

Examination of
Fatigue Effects from
Extended-Time
Accommodations on the
SAT Reasoning Test™

**Cara Cahalan Laitusis, Deanna L. Morgan, Brent Bridgeman,
Jennifer Zanna, and Elizabeth Stone**

The College Board, New York, 2007

Cara Cahalan Laitusis is a research scientist in the Center for Validity Research at Educational Testing Service (ETS).

Deanna L. Morgan is an associate research scientist in the Research and Analysis department at the College Board.

Brent Bridgeman is a principal research scientist in the Center for Validity Research at ETS.

Jennifer Zanna is a data analyst manager in the Center for Statistical Analysis at ETS.

Elizabeth Stone is a senior research associate in the Center for Validity Research at ETS.

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

The College Board: Connecting Students to College Success

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

For further information, visit www.collegeboard.com.

Additional copies of this report (item #060482036) may be obtained from College Board Publications, Box 886, New York, NY 10101-0886, 800 323-7155. The price is \$15. Please include \$4 for postage and handling.

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

Printed in the United States of America.

Contents

Abstract 1

Introduction 1

Method 2

Test Forms 2

Sample 3

Analysis 4

Item Completion 4

Differential Item Functioning 4

Results 5

Item Completion Rates 5

Differential Item Functioning 6

Conclusions 8

References 9

Appendix 10

Tables

1. Order of Test Administration, Length of Sections, and Total Test (in Minutes) by Timing Condition for Form 1 2

2. Order of Test Administration, Length of Sections, and Total Test (in Minutes) by Timing Condition for Form 2 3

3. Order of Test Administration, Length of Sections, and Total Test (in Minutes) by Timing Condition for Form 3 3

4. Sample Sizes and Average Scaled Scores on SAT® by Subgroup 3

5. DIF Analyses 4

6. Summary of DIF Results by Test, Reference–Focal Comparison Groups, and Section Order 7

A1. DIF Results by Item for Form 1 Writing (Standard and Time and a Half) 10

A2. DIF Results by Item for Form 1 Writing (Standard and Double Time) 11

A3. DIF Results by Item for Form 2 Critical Reading (Standard and Time and a Half) 12

A4. DIF Results by Item for Form 3 Critical Reading (Standard and Double Time) 13

Figures

1. Percent reaching rates for Form 1 early writing section (standard and time and a half) 5

2. Percent reaching rates for Form 1 late writing section (standard and time and a half) 5

3. Percent reaching rates for Form 1 early writing section (standard and double time) 6

4. Percent reaching rates for Form 1 late writing section (standard and double time) 6

5. Percent reaching rates for Form 3 early critical reading section (standard and double time) 6

6. Percent reaching rates for Form 3 late critical reading section (standard and double time) 6

7. Percent reaching rates for Form 2 early critical reading section (standard and time and a half) 7

8. Percent reaching rates for Form 2 late critical reading section (standard and time and a half) 7

Abstract

This study examined operational data from the SAT Reasoning Test™ to determine if students who tested under extended-time conditions were suffering from excessive fatigue relative to students who tested under standard-time conditions. Excessive fatigue was defined by significant (a) increases in differential item functioning (DIF) and (b) decreases in item completion rates, for items at the end of testing compared to the beginning of testing. Both of these factors were examined by comparing the performance of students who tested under standard time to students who tested with extended time on items administered early (section position 2 or 3) and different items administered late (section position 8, 9, or 10) during the 10-section test administration. The sample included students with learning disabilities and/or Attention-Deficit/Hyperactivity Disorder (ADHD) who tested with extended time (time and a half or double time) and students without disabilities who tested under standard-time conditions. Analyses were conducted on the critical reading and writing sections of the SAT® and examined item difficulty as well as item completion rates. Results indicated few changes in levels of DIF (early in the test compared to late in the test). In addition, item completion rates for students who received extra time were comparable to (or in some cases higher than) test-takers without disabilities who tested under standard time on both early and late sections.

Introduction

Over the last 20 years, several research studies have examined the impact of extended-time accommodations on the performance of SAT Reasoning Test examinees with and without disabilities (Mandinach, Bridgeman, Cahalan Laitusis, and Trapani, 2005; Camara, Copeland, and Rothschild, 1998; Willingham, Ragosta, Bennett, Braun, Rock, and Powers, 1988). In addition, several research studies have examined the predictive validity of SAT test scores obtained with extended-time accommodations (Cahalan, Mandinach, and Camara, 2002; Braun, Ragosta, and Kaplan, 1986). This prior research focused on ensuring that extended time resulted in comparable scores (relative to test scores obtained under standard-time conditions) and ensuring that students who received extended time were not receiving an unfair advantage relative to students without disabilities who took the test under standard-time conditions. Research examining the negative impact of testing for extended lengths of time has been limited to students without disabilities.

