



Research Report

No. 2004-5

A Study of Fatigue Effects from the New SAT[®]

**Jinghua Liu, Jill R. Allspach,
Miriam Feigenbaum, Hyeon-Joo Oh,
and Nancy Burton**

A Study of Fatigue Effects from the New SAT[®]

Jinghua Liu, Jill R. Allspach, Miriam Feigenbaum,
Hyeon-Joo Oh, and Nancy Burton

Jinghua Liu is a measurement statistician at Educational Testing Service.

Jill R. Allspach is an associate measurement statistician at Educational Testing Service.

Miriam Feigenbaum is a principal statistical associate level II at Educational Testing Service.

Hyeon-Joo Oh is a measurement statistician at Educational Testing Service.

Nancy Burton is a senior research scientist at Educational Testing Service.

Acknowledgments

We are grateful to Michael Walker for his comments on data analyses, and to Rosemary Reshetar and Cathy Wendler for their helpful comments.

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 organization whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 4,500 schools, colleges, universities, and other educational organizations. Each year, the College Board serves over three 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 #040481305) 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.

Copyright © 2004 by College Entrance Examination Board. All rights reserved. College Board, Advanced Placement Program, AP, SAT, and the acorn logo are registered trademarks of the College Entrance Examination Board. Connect to college success, SAT Subject Test, and SAT Reasoning Test are trademarks owned by the College Entrance Examination Board. PSAT/NMSQT is a registered trademark of the College Entrance Examination Board and National Merit Scholarship Corporation. Other products and services may be trademarks of their respective owners. Visit College Board on the Web: www.collegeboard.com.

Printed in the United States of America.

Contents

<i>Abstract</i>	1	<i>Amount of Break Time Desired</i>	7
<i>A Study of Fatigue Effects from the New SAT®</i>	1	<i>Number of Breaks Desired</i>	7
<i>Method</i>	2	<i>Hunger Levels of Test-Takers at the End of the Test</i>	8
<i>Materials</i>	2	<i>Test-Takers' Perception of Whether Hunger Affected Their Performance.</i>	8
<i>Testing Materials</i>	2	<i>Test-Takers' Preferences for Essay Location</i>	8
<i>Score Scales</i>	3	<i>Level of Preoccupation with the Essay Throughout Testing Period</i>	8
<i>Survey Questionnaires</i>	3	<i>Test-Takers' Perception of How Switching Among Content Areas Affected Their Performance</i>	9
<i>Participants</i>	3	<i>Preferences for Essay Location Among Those Who Previously Took the SAT II: Writing Subject Test.</i>	9
<i>Procedure</i>	3	<i>Discussion.</i>	9
<i>Results</i>	3	<i>General Discussion</i>	10
<i>Results of Phase I: Comparisons of Test-Taker Performance</i>	3	<i>References</i>	10
<i>Power Analysis</i>	3	<i>Appendix A: SAT Fatigue Study Survey</i>	11
<i>Effects of Fatigue on SAT Verbal and Math Performance: Essay First Versus No Essay</i>	4	<i>Appendix B: SAT I Fatigue Study Survey</i>	12
<i>Effects of Fatigue on Essay Performance: Essay First Versus Essay Last</i>	4	<i>Tables</i>	
<i>Effects of Motivation on Test Performance</i>	4	1. Mean Scores of Verbal, Math, and Essay Tests by Test-Takers in Three Groups	4
<i>Discussion</i>	5	2. Analysis of Variance of Fatigue Effects on V and M: Essay First Versus No Essay	4
<i>Results of Phase 2: Survey Results</i>	6	3. Analysis of Variance for Fatigue Effects on Essay Scores: Essay First Versus Essay Last	4
<i>Test-Takers' Overall Perceptions of Fatigue at the End of the Test</i>	6	4. Mean SAT Scores from an Actual Administration and from the Study	5
<i>Test-Takers' Perceptions of the Adequacy of the Number and Length of the Breaks Provided</i>	6	5. Analyses of Variance for Mixed Design: Scores Obtained Under Real Administrations Versus Study for Groups with Different Essay Placement.	5
<i>Was the Number of Breaks Sufficient?</i>	6	6. Overall Perception of Fatigue	6
<i>Were the Breaks Placed Appropriately?</i>	7	7. Adequacy of the Number of Breaks Provided.	6

8.	Appropriate Placement of the Breaks Provided	7	14.	Preoccupation with Essay Section Throughout Testing Period	8
9.	Amount of Break Time Desired	7	15.	Perceived Effects of Switching Among Content Areas on How Test-Takers Did on the Test: Test-Takers Who Previously Took the SAT	9
10.	Number of Breaks Desired	7	16.	Essay Location Preference: Test-Takers Who Previously Took the SAT II: Writing Subject Test	9
11.	Overall Level of Hunger	8			
12.	Perceived Negative Effect of Hunger on Performance	8			
13.	Essay Location Preference	8			

Abstract

This study evaluated whether the addition of a writing section to the SAT Reasoning Test™ (referred to as the SAT® in this study) would impact test-taker performance because of fatigue caused by increased test length. The study also investigated test-takers' subjective feelings of fatigue. Ninety-seven test-takers were randomly assigned to three groups: the first group took a current SAT with no essay; the second group took a pseudo new SAT composed of the current SAT plus an essay, with the essay appearing in the first section of the test; and the third group also took the pseudo new SAT with an essay, but with the essay in the last section. Test-taker performance on the verbal and math sections and the essay was then evaluated and compared. The results indicated that while the extended testing time for the new SAT may cause test-takers to feel fatigued, fatigue did not affect test-taker performance.

Keywords: SAT, fatigue, essay placement

A Study of Fatigue Effects from the New SAT®

The effects of fatigue in test-taking have been studied for over a century, generating many discursive articles and reports. The concept of fatigue has been applied to a variety of phenomena: subjective feelings, alternation of behavioral activity, changes of psychological states such as anxiety and/or motivation, and even biochemical changes such as overall arousal level. For example, Bartley and Chute (1947) viewed fatigue as a response involving aversion and a feeling of unwillingness and inadequacy for activity; Spaeth defined fatigue as the decreased capacity to do work as a direct result of having worked (1920). A broad and comprehensive concept of fatigue has been approached from a variety of ways including subjective feelings of tiredness, organic-chemical or physical changes, and changes in the quantity or quality of work output (Bills, 1937; Starch, Stanton, and Koerth, 1936).

