*Mean*	*SD*	*ES*	*N*	*Mean*	*SD*	*ES*
Male	815	5.49	2.11		816	5.58	1.98	
Female	1,102	6.20	1.89	-0.36	1,156	6.28	1.86	-0.36

"Effect Size" (ES) here refers to the standardized mean difference: $(\overline{X}_{Male} - \overline{X}_{Female}) / \sqrt{ms_{WG}}$, where ms_{WG} is a mean square within group.

Table 17

Analysis of Variance for Gender by Essay Prompt Type

Source of Variation	*Sum of Squares*	df	*Mean Square*	F	p
Gender	468.58	1	468.58	123.36	< 0.001
Essay Prompt	6.18	1	6.18	1.63	0.202
Gender × Essay Prompt	0.02	1	0.02	0.01	0.944
Within Group	14,756.97	3,885	3.80		
Total Group	15,231.75	3,888			

Table 18

Mean Essay Scores for Racial/Ethnic Group by Essay Prompt Type

	SAT II Prompt				*Persuasive Prompt*			
	N	*Mean*	*SD*	*ES*	*N*	*Mean*	*SD*	*ES*
Asian Am.	112	6.18	2.06	-0.03	108	6.19	2.72	0.01
African Am.	257	5.22	1.98	0.46	275	5.34	1.88	0.45
Hispanic	231	5.37	1.71	0.39	239	5.41	1.88	0.41
White	1,208	6.12	2.00	—	1,249	6.21	1.90	—

"Effect Size" (ES) here refers to the standardized mean difference: $(\overline{X}_{White} - \overline{X}_{Nonwhite}) / \sqrt{ms_{WG}}$, where ms_{WG} is a mean square within group.

Table 19

Analysis of Variance for Race/Ethnicity by Essay Prompt Type

Source of Variation	*Sum of Squares*	df	*Mean Square*	F	p
Race/Ethnicity	510.99	3	170.33	45.21	< 0.001
Essay Prompt	2.06	1	2.06	1.61	0.460
Race/ Ethnicity × Essay Prompt	0.78	3	0.07	0.07	0.977
Within Group	13,830.72	3,671	3.77		
Total Group	14,344.55	3,678			

prompt types. Table 17 presents the analysis of variance of essay score by gender and prompt type. At an alpha level of .05, the mean difference between gender groups on the essay scores was statistically significant, with F (1, 3,885) = 123.36, p < .001. As in the previous analysis, the difference across essay prompt styles was not statistically significant. There were no interaction effects between gender and essay prompt type.

Race/ethnicity analyses

Table 18 displays the means and standard deviations of the essay scores among the four ethnic groups. The results showed that Asian American and white students did better than African American and Hispanic students on both essays (i.e., persuasive prompt and SAT II prompt). Table 19 presents the analysis of variance for these differences. At an alpha level of .05, the group effect was statistically significant, with F (1, 3,671) = 45.21, p = .001. Follow-up comparisons showed statistically significant essay impact (with white students as the comparison group) for African American and Hispanic students, but not for Asian American students. The essay prompt type effect was not significant. There were no interaction effects between ethnic groups and essay prompt type. Note in particular that the standardized differences for African American test-takers were almost the same for both prompt types. The research by Breland et al. (2003) had indicated a possible increase in the standardized difference for African American test-takers when the persuasive prompt was used.

Language analyses

Table 20 shows the differences in mean essay scores for the language groups (i.e., EFL versus ENFL). Table 21 shows the results of the analysis of variance. As expected, the comparison showed a statistically significant difference in mean essay scores between the EFL and ENFL groups. The mean essay score of the EFL group was higher than the score of the ENFL group, with F (1, 3,856) = 25.05, p < .001. Language group membership accounted for 0.6 percent of the variation in essay scores. The effect of essay prompt type on the essay score was not significant. There were no interaction effects between English proficiency and essay prompt type on the essay performance.