An overview of research on test fatigue was included in Liu, Allspach, Feigenbaum, Oh, and Burton (2004). This review of multiple studies, conducted since the beginning of standardized testing, concluded that fatigue

does not have a major impact on test performance, even during five- to six-hour testing periods, although the impact of fatigue can vary depending on task type. Tasks that are basic, repetitive, and viewed by the test-taker as low stakes can be impacted by fatigue, with performance dropping over time. Tasks that are complex, varied, and viewed as high stakes—such as tests of intelligence, reading, and math for admissions purposes—do not appear to be impacted by fatigue. However, test-takers often reported feeling fatigued over longer testing periods even though their performance did not change. Whether these findings hold for students with disabilities is unclear, and extended-time accommodations for such students can potentially reach or exceed the maximum times used in most of the reviewed studies.

Research on the new SAT Reasoning Test introduced in March 2005 is limited. Wang (2006) examined the effect of increased testing time by comparing four performance indices calculated using randomly equivalent examinee subpopulations on sections of similar content and difficulty administered at different times on three SAT administrations. This study was conducted to address concerns that the increased length of the new SAT Reasoning Test was resulting in increased fatigue and poorer performance. A variety of analyses were conducted in this study, and the researcher found no evidence that the current SAT test length had affected examinee performance at the population level or differentially across gender, racial/ethnic, and language subgroups. On the contrary, this study produced consistent findings, indicating that examinees performed the same on sections of similar content and difficulty, both in terms of direct group comparisons and comparisons conditional on total score, throughout the entire SAT. Furthermore, the findings from the March and October 2005 SAT data were replicated using the May 2002 SAT I data, indicating no significant changes in performance trends between the two tests. The Wang study, however, did not include students with disabilities.

The purpose of this study was to examine the impact of test fatigue on student performance by comparing the performance of students with learning disabilities (LD) or Attention-Deficit/Hyperactivity Disorder who tested under extended-time conditions to the performance of students without disabilities who tested under standard-time conditions. Since it is difficult for experimentally designed research studies to simulate the level of motivation of examinees taking the test at an operational test administration, this study used operational test data to examine test fatigue. Although the use of operational test data provides advantages, one limitation was that the order of sections was not systematically spiraled to result in two test forms with the same sections administered at the beginning of the test in one form and at the end of the test in the other form. In addition, all possible test forms are not administered to students who receive extended-

time test accommodations. Given these limitations, the analyses included in this study examined changes in (a) the number of items flagged for differential item functioning (DIF) and (b) item completion rates between test sections administered early in the test form and test sections administered later in the same test form. Both of these analyses compare students with disabilities who received extra time to students without disabilities who tested under standard-time conditions. These analyses allow us to look at excessive fatigue for students who receive extra time *relative* to students without disabilities who are tested under standard-time conditions, but it is possible that both groups experience fatigue at the end of the test, and this study will not be able to determine if fatigue exists for both groups. In addition, we are making assumptions that the early and late test sections are of comparable content, and this content is of equal difficulty for all groups. Given these assumptions and limitations, the findings from this study should be viewed as a first step in examining fatigue. These findings will not provide a conclusive determination that a fatigue effect exists (or does not exist) for students who receive extended-time accommodations.

Method

Test Forms

All test forms and spirals administered from March and May 2005 were examined to determine the ideal forms for analysis based on order of sections and where the test form was administered (school or center).¹ Either 9 or 10 sections were administered in each of the test forms considered for inclusion in this study. These sections included three writing sections (one essay and two multiple-choice sections), three mathematics sections, and three critical reading sections. Students who tested under standard-time conditions also received an additional 25-minute variable section for item pretesting. Students who tested under the time-and-a-half (1.5x) and double-time (2x) conditions did not receive the variable section. Since the essay always comes first, we defined “early” sections as sections administered in positions 2 or 3 and “late” sections as positions 8, 9, or 10.²