The literature shows that when mental tasks are performed, the primary fatigue factors include time of day, testing time, type of task, personal states such as anxiety or motivation, and external circumstances. A review of literature on diurnal variations in performance (Smith, 1989) showed that overall arousal increases over the day from 8 a.m. to about 10 p.m. In general, increased arousal is associated with increased performance. Krueger's (1994) review of performance throughout

the day identifies the hours of 7:30 a.m. to 1:00 p.m. (if lunch is not taken) as being, in general, a period of particularly good performance. Other researchers found that repetitive, homogeneous work is more fatiguing than a heterogeneous mix of tasks (e.g., Mednick, Nakayama, Cantero, Atienza, Levin, Pathak, and Stickgold, 2002; Newburger, 1942; Thorndike, 1921). A review of the literature on fatigue in medical residents, who at the time averaged 36 consecutive hours on the job and 100-hour work weeks, concluded that the reported effects on mood and performance of tasks requiring sustained vigilance were of practical significance (Samkoff and Jacques, 1991).

Several experiments have been conducted to explore the effects of fatigue on test performance. Carmichael and Dearborn (1947) had high school and Harvard College students read for six hours, with a large number of multiple-choice questions interspersed and a multiple-choice comprehension test immediately after. Students' level of comprehension did not change over the six hours, and their reports of fatigue were unrelated to performance. Tucker (1947) studied students who took College Board afternoon Achievement Tests only versus those who had also taken the SAT in the morning. He found no performance difference between the groups. Poffenberger (1928) studied subjects (age unstated) who worked continuously for 5½ hours on one of four kinds of tasks (addition, sentence completion, intelligence test items, or judging compositions). Feelings consistently declined, but only arithmetic performance declined; performance increased over the time on the intelligence test tasks.

All of the above studies focused on tests with durations of more than five hours. Researchers have also conducted studies on tests less than five hours. Noll (1932) studied teachers-college students planning to transfer to a state university after one or two years of study. They took an equation completion test before and after a college-ability test. He reported that their efficiency appeared higher after the three-hour ability test than before. However, an attempt to increase test motivation for the group who took the equation test after, by stating that their performance on the first-taken ability test would depend upon their performance on the second-taken equation test, seemed to have a negative effect on performance. Massey (1977) studied experimental forms of the British General Certificate of Education (GCE) chemistry test that were 1¼ hours long. They were administered in original and reverse orders in unspeeded conditions. The investigator found no evidence of performance decline at the end of the test. Mollenkopf (1950) did a similar experiment with verbal analogy and mathematical items. He found a fatigue effect for analogy items, but none for mathematical items.

Overall, there is some indication that there is more fatigue associated with simple tasks than with complex tasks, and with low-stakes tasks compared with high-stakes tasks. Tests of intelligence, reading, and mathematics all seemed resistant to fatigue, at least for periods of up to five or six hours. While most test-takers report feelings of fatigue, these perceptions seem to have little relation to performance, again for periods of five to six hours. However, test-takers' perception that longer tests are more tiring is an issue in itself. Physiological studies suggested that breaks of five minutes per half hour or ten minutes per hour would be beneficial. Other research, however, indicated that varying the tasks might partially reduce the need for such frequent breaks.

The research on fatigue effects has important implications for the changes to the SAT Reasoning Test (referred to as the SAT in this study). To strengthen the alignment of the SAT to curriculum and instructional practices in high schools and colleges, changes are being made to the test. A new writing section will be added to the SAT test battery, including multiple-choice questions and a student-written essay. The critical reading section, currently known as the verbal section, will have analogy items eliminated, and short reading passages will be added along with the existing long reading passages. Math content will be expanded, and quantitative comparison questions will be eliminated. With the addition of a writing section, total testing time will be increased to 3 hours and 45 minutes, as compared to the current 3-hour test.¹ If one includes time for material distribution and collection, instruction, and breaks, the total administration time will be more than four hours.

This extended testing time may cause a certain amount of fatigue. In addition, dropping two item types will make the test more homogeneous, which may result in greater fatigue. Conversely, the possible fatigue induced by dropping item types may be counteracted by the addition of an entirely new writing measure. Thus, the changes associated with the new test may result in several countervailing influences on test-taker fatigue. The purpose of this study, therefore, was to gather empirical evidence about the possible effects of fatigue on the new SAT.

In the summer of 2002, Educational Testing Service (ETS) undertook a fatigue study on behalf of the College Board. The study explored the effect of increased testing time on performance and test-taker perceptions of fatigue resulting from the increased time. Specifically, the study investigated the following questions:

1. Were SAT verbal (V) and math (M) scores different for test-takers who did not write an essay

from the scores of those who did? Interactions among test content were not anticipated, nor are they of interest here. Rather, the real issue was whether the increased length of the test would adversely affect performance. It was hypothesized that the introduction of the essay would not significantly alter performance when total testing time increased. This would be consistent with the previous research showing that test-takers' feelings of fatigue seem to have little relation to their performance.

2. Were there differences in essay performance between the group that had an SAT with the essay section first and the group that had the test with the essay last? Both groups had identical testing time, so the critical issue was whether essay placement could affect performance. Those who wrote the essay last could show a decreased level of performance on the essay because of fatigue. It was anticipated that the fatigue would not significantly alter test-taker performance on the essay.
3. Looking at students who would have or had V and M scores from actual SAT administrations, was there a difference in their performance in the actual administration vis-à-vis the study? This question assessed issues of motivation and other factors that could render the study results nonrepresentative of the data from a general administration.
4. What were the test-takers' overall perceptions of fatigue at the end of the test?