Table 20

Mean Essay Scores for Language Group by Essay Prompt Type

	SAT II Prompt				*Persuasive Prompt*			
	N	*Mean*	*SD*	*ES*	*N*	*Mean*	*SD*	*ES*
EFL	1,580	5.96	2.01		1,609	6.08	1.91	
ENFL	325	5.63	2.06	0.17	346	5.57	2.03	0.26

"Effect Size" (ES) here refers to the standardized mean difference: $(\overline{X}_{EFL} - \overline{X}_{ENFL}) / \sqrt{ms_{WG}}$, where ms_{WG} is a mean square within group.

Table 21

Analysis of Variance for Language by Essay Placement

Source of Variation	*Sum of Squares*	df	*Mean Square*	F	p
Language	97.67	1	97.67	25.05	< 0.001
Essay Prompt	0.52	1	0.52	0.13	0.714
Language × Essay Prompt	4.68	1	4.68	1.20	0.273
Within Group	15,033.95	3,856	3.90		
Total Group	15,136.82	3,859			

Table 22

Mean Essay Scores for School Type by Essay Prompt Type

	SAT II Prompt			*Persuasive Prompt*		
	N	*Mean*	*SD*	*N*	*Mean*	*SD*
Public	1,281	5.73	2.00	1,309	5.82	1.91
Private	637	6.25	2.01	663	6.31	1.96

Table 23

Analysis of Variance for School Type by Essay Prompt Type

Source of Variation	*Sum of Squares*	df	*Mean Square*	F	p
School Type	221.70	1	221.70	57.47	< 0.001
Essay Prompt	5.34	1	5.34	1.38	0.240
School Type × Essay Prompt	0.31	1	0.31	0.08	0.776
Within Group	14,990.87	3,886	3.86		
Total Group	15,218.22	3,889			

Table 24

Mean Essay Scores for School Location by Essay Prompt Type

	SAT II Prompt			*Persuasive Prompt*		
	N	*Mean*	*SD*	*N*	*Mean*	*SD*
Large City	519	5.96	2.08	542	6.06	2.05
Medium City	243	5.85	2.04	252	5.93	1.80
Small City	417	5.70	2.05	419	5.70	1.81
Suburban	521	6.19	1.94	535	6.30	2.00
Rural	160	5.41	1.87	165	5.54	1.77

Table 25

Analysis of Variance for School Location by Essay Prompt Type

Source of Variation	*Sum of Squares*	df	*Mean Square*	F	p
School Location	225.14	4	56.28	14.46	< 0.001
Essay Prompt	5.62	1	5.62	1.44	0.230
Location × Essay Prompt	1.94	4	0.49	0.13	0.974
Within Group	14,640.30	3,763	3.89		
Total Group		3,772			

School type analyses

Table 22 displays results of the comparisons of the essay scores between the school type (i.e., public versus private). Mean scores of private school students on both essay prompts were higher than the mean scores of public school students. The analysis of variance shown in Table 23 revealed statistically significant school type differences in students' essay performance, with F (1, 3,886) = 57.47, p = .001.

There were no statistically significant differences in essay score by essay prompt type. There were no interaction effects between school type and essay prompt type on the essay performance.

School location analyses

Table 24 displays the mean essay scores for the five school locations. The results observed in the prompt type study are similar to those observed in the essay placement study. The mean essay score for the suburban area was the highest score among the five school locations. Analysis of variance results in Table 25 confirmed these differences, with F (1, 3,763) = 14.46, p = .001. A follow-up Bonferroni test revealed that suburban school students outperformed the students in other school locations. There was no statistically significant prompt type difference on the essay scores, nor any interaction effect between school location and essay prompt type.

Regression Analyses for School Variables

Results from the analyses regarding school type and school location revealed main effects for both variables. In general, private school students performed better than public school students on the SAT measures. Suburban area students did slightly better than other area students on the SAT measures. Regression analyses were conducted on school type and location simultaneously to investigate school variable effects on students' performance on the SAT. The results showed that only 1 to 2 percent of unique variance in the critical reading, mathematics, writing MC, and essay scores were explained by school type. The results also showed that only 1 percent of unique variance in the SAT measures were explained by school location. The results indicated that although there were group differences among school locations, these differences represented only a small amount of the variability in the test-takers' performance on the SAT. Table 26 shows the unique variance (i.e., semipartial R) in essay score explained by school type and school location, as estimated by the regression analyses.