After examining all possible test forms and spirals, three tests forms were selected for analyses. Since none of the test forms administered from March and May 2005

included both an early and late mathematics section, this study only examines fatigue on the critical reading and writing sections of these three forms. This report will provide results on analyses conducted on the writing (SAT-W) section of a test form administered in March 2005, the critical reading (SAT-CR) section of another test form administered in March 2005, and the critical reading section of a test form administered in May 2005. The three forms will be referred to as Form 1, Form 2, and Form 3, respectively. Specifically, Form 1 was administered in March 2005 to five subgroups (i.e., nondisabled, LD 1.5x,³ LD 2x,⁴ ADHD 1.5x,⁵ and ADHD 2x⁶) and included early and late writing sections. Form 2 was administered in March 2005 to three subgroups (nondisabled, LD 1.5x, and ADHD 1.5x) and included early and late critical reading sections. Form 3 was administered in May 2005 to three subgroups (nondisabled, LD 2x, and ADHD 2x) and included early and late critical reading sections. Tables 1 through 3 display the order of sections in each of the test forms analyzed for this report as well as the length

Table 1

Order of Test Administration, Length of Sections, and Total Test (in Minutes) by Timing Condition for Form 1

Section Order	Section	Standard Time	Time and a Half	Double Time
1	Essay	25	38	50
2	Mathematics 2	25	38	50
	Break	5	5	5
3	Writing 1	25	38	50
4	Reading 2	25	38	50
	Break	1	1	NA
5	Mathematics 1	25	38	50
6	Reading 1	25	38	50
	Break	5	5	5
7	Variable	25	NA	NA
8	Mathematics 3	20	30	40
9	Reading 3	20	30	40
10	Writing 2	10	15	20
	Total Day 1	236	314	205
	Total Day 2	0	0	205
	Total	236	314	410

Note: NA = not applicable because variable section was not administered to extended-time test-takers and breaks varied by timing condition.

¹Double-time accommodations are almost universally provided in administrations at schools, and time-and-a-half accommodations are nearly exclusively provided at test centers.

²The last section in each test form is fixed as a multiple-choice writing section, so late critical reading sections were positioned slightly earlier than late writing sections.

³LD 1.5x refers to students with learning disabilities who received a time-and-a-half accommodation.

⁴LD 2x refers to students with learning disabilities who received a double-time accommodation.

⁵ADHD 1.5x refers to students with ADHD who received a time-and-a-half accommodation.

⁶ADHD 2x refers to students with ADHD who received a double-time accommodation.

Table 2

Order of Test Administration, Length of Sections, and Total Test (in Minutes) by Timing Condition for Form 2

Section Order	Section	Standard Time	Time and a Half
1	Essay	25	38
2	Reading 1	25	38
	Break	5	5
3	Mathematics 2	25	38
4	Writing 1	25	38
	Break	1	1
5	Reading 2	25	38
6	Mathematics 1	25	38
	Break	5	5
7	Variable	25	NA
8	Reading 3	20	30
9	Mathematics 3	20	30
10	Writing 2	10	15
	Total	236	314

Note: NA = not applicable because variable section was not administered to extended-time test-takers.

(in minutes) for each section. The double-time condition is administered across two days with sections 1 through 4 being administered on the first day and sections 5 through 9 being administered on the second day.

Table 3

Order of Test Administration, Length of Sections, and Total Test (in Minutes) by Timing Condition for Form 3

Section Order	Section	Standard Time	Double Time
1	Essay	25	50
2	Reading 1	25	50
	Break	5	5
3	Mathematics 1	25	50
4	Variable	25	NA
	Break	1	NA
5	Reading 2	25	50
6	Mathematics 2	25	50
	Break	5	NA
7	Writing 1	25	50
	Break	NA	5
8	Reading 3	20	40
9	Mathematics 3	20	40
10	Writing 2	10	20
	Total Day 1	236	205
	Total Day 2	0	205
	Total	236	410

Note: NA = not applicable because variable section was not administered to extended-time test-takers and breaks varied by timing condition.

Sample

The sample for this study included students who took one of the three test forms described above but was limited to test-takers without disabilities who tested under standard-time conditions, test-takers with learning disabilities (LD) and no other disability who tested under extended-time conditions, and test-takers with Attention-Deficit/Hyperactivity Disorder (ADHD) and no other disability who tested under extended-time conditions. Test-takers with disabilities (LD or ADHD) were further subdivided by the amount of extra time they received, time and a half and double time. Table 4 displays the sample sizes and SAT scores across test form, disability group, and timing condition. The average scores for all groups are slightly higher for Forms 1 and 2 (March administration) than for Form 3 (May administration). However, the average performance between subgroups is fairly consistent across forms, with students with learning disabilities performing worse on average than students with ADHD and students without disabilities. Although the average scaled scores are fairly equivalent for students without disabilities and students with ADHD, the sample of students with learning disabilities performs between .3 and .5 standard deviation units below the students with no disability and students with ADHD.