Method

Materials

Testing Materials

Three test books were assembled for three different groups: (1) The "No Essay" group had a current SAT book with no essay; (2) the "Essay First" group had an identical current SAT book plus an essay, with the essay appearing in the first section of the test; and (3) the "Essay Last" group had the same book as the "Essay First" group, except that the essay was given in the last section. When the study was conducted, there were no actual new SAT books assembled—we simply added an essay to the current SAT book and replaced the variable section in the current book with a writing multiple-choice section to simulate the new test. Test-takers who took the essay were tested for 25 minutes longer.

¹ When the study was conducted, the configuration of the new SAT writing section included a 25-minute essay and a 25-minute multiple-choice section. The writing prototype was revised later to increase the length of the multiple-choice section to 35 minutes, in an effort to increase test reliability, resulting in a total testing time of 3 hours, 45 minutes.

Score Scales

For V and M, scores were reported on the 200-to-800 SAT scale in 10-point units (200, 210...790, 800). Essay scores were reported on a 20-to-80 scale in 1-point units, in alignment with the SAT Writing Subject Test (now called the SAT Subject Test™ in Writing). Writing multiple-choice questions were presented only to simulate a new SAT, and did not contribute to any of the scores created.

Survey Questionnaires

The survey asked participants to indicate their perceptions of fatigue and hunger; reactions to the number and placement of breaks; preferences for essay location; and perceived effects of switching among the three content areas. Respondents were also asked to compare this testing experience with their previous experiences taking the SAT or the SAT Writing Subject Test. The questionnaire in Appendix A was used by participants who took the test with an essay; the questionnaire in Appendix B by those whose test did not include an essay.

Participants

Students who lived within 40 miles of ETS and who were registered for the October 2002 SAT administration as of July 2002 were sent invitation letters and/or e-mails describing the study. Ultimately, a total of 97 students participated in the study: 45 males and 52 females. Because the sample used was small, results should be considered preliminary. In addition, the small sample sizes did not allow an exploration of subgroup differences. Information on race/ethnicity was not collected.

As an incentive, all participants were given their test book, a copy of their answer sheet, and the answer key and conversion table for their test form upon completion of the test. This allowed them to score and analyze their work immediately. The essays were scored later by ETS, and unofficial score reports were sent to the test-takers by mail. Additionally, participants were told when they signed up for the study that they would receive a \$50 American Express Gift Certificate if their scores on this pilot test were within expected ranges. All participants received certificates upon completion of the test with the explanation that ETS believed that they had made their best effort and their willingness to participate was appreciated.

Procedure

All testing was done at ETS in August 2002. Half of the participants took the test on Day 1, and the other half on Day 2. The testing occurred over the course of two days in order to accommodate more participants within the testing rooms, and to accommodate their summer schedules with more ease and flexibility. All participants were randomly assigned to three groups prior to the

testing; on the day of the testing, they went to their assigned room. Each group was tested in a separate room, so participants were unaware that testing conditions for others were different.

The study involved two phases. In the first phase, each group of test-takers took the test. In the second phase, test-takers completed a survey upon completion of the test. All participants were allowed to take several scheduled breaks during the test. The “No Essay” group took one 5-minute break and one 1-minute break. The “Essay First” and the “Essay Last” groups had two 5-minute breaks.

ETS staff observed participants and videotaped them throughout the testing so that their actions could be assessed at a later date, if necessary.

Results

The results section is organized as follows. First, we present the results of the first phase of the study: the comparisons of test-taker performance on all three measures. Second, we discuss the results from the second phase: the survey results.

Results of Phase 1: Comparisons of Test-Taker Performance

Power Analysis

The sample sizes for the three groups in the study ranged from 31 to 35. With a significance level of .05, the study design would allow detection of mean scaled score differences among population groups of 70 points (on the 200-to-800 scale) with a power of .80 (i.e., if the largest mean difference among the population groups is as large as 70 scaled score points, then we would obtain a statistically significant result from an analysis of variance 80 percent of the time). We would be able to detect mean scaled score differences of 80 points among population groups with a power of .90. Given that the population standard deviations of the verbal, math, and writing sections are close to 110 scaled score points, these target differences represent standardized differences between groups of .64 and .73, respectively.

These power analyses reinforce the likelihood, given the small sample sizes, of not obtaining statistically significant results in this study unless the population effects are fairly large. Given this information, we should not take lack of statistical significance as an indication that no population differences are present, or even that no large differences exist. All statistical tests are augmented by effect size measures, which reflect more adequately than probability values the importance of the results of the study. Because most people affiliated with college

admissions testing are familiar with the 200-to-800 SAT scale, such that they can readily interpret scaled score differences, effect sizes will generally be reported as 90 percent confidence intervals for the population mean differences.

Effects of Fatigue on SAT Verbal and Math Performance: Essay First Versus No Essay

To detect the fatigue effects on V and M of taking a longer test, we compared performance on V and M between the “Essay First” group and the “No Essay” group.

The first set of findings are given in Table 1, which presents mean scores on V and M for each group. As expected, the addition of an essay did not seem to alter test performance, as the mean scores for V and M for the “Essay First” group were not lower than the scores of the “No Essay” group. In fact, test-takers in the “Essay First” group performed better than those in the “No Essay” group, even though the former tested for a longer time. Effect size estimates indicated that the mean population differences could be anywhere from -2 to 84 scaled score points on Verbal, and from -1 to 101 on Math.

Further analyses were performed to explore whether or not those differences were statistically significant. Table 2 shows the result of the analysis of variance comparing performance differences on V and M between the “Essay First” group and the “No Essay” group. At an alpha level of .05, group differences for both V and M were not statistically significant: $F(1, 64) = 2.44, p = .12$, and $F(1, 64) = 2.71, p = .11$, respectively.

Effects of Fatigue on Essay Performance: Essay First Versus Essay Last

Using performance data from when essays were given prior to V and M (essay first) versus when essays were given after V and M (essay last), allows us to explore possible fatigue effects on the essay. As can be seen in Table 1, the mean essay score of the “Essay First” group and that of the “Essay Last” group were both 53.7. The 90 percent confidence interval for the population mean difference on the essay is 15 to 22 points.