Table 26

Regression Analyses of School Type and School Location

Test	*Predictors*	*Semipartial* R
Critical Reading	School Type	0.143
	School Location	0.100
	Medium City	-0.005
	Small City	0.017
	Suburban	0.085
	Rural	-0.002
Mathematics	School Type	0.080
	School Location	0.114
	Medium City	0.000
	Small City	0.024
	Suburban	0.097
	Rural	-0.006
Writing MC	School Type	0.152
	School Location	0.089
	Medium City	-0.006
	Small City	0.018
	Suburban	0.073
	Rural	-0.009
Essay	School Type	0.089
	School Location	0.101
	Medium City	-0.012
	Small City	-0.033
	Suburban	0.038
	Rural	-0.044

Discussion

The essay placement study explored whether essay placement may have any effect on test-takers' performance on the SAT. The results of this study indicate that *essay placement* does not impact test-takers' performance on the critical reading, mathematics, and writing MC sections. However, a mean essay score difference was observed between the Essay First and Essay Last group. The effect size was small but potentially meaningful, representing one-third of a standard deviation difference in the performance of the two groups. In general, students tended to do better on the essay when the essay was presented first. This mean essay score difference was evident in the subgroup analyses (i.e., gender, race/ethnicity, language, school type, and school location). All subgroups did better on average when they wrote the essay first than when they wrote the essay last.

The nonsignificant effects of essay placement on the critical reading, mathematics, and writing MC scores found in the present study were consistent with the findings of a previous study (Liu and Oh, 2003), which indicated that essay placement had no effects on the performance of the SAT. The previous study also found nonsignificant effects of essay placement on the essay score, but the statistical test had very low power because of the small sample size. As mentioned by the authors of the previous study, "the strength of the data was adversely affected by the small sample size...because the sample size was small and non-representative, generalizations of the results are limited" (Liu and Oh, 2003). Given the small sample size (i.e., N = 97), the design of the previous study would only detect effect sizes larger than 0.7.

A further comparison of the current study to the Liu and Oh (2003) study may suggest that the essay placement effects reported here could be unique to the field trial. The effect size reported in the Liu and Oh study was very nearly zero (the upper bound of the 95 percent confidence interval for the effect size was .12). By contrast, an effect size of .29 was found in the current study. Whereas the students in the Liu and Oh study were motivated by a desire for practice on the SAT and by monetary incentives, this was not the case in the field trial. Thus, students in the current study possibly experienced decreased motivation toward the end of the testing period, which may have affected their performance on the essay when it was administered last. During an actual administration of the test, such a decline in motivation may not manifest itself. The results are also compatible with the notion that test-takers who wrote the essay last might be slightly fatigued as a result of taking other sections of the test (i.e., critical reading, mathematics, and writing MC). The results of the field trial provide support for placing the essay in the first section of the SAT.

The results from the essay prompt study indicated that *prompt type* did not affect test-takers' essay performance across the studied subgroups. Although differences in mean essay scores were observed for different racial/ethnic, gender, language, school type, and school location groups, no statistically significant group differences were observed between the two prompt types. This finding reinforces the previous finding of Breland et al. (2003) of no significant differences in essay scores between the two prompt types.

The gender differences observed in the prompt study are similar to those observed in previous studies of essay writing performance (Breland et al., 2003; Breland et al., 1995; Bridgeman and McHale, 1996; Schaeffer et al., 2001). In all these previous studies, female test-takers outperformed male test-takers. The racial/ethnic differences observed in the present study are also similar to those of the previous studies (Breland et al., 2003; Breland et al., 1995; Bridgeman et al., 1996; Schaeffer et al., 2001). In general, Asian American and white students performed better than African American and Hispanic students.