Table 4

Sample Sizes and Average Scaled Scores on SAT® by Subgroup

Disability	Timing	SAT-CR			SAT-M		SAT-W	
		n	M	SD	M	SD	M	SD
Form 1: Writing								
None	Standard	126,111	540	106	560	105	542	105
LD	Time and a Half	628	505	109	531	112	509	108
LD	Double Time	799	510	120	528	132	513	115
ADHD	Time and a Half	281	548	107	576	115	557	101
ADHD	Double Time	194	545	113	561	118	551	107
Form 2: Critical Reading								
None	Standard	121,955	540	105	560	105	540	105
LD	Time and a Half	596	500	113	527	125	498	117
ADHD	Time and a Half	306	548	90	578	104	554	94
Form 3: Critical Reading								
None	Standard	104,833	513	108	538	110	513	104
LD	Double Time	534	474	116	492	130	470	120
ADHD	Double Time	123	526	112	544	116	525	116

Analysis

In order to examine test fatigue with operational test data, we employed two different analyses: item completion rates by disability group and testing accommodation and differential item functioning (DIF) on items placed early and late in the test form. These analyses examined the differences in item completion rates and differential item difficulty between test-takers without disabilities (who took the SAT with standard time limits) and the four disability/accommodation subgroups: (1) students with learning disabilities who took the SAT with time and a half (LD 1.5x), (2) students with learning disabilities who took the SAT with double time (LD 2x), (3) students with Attention-Deficit/Hyperactivity Disorder who took the SAT with time and a half (ADHD 1.5x), and (4) students with Attention-Deficit/Hyperactivity Disorder who took the SAT with double time (ADHD 2x). Since students were not randomly assigned to the timing conditions, it is possible that disability and accommodation may be intertwined (e.g., students with learning disabilities who received extra time may have performed differently on test sections administered late whether or not they had extra time); however, if relatively few items perform differently (in terms of differential item difficulty), this study will provide some evidence that a fatigue effect is not contributing to score differences between standard- and extended-time administrations.

Item Completion

The first set of analyses examined item completion rates across the different disability subgroups for the sections of the test given early and the sections of the test given later in the administration to determine if the additional testing time resulted in lower item completion rates on later sections. For this report, *item completion rate* refers to the extent to which test-takers are able to complete each item within a test section in the allotted time. The item completion rates were calculated using the same procedure used with test data from operational administrations of the SAT Reasoning Test (see Cahn and Kelly, 2006), and these rules require the identification of reached and not-reached items. Reached items are defined as any item that the test-taker answers or any unanswered items that have

subsequent items answered in that section (e.g., if item 5 is blank and item 6 is answered, then item 5 would be considered to have been reached). These reached items are considered to have been completed. Items defined as not reached include items that are not answered where all subsequent items are also not answered (e.g., if items 5–20 on a 20 item section are not answered, then the last item reached would be item 4, and items 5–20 would be considered not reached). These not-reached items are considered to have been not completed.

Differential Item Functioning

The second set of analyses examined differential item functioning (DIF) between students with and without disabilities to determine if items in the same content area seen on later sections of the form exhibited more DIF than items seen on earlier sections of the form. For these analyses, students were matched on their scaled score in the same content area (e.g., when examining performance on early and late critical reading items, students were matched on their SAT critical reading scaled score [SAT-CR]). While a variety of statistical procedures exist for examining DIF, this study employed the Mantel-Haenszel D-DIF statistic (MH D-DIF), which is used operationally on the SAT Reasoning Test (see Holland and Thayer, 1988 for a full explanation of DIF analyses). The DIF statistic is expressed as differences in item difficulty on the delta scale, where delta is an ETS index used to characterize item difficulty. A negative value means that the test question is differentially more difficult for the focal group (i.e., in this case students with a disability who received extra time), while a positive value indicates that the item is differentially more difficult for the reference group (i.e., students without disabilities). Items are then categorized into three categories (i.e., A, B, and C). Category A contains items with negligible DIF, category B contains items with slight to moderate levels of DIF, and category C contains items with moderate to large levels of DIF. For this study, we look at both category B items and category C items with negative and positive DIF values (e.g., an item with a large DIF value favoring the reference group is referred to as “negative C-DIF” or “C-”). See Table 5 for a complete list of the reference-focal comparisons by content area and test form.