Table 3 presents the analysis of variance for essay scores. At an alpha level of .05, the difference between the “Essay First” group and the “Essay Last” group was not statistically significant: $F(1, 64) = 0.00, p = .99$. Thus, the data offered no evidence that fatigue affected essay scores.

Effects of Motivation on Test Performance

It was difficult to be certain whether or not the participants in this study were properly motivated to take the test. One way to detect motivation effects is to compare test-takers’ performance in the study with their performance in a national SAT administration. Of the 97 participants, 23 had not taken any national SAT

Table 1

Mean Scores of Verbal, Math, and Essay Tests by Test-Takers in Three Groups			
	Essay First	Essay Last	No Essay
Verbal			
N	35	31	31
Mean	560	562	519
SD	109	95	102
Math			
N	35	31	31
Mean	582	573	532
SD	116	90	131
Essay			
N	35	31	
Mean	53.7	53.7	
SD	9.4	9.0	

Table 2

Analysis of Variance of Fatigue Effects on V and M: Essay First Versus No Essay

Source of Variation	Sum of Squares	df	Mean Square	F	p
Verbal					
Between group	27207.65	1	27207.65	2.44	0.12
Within group	714168.11	64	11158.88		
Total	741375.76	65			
Math					
Between group	41143.95	1	41143.95	2.71	0.11
Within group	972159.08	64	15189.99		
Total	1013303.03	65			

Table 3

Analysis of Variance for Fatigue Effects on Essay Scores: Essay First Versus Essay Last

Source of Variation	Sum of Squares	df	Mean Square	F	p
Between group	0.01	1	0.01	0.00	0.99
Within group	5467.93	64	85.44		
Total	5467.94	65			

administration as of January 2003. Seventy participants took the test in October 2002, two in November 2002, and two in December 2002. If test-takers took the SAT at more than one fall administration, only the scores from the first administration were used in subsequent data analyses.

However, when we selected those participants who had operational SAT scores from the three randomly assigned groups, there was some possibility that the equivalence of the groups might be destroyed. We need to keep this in mind when we interpret the results in the following analyses.

Table 4 displays descriptive statistics for the participants who had taken the SAT at an actual administration, compared with their scores from the study. Note that the sample sizes were now reduced across the three groups. As shown in Table 4, all participants scored higher on the actual test than they did in the study, which was reasonable given the different motivational levels likely experienced in a real high-stakes testing situation versus a study.

The performance differences between an actual administration and the study were compared using a mixed design analysis of variance, in which group (essay first, essay last, and no essay) was a between-group factor, and score (administration versus study) was a within-group factor. Table 5 presents the results of comparisons. For Verbal, there was no significant group difference with respect to essay appearance, $F(2, 69) = .97, p = .39$. Second, there was a significant difference between the scores from an actual administration and the scores from the study, $F(1, 69) = 4.19, p < .05$. Participants did better in actual administrations than in the study. Third, there was no significant interaction between group and score. The results for Math were similar to those for Verbal: there was no significant group difference with respect to essay appearance, and

Table 4

Mean SAT Scores from an Actual Administration and from the Study

	Essay First		Essay Last		No Essay	
	Admin.	Study	Admin.	Study	Admin.	Study
Verbal						
N	27	27	23	23	22	22
Mean	571	567	567	560	543	519
SD	116	122	99	91	84	96
Math						
N	27	27	23	23	22	22
Mean	601	590	597	579	565	535
SD	107	122	94	97	133	138

Table 5

Analyses of Variance for Mixed Design: Scores Obtained Under Real Administrations Versus Study for Groups with Different Essay Placement

Source of Variation	Sum of Squares	df	Mean Square	F	p
Verbal					
Between Group					
Group	39378.59	2	19689.29	.97	.39
Error	1406687.38	69	20386.77		
Within group					
Score	4334.03	1	4334.03	4.19	.04*
Group × Score	2763.27	2	1381.63	1.34	.27
Error	71352.70	69	1034.10		
Math					
Between Group					
Group	56382.55	2	28191.28	1.10	.34
Error	1766511.20	69	25601.61		
Within group					
Score	13417.36	1	13417.36	11.18	.00**
Group × Score	2033.28	2	1016.64	.85	.43
Error	82799.36	69	1199.99		

* $p < .05$; ** $p < .01$

there was no significant interaction between group and score. However, participants did significantly better in actual administrations. Even though there was no significant interaction detected, it seemed that the discrepancy between actual scores and study scores for the “No Essay” group was larger than for the other two groups. It may be that the test-takers of the “No Essay” group were motivated to participate in the study because they anticipated they would be taking an essay, and therefore became less motivated when they found out that there was no essay in their test book.

Discussion

This study explored whether or not fatigue may have any detrimental effects on test-taker performance on the new SAT due to increased test length. The results indicate no problems related to the increased test length that would seriously hamper test-taker performance. In general, participants who tested for a longer period did not show evidence of a decline on V and M, relative to other test-takers in the study who were tested for a shorter time (Essay First versus No Essay), and participants who wrote the essay last did not show a decrease in essay score

compared with test-takers in the “Essay First” group. This was consistent with previous research that found that fatigue does not have a great effect on an individual’s performance on tests or other mental tasks during an extended period (Tucker, 1947; Wohlhueter, 1966).

The results also suggest that motivational factors may play a role in testing situations. The participants performed significantly better in an SAT operational administration than they did in the study. The discrepancy between actual scores and the study scores for the “No Essay” group seemed to be larger than for the other two groups, even though there was no significant difference detected. This might be explained by the hypothesis that strong motivational factors will counteract any detrimental effects that fatigue may have upon an individual’s performance (Wohlhueter, 1966), since the participants in the two essay groups might have been more motivated when they took the test with an essay.