The language group differences obtained for the present study were generally consistent with the previous

studies (Breland et al., 2003; Pomplun et al., 1992[4]). That is, typically EFL test-takers did better than ENFL test-takers when writing essays. Note, however, that the Breland et al. study showed a very small difference between those reporting that they learned English first and those who reported learning another language first, although the difference was quite large between those reporting English as their best language and those reporting another language as their best language in Table 20. The SAT II prompt means represent a standardized difference of .17 between EFL and ENFL students, while the persuasive prompt means represent a standardized difference of .26. This finding contrasts with the results of Breland et al., who report standardized differences of .04 on the SAT II prompt and -.01 on the persuasive prompt.

Results from the subgroup analyses regarding school type and school location revealed school type and school location effects. In general, private school students performed better than public school students on the SAT measures. Suburban area students also did better than students from other areas on the SAT measures. The results of the regression analyses indicated that school type and school location had small effects on SAT performance. These results are consistent with patterns found in SAT data for the 2003 College-Bound Seniors (College Board, 2003).

The results of the essay placement and essay prompt studies have important implications for the very large number of test-takers who will take the essay section for the first time on the new SAT. SAT takers will also respond to the persuasive prompt for the first time on the new SAT. Thus, a decision on essay placement and essay prompt type was critical for the prospective college-bound seniors after the year 2005.

The findings in this study indicate that the decision to place the essay prompt at the beginning of the SAT battery is a sound one in terms of potential impact on test-taker performance, for two reasons. First, the study did not reveal differential effects of essay placement on performance on the multiple-choice sections of the test.[5] Second, the study showed superior performance on the essay when the essay was placed first. Thus, placing the essay first would appear to maximize average essay scores.[6]

References

Breland, H., Bonner, M. W., & Kubota, M. (1995). *Factors in performance on brief, impromptu essay examinations* (College Board Research Report No. 95-4, ETS RR No. 95-41). New York: The College Board.

Breland, H., Kubota, M., Nickerson, K., Trapani, C., & Walker, M. (2003). *SAT Writing Prompt Study*. Educational Testing Service. Prepared for the College Board.

Bridgeman, B., & McHale, F. (1996). *Gender and ethnic group differences on the GMAT Analytical Writing Assessment* (ETS RR-96-2). Princeton, NJ: Educational Testing Service.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

College Board (2003). *2003 college-bound seniors: A profile of SAT Program test takers, total group report*. New York: author.

Kirk, R. E. (1995). *Experimental design: procedures for the behavioral sciences* (3rd ed.). St. Paul, MN: Brooks/Cole Publishing Company.

Liu, J., & Feigenbaum, M. (2003). *Prototype analysis of spring 2003 new SAT field trial* (ETS SR-03-69). Princeton, NJ: Educational Testing Service.

Liu, J., & Oh, H.-J. (2003). *A study of fatigue effects from the new SAT*. Educational Testing Service. Prepared for the College Board.

Pomplun, M., Wright, D., Oleka, N., & Sudlow, M. (1992). *An analysis of English composition essay prompts for differential performance* (College Board Research Report No. 92-4). New York: The College Board.

Schaeffer, G. A., Briel, J. B., & Fowles, M. E. (2001). *Psychometric evaluation of the new GRE Writing Assessment* (GRE Report No. 96-11P, ETS RR-01-08). Princeton, NJ: Educational Testing Service.

[4] It is important to note that the sample used by Pomplun et al. was from the Subject Test populations, which might be considerably different from the national random samples of high school students used in Breland et al. and in the present study. It might be expected that Subject Test populations would show larger language-group effects.

[5] One might argue that a statistically significant difference in performance on these sections across essay placement conditions might have called into question the comparability of scores on the current SAT Reasoning Test to scores on the former SAT I: Reasoning Test. However, the new tests have been equated to the old tests, so that both the new and old tests are on the same scale. Even if there were evidence of systematic differences in test scores when an essay was included in the test, the effects of such differences would be removed through the equating process. Only substantial interactions of essay placement with major subgroups of interest might be cause for any great concern. With the exception of the one statistically significant but practically negligible interaction effect of essay position with race/ethnicity on mathematics scores, no such effects were found.