In addition to motivational issues, performance variations between different groups might be attributed to the homogeneity or heterogeneity of tasks. Researchers found that repetitive stimulation of a particular mental task—homogeneity—produces more fatigue than a heterogeneous task (Newburger, 1942; Thorndike, 1921). Although both the existing SAT and the proposed new SAT are homogeneous from the point of view that they both measure reasoning skills, they are indeed heterogeneous in terms of item types and content. For example, the existing SAT V and M both have three item types requiring different mental functions, crossing different difficulty levels, and covering a variety of content. In this study, the test becomes even more heterogeneous since an entirely new Writing measure is added. Writing an essay, which is a different cognitive task from answering multiple-choice items, and switching among three different subjects (Verbal, Math, and Writing) may have contributed to reducing the possible effects of test-taker fatigue.

Results of Phase 2: Survey Results

Survey results were compared among the three groups. The “Essay First” group included 35 test-takers (17 males and 18 females); the “Essay Last” group, 31 test-takers (9 males and 22 females); and the “No Essay” group, 31 test-takers (19 males and 12 females). Because the sample was small, the results should be considered preliminary. Again, the small sample sizes did not allow for an exploration of subgroup differences.

Test-Takers’ Overall Perceptions of Fatigue at the End of the Test

As shown in Table 6, the percentage of responses in each category was similar. The majority of test-takers were

Table 6

Overall Perception of Fatigue						
	Essay First		Essay Last		No Essay	
	N	%	N	%	N	%
Very tired	9	26	10	32	6	19
Somewhat tired	22	63	21	68	23	74
Not at all tired	4	11	0	0	2	6

Note: Summation of percentages may not equal 100 percent due to rounding.

either very or somewhat tired. All of the test-takers in the “Essay Last” group and approximately 90 percent of the test-takers in each of the other groups indicated some degree of fatigue.

Test-Takers’ Perceptions of the Adequacy of the Number and Length of the Breaks Provided

During the test administration, the groups with an essay received two 5-minute breaks, for a total of 10 minutes, while the group without an essay received one 5-minute break and one 1-minute break, for a total of 6 minutes. The survey asked test-takers whether they thought the number of breaks was sufficient, whether the breaks were placed at the appropriate sections, and the number and length of breaks desired.

Was the Number of Breaks Sufficient?

Table 7 provides the results of the test-takers’ perceptions of the number of breaks provided during the test. Most test-takers in the “Essay First” (69 percent) and “Essay Last” (52 percent) groups indicated there were enough breaks. In the “Essay First” group, only 14 percent felt there were not enough breaks, while 17 percent were unsure. In the “Essay Last” group, 29 percent felt there were not enough breaks, and 19 percent indicated that they were unsure. The “No Essay” group was slightly different, with 55 percent indicating that there were not enough breaks. Only 32 percent of test-takers felt enough breaks were provided, and 13 percent indicated that they were unsure.

Table 7

Adequacy of the Number of Breaks Provided						
	Essay First		Essay Last		No Essay	
	N	%	N	%	N	%
Enough	24	69	16	52	10	32
Not enough	5	14	9	29	17	55
Don’t know	6	17	6	19	4	13

Were the Breaks Placed Appropriately?

Table 8 provides results regarding placement of the breaks. All three groups agreed that the breaks were placed at the appropriate sections in the test: specifically, 77 percent in the “Essay First” and “Essay Last” groups and 55 percent in the “No Essay” group. However, greater differences were seen in the percentage of test-takers who felt the breaks were placed inappropriately. In the “Essay First” group, 17 percent indicated that the breaks were placed inappropriately, and 6 percent indicated that they were unsure. In the “Essay Last” group, 6 percent indicated that the breaks were placed inappropriately, and 16 percent indicated that they were unsure. More test-takers in the “No Essay” group (39 percent) felt that the breaks were placed inappropriately, and 6 percent indicated they were unsure.

Amount of Break Time Desired.

Test-takers who indicated either that there were not enough breaks or that the breaks were not placed appropriately were asked to answer an additional question about the length of desired breaks. Table 9 presents results regarding the amount of break time desired. The majority of test-takers in the “Essay First” (80 percent) and “Essay Last” (71 percent) groups indicated that the break time provided (a total of 10 minutes) was sufficient. The next most popular response favored an increase in the amount of break time: 14 percent in the “Essay First” group and 29 percent in the “Essay Last” group indicated a desire for 11 or more minutes of break time. In the “No Essay” group, 48 percent indicated that they were satisfied with the amount of break time provided (a total of 6 minutes), but 49 percent indicated that increased break time was desired (of these test-takers, 36 percent indicated a desire for between 7 and 10 minutes of break time).

Table 9 also provides statistics on the amount of break time desired. For both of the essay groups, the mean value was similar to the actual break time provided (10 minutes), though the mean for the “Essay Last” group was slightly higher at 11 minutes. For the “No Essay” group, the mean was much larger at 8.6, in comparison with the six-minute break time provided. This was probably due to an extreme value. The “Essay Last” group was the only one in which the minimum amount of break time desired was equivalent to the amount of break time provided (10

Table 8

Appropriate Placement of the Breaks Provided

	Essay First		Essay Last		No Essay	
	N	%	N	%	N	%
Yes	27	77	24	77	17	55
No	6	17	2	6	12	39
Don't know	2	6	5	16	2	6

Note: Summation of percentages may not equal 100 percent due to rounding.

Table 9

Amount of Break Time Desired

Minutes	Essay First		Essay Last		No Essay	
	N	%	N	%	N	%
5 or less	1	3	0	0	1	3
6	1	3	0	0	15	48
7 to 9	0	0	0	0	4	13
10	28	80	22	71	7	23
11 or more	5	14	9	29	4	13
N	35		31		31	
Mean	10.3		11.0		8.6	
Minimum	2		10		1	
Maximum	20		18		30	

minutes). In the “Essay First” and “No Essay” groups, the minimums were 2 minutes and 1 minute, respectively. The maximums for both of the essay groups were similar, with the “Essay First” group indicating 20 minutes and the “Essay Last” group 18 minutes. However, the “No Essay” group requested a maximum of 30 minutes, which was the highest amount suggested in the survey. Only one test-taker gave this response, which could be deemed an outlier, as the next highest response was 18 minutes.

Number of Breaks Desired.

Table 10 presents results regarding the number of breaks desired. Results showed that, for all groups, the majority of test-takers felt two breaks were desirable (86 percent in the “Essay First” group and 71 percent in both the “Essay Last” and “No Essay” groups), while some test-takers indicated a desire for three or more breaks. The means for each of the three groups indicated that test-takers desired slightly more than two breaks: 2.4 for the “Essay First”

Table 10

Number of Breaks Desired

Number of Breaks Desired	Essay First		Essay Last		No Essay	
	N	%	N	%	N	%
1	0	0	0	0	1	3
2	30	86	22	71	22	71
3	1	3	4	13	4	13
4	2	6	3	10	0	0
5	0	0	0	0	2	6
6	2	6	2	6	2	6
N	35		31		31	
Mean	2.4		2.6		2.6	
Minimum	2		2		1	
Maximum	6		6		6	

Note: Summation of percentages may not equal 100 percent due to rounding.

group, 2.6 for the “Essay Last” group, and 2.6 for the “No Essay” group. Overall, results show that the number of desired breaks was between two and three.

Hunger Levels of Test-Takers at the End of the Test

Information on the hunger question appears in Table 11. The majority of test-takers in the essay groups indicated that they were very hungry: 51 percent in the “Essay First” group and 55 percent in the “Essay Last” group. However, in the “No Essay” group, the majority of test-takers (65 percent) indicated that they were somewhat hungry. Overall, 94 percent of test-takers in the “Essay First” group indicated some degree of hunger, while in the “Essay Last” and “No Essay” groups, 100 percent indicated some degree of hunger. It should be noted that participants were required to report to the testing site at 8:15 a.m., and the approximate ending time was 1:00 p.m., similar to what will occur when actual test centers are administering the new SAT. In light of this, it is not surprising that nearly all of the participants indicated some degree of hunger.

Test-Takers’ Perception of Whether Hunger Affected Their Performance

More than half of the test-takers across all groups reported that hunger had a somewhat negative effect on their performance. As shown in Table 12, responses were similar for all of the groups: In the “Essay First” group, 66 percent indicated that hunger affected their performance; in the “Essay Last” group, the percentage was 64 percent; and in the “No Essay” group, 58 percent. The percentage of test-takers who indicated that hunger had not affected their performance was highest in the “No Essay” group (42 percent). In the “Essay First” and “Essay Last” groups, the percentages were similar at 34 percent and 35 percent, respectively.

Test-Takers’ Preferences for Essay Location

Participants in the “Essay First” and “Essay Last” groups were asked their preferences for the location of the essay, and the results are presented in Table 13. Test-takers’ preferences mimicked the position of the essay used in their assigned testing groups: 77 percent of the “Essay First” test-takers indicated they preferred the essay at the beginning of the test, and 55 percent of those in the “Essay Last” group indicated they preferred the essay at the end of the test.

Level of Preoccupation with the Essay Throughout Testing Period

Table 14 shows that approximately 68 percent of test-takers in the “Essay First” group and 61 percent in the “Essay Last”

Table 11

Overall Level of Hunger						
	Essay First		Essay Last		No Essay	
	N	%	N	%	N	%
Very hungry	18	51	17	55	11	35
Somewhat hungry	15	43	14	45	20	65
Not at all hungry	2	6	0	0	0	0

Table 12

Perceived Negative Effect of Hunger on Performance						
	Essay First		Essay Last		No Essay	
	N	%	N	%	N	%
Yes, very much	3	9	6	19	3	10
Yes, somewhat	20	57	14	45	15	48
No	12	34	11	35	13	42

Note: Summation of percentages may not equal 100 percent due to rounding.

Table 13

Essay Location Preference				
	Essay First		Essay Last	
	N	%	N	%
At the beginning of the test	27	77	5	16
In the middle of the test	4	11	6	19
At the end of the test	1	3	17	55
I don’t know/I don’t care	3	9	2	6
Either at the beginning or in the middle of the test	0	0	1	3

Note: Summation of percentages may not equal 100 percent due to rounding.

Table 14

Preoccupation with Essay Section Throughout Testing Period				
	Essay First		Essay Last	
	N	%	N	%
Agree completely	1	3	2	6
Agree somewhat	6	17	3	10
Neither agree nor disagree	4	11	7	23
Disagree somewhat	5	14	6	19
Disagree completely	19	54	13	42

Note: Summation of percentages may not equal 100 percent due to rounding.

group indicated they were not concerned about the essay while working on other parts of the test. Only 20 percent of test-takers in the “Essay First” group and 16 percent in the “Essay Last” group indicated some preoccupation with the essay section during the rest of the test.

Test-Takers’ Perception of How Switching Among Content Areas Affected Their Performance

Participants who had taken the SAT before were asked if they felt that switching among three content areas (V, M, and W) rather than between two (V and M) helped or hurt their overall performance. As can be seen in Table 15, the most common response for the “Essay First” and “No Essay” groups was “had no effect” (40 percent of the “Essay First” group and 38 percent of the “No Essay” group gave this reply). In each of those groups, the number of test-takers indicating “sort of hurt” and “sort of helped” was similar. Responses from test-takers in the “Essay Last” group were virtually equal among “sort of hurt” (32 percent), “had no effect” (32 percent) and “sort of helped” (28 percent). Overall, the majority of test-takers across all groups felt that switching among sections had no effect, or a positive impact, on their test performance.

Preferences for Essay Location Among Those Who Previously Took the SAT II: Writing Subject Test

Virtually all of the participants in the essay groups who had previously taken an essay as part of an SAT II: Writing Subject Test administration indicated that the essay was at the beginning of the test when they took that test. This was a good indication that test-takers responded sincerely to the survey, as the essay is always first for the SAT II: Writing Subject Test. For the test-takers in the “Essay Last” group, 56 percent preferred to take the essay last, 33 percent preferred to take it first, and 11 percent did not have an opinion. The results are shown in Table 16.