[6] Note that this is not an issue of fairness. All test-takers will receive the essay in the same position on the SAT. In the absence of any interaction effect of essay position with other factors on essay performance, we may conclude that all groups of interest are on average equally advantaged by placing the essay in the first position.

Appendix A: PSAT/NMSQT® Essay Placement Analyses

Table A1 displays the results of the comparisons of the previous PSAT/NMSQT (P/N) critical reading, mathematics, and writing scores between the two essay groups (i.e., Essay First versus Essay Last). The analyses were based on 3,423 out of 4,007 participants who had taken the PSAT/NMSQT in either 2001 or 2002 for the essay placement study. The mean P/N scores of the Essay First group and those of the Essay Last group were very close across critical reading, mathematics, and writing. Table A2 shows the analysis of variance to verify the equivalence of the two groups. At an alpha level of .05, the group differences for the Essay First group and the Essay Last group were not statistically significant on any of the P/N measures: F (1, 3,421) = 0.69, p = .405 for verbal, F (1, 3,421) = 0.34, p = .559 for mathematics, and F (1, 3,421) = 0.34, p = .559 for writing. Thus, the data offer no evidence that the ability level of the test-takers who participated in the study were different. The results also suggest that spiraling procedures for the field trial were successful in ensuring equivalent groups.

Tables A3 and A4 display the results of descriptive statistics and analysis of variance, respectively, for the two gender groups by essay placement. Male test-takers performed better on the verbal and mathematics sections while female test-takers performed better on the writing section. The differences were statistically significant at an alpha level of .05. However, no statistically significant essay placement effects were found for any of the P/N measures by essay placement.

Table A5 shows the descriptive statistics for the P/N scores among the four racial/ethnic groups. The comparisons showed that Asian American and white students performed better than African American and Hispanic students on all three P/N measures. Table A6 presents the analysis of variance for these differences. At an alpha level of .05, the racial/ethnic group differences were statistically significant for all three P/N measures. Table A6 also shows that the group differences for the Essay First group and the Essay Last group were not statistically significant on any of the P/N measures at an alpha level of .05.

Table A1

Mean P/N Scores of Critical Reading, Mathematics, and Writing by Groups of Essay Placement

	Essay First			*Essay Last*			*Effect Size*
	N	***Mean***	***SD***	***N***	***Mean***	***SD***	
P/N Verbal	1,768	48.85	10.72	1,655	49.14	10.05	-0.03
P/N Mathematics	1,768	49.67	10.53	1,655	49.46	10.69	0.02
P/N Writing	1,768	50.07	10.18	1,655	49.87	10.05	0.02

"Effect Size" here refers to the standardized mean difference: $(\overline{X}_{First} - \overline{X}_{Last}) / \sqrt{ms_{WG}}$, where ms_{WG} is a mean square within group.

Table A2

Analysis of Variance for Essay Placement on P/N Scores

Source of Variation	*Sum of Squares*	df	*Mean Square*	F	p
P/N Verbal					
Between Group	75.14	1	75.14	0.69	0.405
Within Group	370,200.52	3,421	108.21		
Total Group	370,275.66	3,422			
P/N Mathematics					
Between G roup	38.40	1	38.40	0.34	0.559
Within Group	385,157.46	3,421	112.59		
Total Group	385,195.85	3,422			
P/N Writing					
Between Group	34.93	1	34.93	0.34	0.559
Within Group	349,896.24	3,421	102.28		
Total Group	349,931.18	3,422			

Table A3

Mean P/N Scores of Critical Reading, Mathematics, and Writing by Gender by Essay Placement