Table 15

Perceived Effects of Switching Among Content Areas on How Test-Takers Did on the Test: Test-Takers Who Previously Took the SAT

	Essay First		Essay Last		No Essay	
	N	%	N	%	N	%
Really hurt how I did on the test	0	0	1	4	0	0
Sort of hurt how I did on the test	8	27	8	32	7	29
Had no effect on how I did on the test	12	40	8	32	9	38
Sort of helped how I did on the test	7	23	7	28	7	29
Really helped how I did on the test	1	3	0	0	0	0

Note: Summation of percentages may not equal 100 percent due to rounding and/or the inclusion of missing data into the percentages.

Table 16

Essay Location Preference: Test-Takers Who Previously Took the SAT II: Writing Subject Test

	Essay First		Essay Last	
	N	%	N	%
The location of the essay was the same today and on my previous SAT II: Writing Test	12	86	0	0
I preferred the location of the essay in today’s test	0	0	5	56
I preferred the location of the essay in my previous SAT II: Writing Test	1	7	3	33
I don’t know/I don’t care where the essay is located	0	0	1	11

Note: Summation of percentages may not equal 100 percent due to rounding and/or the inclusion of missing data into the percentages.

Discussion

The results of the survey indicate that, in general, test-takers do not appear to feel that the change from the current 3-hour test to the new 3-hour and 45-minute test involves detrimental fatigue. Even though test-takers appeared to be slightly more fatigued when they took the new, longer test, they did not feel that fatigue negatively affected their performance. Also, test-takers felt that switching among the three content areas had no effect on their performance or even helped.

This finding is consistent with previous research that changing the nature of tasks performed seems to help reduce the perceived fatigue experienced by participants (Newburger, 1942). Newburger made a similar point in a study examining the mental fatigue of college test-takers related to the difficulty and homogeneity of tasks. The findings suggested that the difficulty of tasks is not as important as the variety of tasks. In the groups where the mental tasks were difficult but heterogeneous, there was less mental fatigue observed than in cases where the tasks were difficult and homogeneous. In the current study, when test-takers were engaged in a wider variety of tasks, including writing an essay and answering multiple-choice questions in all of the Verbal, Math, and Writing content areas, they seemed to experience a reduced perception of fatigue.

When we questioned students who had an essay about their preferences for essay location, the answers were dependent upon the location of the essay in the test they took. Generally they did not seem to feel that essay placement had a great effect on their performance. However, it should be noted that students writing the essay last did report hunger and fatigue more frequently, compared with those who wrote the essay first.

In summary, it appears that students taking the new SAT do not feel dramatically increased levels of fatigue or hunger compared with those taking the current form of the test. Further, test-takers do not seem to feel that the placement of the essay has a great impact on their performance. Finally, it must be noted that because the sample size was small and not representative of the College-Bound Senior cohort, these results provide some guidance, but further study is warranted.

General Discussion

In the current study, we were concerned with subjective feelings of fatigue resulting from prolonged mental tasks on the new SAT, and the effects of fatigue, if there are any, on the performance of the test-takers. The preliminary results of the study indicated that fatigue would not alter test-taker performance on the pseudo new SAT, even with an increase in the total testing time. Even though test-takers appeared to be slightly more fatigued when they took the longer test, they did not feel that fatigue negatively affected their performance. Placement of the essay did not affect test-taker performance, and generally test-takers did not seem to feel that essay placement had a great effect on their performance.

This study was designed to gather preliminary data on performance and fatigue level when students take the longer new SAT. Due to scheduling limitations and the difficulty of recruiting participants during the summer, only a small sample size was available, and testing materials only approximated the new SAT test. Consequently, the three groups might not necessarily be equivalent in terms of overall ability level, even though the samples were formed by random assignment. In addition, the strength of the data was adversely affected by the small sample size. Finally, because the sample size was small and nonrepresentative, generalizations of the results are limited.

This study is a pilot study and should be followed up by collecting and analyzing data from intact tests given to a larger group at actual administrations. It is also recommended that an analysis of the data to determine if fatigue affects test-takers' performance also take into consideration factors such as subjective feelings of fatigue and individual motivations for taking the test, thus investigating the possible interaction between test-taker performance and subjective feelings of fatigue. The information derived from such analyses should give us a clearer understanding of the effect fatigue has upon the performance of test-takers taking the new SAT.

References

- Bartley, S. H., & Chute, E. (1947). *Fatigue and impairment in man*. New York: McGraw-Hill.
- Bills, A. G. (1937). Fatigue in mental work. *Physiological Review*, 17, 436–453.
- Carmichael, L., & Dearborn, W. F. (1947). *Reading and visual fatigue*. Cambridge, MA: Houghton Mifflin.
- Krueger, G. P. (1994). Fatigue, performance, and medical error. In M. S. Bogner (Ed.), *Human error in medicine*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Massey, A. J. (1977). Candidate fatigue and performance on GCE objective tests. *British Journal of Educational Psychology*, 47, 203–208.
- Mednick, S. C., Nakayama, K., Cantero, J. L., Atienza, M., Levin, A. A., Pathak, N., & Stickgold, R. (2002). The restorative effect of naps on perceptual deterioration. *Nature Neuroscience*, 5(7), 677–681.
- Mollenkopf, W. G. (1950). An experimental study of the effects on item analysis data of changing item placement and test time limit. *Psychometrika*, 15, 291–315.
- Newburger, M. (1942). The relative importance of homogeneity and difficulty in the development of mental fatigue at two different levels of intelligence. *Journal of Applied Psychology*, 26, 81–93.
- Noll, V. H. (1932). A study of fatigue in three-hour college ability tests. *Journal of Applied Psychology*, 16, 175–183.
- Poffenberger, A. T. (1928). The effects of continuous work upon output and feelings. *Journal of Applied Psychology*, 12, 459–467.
- Samkoff, J. S., & Jacques, C. H. M. (1991). A review of studies concerning effects of sleep deprivation and fatigue on residents' performance. *Academic Medicine*, 66(11), 687–693.
- Smith, A. (1989). Diurnal variations in performance. In A. M. Colley & J. R. Beech (Eds.), *Acquisition and performance of cognitive skills* (pp. 301–325). New York: John Wiley & Sons.
- Spaeth, R. A. (1920). The problem of fatigue. *Journal of Industrial Hygiene*, 1, 22–53.
- Starch, O., Stanton, H. M., & Koerth, W. (1936). *Controlling human behavior*. New York: Macmillan.
- Thorndike, E. L. (1921). *Educational psychology, Vol. III*. New York: Teacher's College Press.
- Tucker, L. R. (1947). *Memorandum concerning study of effects of fatigue on afternoon achievement scores due to Scholastic Aptitude Test being taken in the morning*. Research Memorandum RM-48-2. Princeton, NJ: Educational Testing Service.
- Wohlhueter, J. F. (1966). *Fatigue in testing and other mental tasks: A literature survey*. Research Memorandum RM-66-6. Princeton, NJ: Educational Testing Service.

Appendix A: SAT Fatigue Study Survey

Name (optional): _____

Thank you for participating in this survey about the test you have just completed. It should only take a few minutes of your time to answer it.

- 1) Now that you have completed the test, would you say you feel...
 - a) Very tired
 - b) Somewhat tired
 - c) Not at all tired
- 2) You were given two five-minute breaks during the course of this test. Would you say the number of breaks provided was...
 - a) Enough
 - b) Not enough
 - c) Don't know
- 3) Would you say the breaks were placed at the appropriate sections of the test?
 - a) Yes
 - b) No
 - c) Don't know
- 4) If you answered "b" to questions 2 or 3 above, please fill in the chart below indicating where you think the breaks should have been placed. Check 5-minute, 1-minute, or no break for each section.

After	5-minute break	1-minute break	No break
Section 1			
Section 2			
Section 3			
Section 4			
Section 5			
Section 6			
Section 7			

- 5) At the moment would you say you are...
 - a) Very hungry
 - b) Somewhat hungry
 - c) Not at all hungry
- 6) Do you think your level of hunger had a negative effect on your performance toward the end of the test?
 - a) Yes, very much
 - b) Yes, somewhat
 - c) No

The next three questions deal with the **essay section** of the test.

- 7) Where was the essay located in your test?
 - a) At the beginning
 - b) Somewhere in the middle
 - c) At the end
 - d) I can't remember
- 8) Where do you think the essay should be placed in the test?
 - a) At the beginning of the test
 - b) In the middle of the test
 - c) At the end of the test
 - d) I don't know/I don't care
- 9) How much do you agree or disagree with the statement: "I was concerned about the essay section while working on other parts of the test."
 - a) Agree completely
 - b) Agree somewhat
 - c) Neither agree nor disagree
 - d) Disagree somewhat
 - e) Disagree completely

10) Have you taken the SAT II: Writing Test before? If so, please indicate the month and year in which you last took it.

- a) No
- b) Yes; I took it in _____

If you answered "No" to question 10, please skip question 11.

11) Think back to when you previously took the SAT II: Writing Test and answer this question: In what location of the test would you prefer to have the essay section?

- a) The location of the essay was the same today and on my previous SAT II: Writing Test
- b) I preferred the location of the essay in today's test
- c) I preferred the location of the essay in my previous SAT II: Writing Test
- d) I don't know/I don't care where the essay is located

12) Do you plan to take the SAT I test during the coming school year? If so, please indicate the month and year in which you plan to take it.

- a) No
- b) Yes; I plan to take it in _____

13) Have you taken the SAT I test before? If so, please indicate the month and year in which you last took it.

- a) No
- b) Yes; I took it in _____

If you answered "No" to question 13, please skip question 14.

14) Thinking back on your previous SAT I testing, with which of the following statements would you agree most concerning today's testing situation?

"I think that switching among the three subject areas (Verbal, Math, and Writing)..."

- a) Really hurt how I did on the test
- b) Sort of hurt how I did on the test
- c) Had no effect on how I did on the test
- d) Sort of helped how I did on the test
- e) Really helped how I did on the test

In your opinion, what could be done to improve the content or the administration of the test?

Thank you very much for your time and cooperation.

Appendix B: SAT I Fatigue Study Survey

Name (optional): _____

Thank you for participating in this survey about the test you have just completed. It should only take a few minutes of your time to answer it.

1) Now that you have completed the test, would you say you feel...

- a) Very tired
- b) Somewhat tired
- c) Not at all tired

2) You were given one five-minute and a one-minute break during the course of this test. Would you say the number of breaks provided was...

- a) Enough
- b) Not enough
- c) Don't know

3) Would you say the breaks were placed at the appropriate sections of the test?

- a) Yes
- b) No
- c) Don't know

4) If you answered "b" to questions 2 or 3 above, please fill in the chart below indicating where you think the breaks should have been placed. Check 5-minute, 1-minute, or no break for each section.

After	5-minute break	1-minute break	No break
Section 1			
Section 2			
Section 3			
Section 4			
Section 5			
Section 6			
Section 7			

-
- 5) At the moment would you say you are...
- a) Very hungry
 - b) Somewhat hungry
 - c) Not at all hungry
- 6) Do you think your level of hunger had a negative effect on your performance toward the end of the test?
- a) Yes, very much
 - b) Yes, somewhat
 - c) No
- 7) Do you plan to take the SAT I test during the coming school year? If so, please indicate the month and year in which you plan to take it.
- a) No
 - b) Yes; I plan to take it in _____
- 8) Have you taken the SAT I test before? If so, please indicate the month and year in which you last took it.
- a) No
 - b) Yes; I took it in _____

Thank you very much for your time and cooperation.

If you answered "No" to question 8, please skip question 9.

- 9) Thinking back on your previous SAT I testing, with which of the following statements would you agree most concerning today's testing situation?

"I think that switching among the three subject areas (Verbal, Math, and Writing)...

- a) Really hurt how I did on the test
- b) Sort of hurt how I did on the test
- c) Had no effect on how I did on the test
- d) Sort of helped how I did on the test
- e) Really helped how I did on the test

In your opinion, what could be done to improve the content or the administration of the test?