	Essay First			*Essay Last*		
	Male	***Female***	***Total***	***Male***	***Female***	***Total***
Critical Reading						
N	722	1,042	1,764	680	971	1,651
Mean	49.25	48.56	48.84	49.87	48.66	49.16
SD	11.01	10.51	10.72	10.10	10.00	10.06
Effect Size			.07			0.12
Mathematics						
N	722	1,042	1,764	680	971	1,651
Mean	51.42	48.47	49.68	51.56	48.00	49.47
SD	10.78	10.39	10.69	10.73	10.31	10.60
Effect Size			0.28			0.34
Writing						
N	722	1,042	1,764	680	971	1,651
Mean	49.06	50.76	50.06	49.46	50.18	49.88
SD	10.18	9.96	10.05	10.15	10.05	10.11
Effect Size			-0.17			-0.07

"Effect Size" here refers to the standardized mean difference: $(\overline{X}_{Male} - \overline{X}_{Female}) / \sqrt{ms_{WG}}$, where ms_{WG} is a mean square within group.

Table A4

Analysis of Variance for Gender by Essay Placement on P/N Scores

Source of Variation	*Sum of Squares*	df	*Mean Square*	F	p
Critical Reading					
Gender	736.71	1	736.71	6.82	0.009
Essay Placement	106.79	1	106.79	0.99	0.320
Gender × Essay Placement	57.04	1	57.04	0.53	0.468
Within Group	368,519.86	3,411	108.04		
Total Group	369,420.40	3,414			
Mathematics					
Gender	8,754.38	1	8,754.38	79.61	< 0.001
Essay Placement	21.55	1	21.55	0.20	0.658
Gender × Essay Placement	74.70	1	74.70	0.68	0.410
Within Group	375,099.78	3,411	109.97		
Total Group	383,950.41	3,414			
Writing					
Gender	1,200.44	1	1,200.44	11.78	0.001
Essay Placement	6.84	1	6.84	0.06	0.796
Gender × Essay Placement	197.42	1	197.42	1.94	0.164
Within Group	347,544.23	3,411	101.89		
Total Group	433,880.37	3,414			

Table A5

Mean P/N Scores of Critical Reading, Mathematics, and Writing by Ethnic Group

	Essay First				*Essay Last*			
	N	*Mean*	*SD*	*ES*	*N*	*Mean*	*SD*	*ES*
Critical Reading								
Asian Am.	103	48.00	11.92	0.32	98	47.70	11.30	0.36
African Am.	213	41.82	10.19	0.92	212	43.38	10.33	0.78
Hispanic	204	42.82	10.22	0.83	185	43.79	8.51	0.74
White	1,162	51.31	9.68		1,066	51.37	9.19	
Mathematics								
Asian Am.	103	53.76	11.00	-0.18	98	51.78	10.94	-0.03
African Am.	213	41.79	10.08	1.03	212	43.21	10.30	0.83
Hispanic	204	43.44	9.21	0.86	185	44.69	9.27	0.68
White	1,162	52.01	9.63		1,066	51.46	10.18	
Writing								
Asian Am.	103	50.08	11.46	0.13	98	48.30	10.36	0.37
African Am.	213	43.43	8.19	0.91	212	44.29	9.21	0.79
Hispanic	204	45.19	8.47	0.73	185	46.17	8.17	0.59
White	1,162	52.16	9.81		1,066	51.85	9.83	

"Effect Size" (ES) here refers to the standardized mean difference: $(\overline{X}_{White} - \overline{X}_{Nonwhite}) / \sqrt{ms_{WG}}$, where $ms_{WG}G$ is a mean square within group.

Table A6

Analysis of Variance for Race/Ethnicity by Essay Placement on P/N Scores

Source of Variation	*Sum of Squares*	df	*Mean Square*	F	p
Critical Reading					
Race/Ethnicity	42,450.49	3	14,150.16	150.25	< 0.001
Essay Placement	126.19	1	126.19	1.34	0.247
Race/Ethnicity × Placement	261.30	3	87.10	0.93	0.428
Within Group	304,666.23	3,235	105.49		
Total Group	347,504.21	3,242			
Mathematics					
Race/Ethnicity	45,440.69	3	15,146.90	153.70	< 0.001
Essay Placement	0.46	1	0.46	0.01	0.946
Race/Ethnicity × Placement	705.54	3	235.18	2.39	0.067
Within Group	318,805.43	3,235	98.55		
Total Group	364,952.12	3,242			
Writing					
Race/Ethnicity	32,351.56	3	10,783.85	117.30	< 0.001
Essay Placement	1.38	1	1.38	0.02	0.903
Race/Ethnicity × Placement	379.85	3	126.62	1.38	0.248
Within Group	297,414.66	3,235	91.94		
Total Group	330,147.45	3,242			

Appendix B: PSAT/NMSQT Prompt Type Analyses

Table B1 displays the differences in the P/N critical reading, mathematics, and writing scores between the two prompt type groups (i.e., SAT II versus persuasive). The analyses were based on 3,353 out of 3,896 participants who had taken the PSAT/NMSQT in either 2001 or 2002 for the essay prompt study. The mean scores of the SAT II prompt group and those of the persuasive prompt group were very close across critical reading, mathematics, and writing. Table B2 presents the results of analysis of variance. At an alpha level of .05, the group differences for the SAT II prompt and the persuasive prompt were not statistically significant on any of the measures: F (1, 3,351) = 0.09, p = .764 for critical reading, F (1, 3,351) = 0.47, p = .493 for mathematics, and F (1, 3,351) = 0.00, p = .976 for writing. Thus, the results suggest that the test-takers' abilities for the essay prompt type study were equivalent between the two groups.

Table B1

Mean P/N Scores of Critical Reading, Mathematics, and Writing by Groups of Essay Prompt Type

	SAT II Type			*Persuasive Type*			*Effect Size*
	N	*Mean*	*SD*	*N*	*Mean*	*SD*	
P/N Verbal	1,655	49.14	10.05	1,698	49.25	10.11	-0.01
P/N Mathematics	1,655	49.46	10.69	1,698	49.71	10.44	-0.02
P/N Writing	1,655	49.87	10.05	1,698	49.88	10.01	-0.00

"Effect Size" here refers to the standardized mean difference: $(\overline{X}_{SATII} - \overline{X}_{Persuasive}) / \sqrt{ms_{WG}}$, where ms_{WG} is a mean square within group.

Table B2

Analysis of Variance for Essay Prompt Type on P/N Scores

Source of Variation	*Sum of Squares*	df	*Mean Square*	F	p
P/N Verbal					
Between Group	9.19	1	9.19	0.09	0.764
Within Group	340,649.68	3,351	101.66		
Total Group	340,658.87	3,352			
P/N Mathematics					
Between Group	52.53	1	52.53	0.47	0.493
Within Group	373,876.79	3,351	111.57		
Total Group	373,929.32	3,352			
P/N Writing					
Between Group	0.09	1	0.09	0.00	0.976
Within Group	336,901.78	3,351	100.54		
Total Group	336,901.87	3,352			

Appendix C: Sample Essay Prompts

SAT II Type Essay Prompt

Consider carefully the following incomplete statement and the assignment below it. Then plan and write your essay as directed.

> To me, the wisest person is _____.

Assignment: Complete the statement. In an essay, explain your response, using an example (or examples) from history, philosophy, literature, the arts, politics, science and technology, or your experience or observations.

Persuasive Type Essay Prompt

Consider carefully the following excerpt and the assignment below it. Then plan and write an essay that explains your ideas as persuasively as possible. Keep in mind that the support you provide—including reasons and examples—will help make your view convincing to the reader.

> "We need to remember that wisdom is not just about what we think or know, but more importantly, how we act. Simply being smart is not enough. I define wisdom as the application of intelligence and experience toward the attainment of a common good. In other words, the wisest people are those who look out not just for themselves but for others."

Adapted from Robert J. Sternberg, "Teaching for Wisdom in Our Schools."

Assignment: In your opinion, what are the characteristics of a wise person, and who do you think best displays these characteristics? In an essay, support your position using an example (or examples) from literature, the arts, history, current events, politics, science and technology, or your experience or observations.


CollegeBoard
connect to college success™