Table 5

DIF Analyses

Reference		Focal		Test Form		
Disability	Time	Disability	Time	Form 1	Form 2	Form 3
None	Standard	LD	Double Time	Writing	NA	Critical Reading
None	Standard	ADHD	Double Time	Writing	NA	Critical Reading
None	Standard	LD	Time and a Half	Writing	Critical Reading	NA
None	Standard	ADHD	Time and a Half	Writing	Critical Reading	NA

Note: NA = not applicable because the test form was not administered under this specific timing condition.

Results

Item Completion Rates

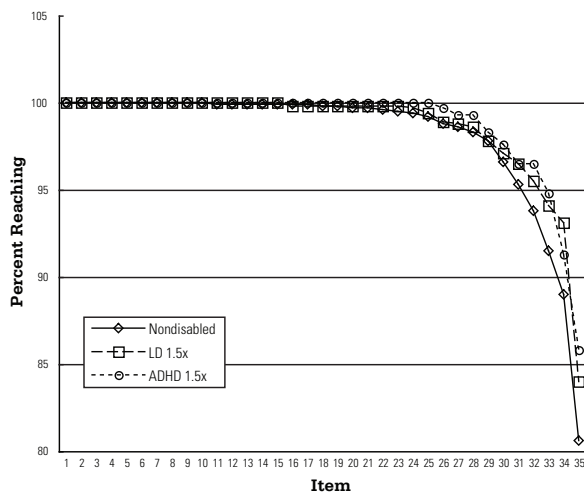
The percentage of students completing each item in the early test sections and late test sections are displayed in Figures 1 through 8. For a particular form and section, the percentage of students with disabilities completing items at the end of that section is generally larger than the percentage of students without disabilities completing the same items at the end of the section. For instance, in Figure 1 (when the section is given early) nearly all students complete items 1 through 14, but starting with item 15, the percentage of students without disabilities who complete this item and subsequent items begins to gradually decrease, while the same gradual decrease does not occur until item 19 for the LD 1.5x group and item 26 for the ADHD 1.5x group. At the end of the section, the percentage of extended-time students completing the final item (item 35) is greater than the percentage of standard-time students completing the final item (by 3 percent for LD 1.5x and by 5 percent for ADHD 1.5x).

Although the item completion rates are generally greater for students in the extended-time conditions than students without disabilities (standard time), the gap in item completion rates between test-takers without disabilities and test-takers with disabilities who test under extended-time conditions appears to close when comparisons are made between the early and late test sections of the same form. In Figure 2 (the later test section of the same test form described above), the slope of the line is more gradual for the ADHD 1.5x group⁷ than

the LD 1.5x or standard-time groups (which are similar to each other). This change in the completion rates between the early and late sections on Form 1 (writing) (compare Figures 1 and 2) may indicate some slight fatigue relative to the standard-time condition, but test-takers in the extended-time conditions (LD 1.5x and ADHD 1.5x) are still completing the later test section at higher or equal rates than the test-takers who test under standard-time conditions.

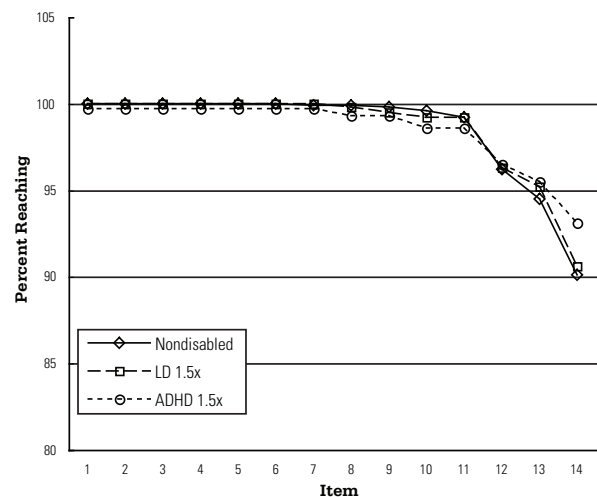
The pattern of ADHD and LD test-takers having higher or comparable item completion rates than the test-takers without disabilities is also found when comparing students in the double-time condition (see Figures 3, 4, 5, and 6), which was expected to have no difference in item completion rates since students who received double time were tested over two days. Comparisons between standard- and double-time conditions on Form 1 (writing) and Form 3 (critical reading) indicate similar findings as those found above, with students who received double time (LD and ADHD) completing both early and late test sections at a higher rate than test-takers who received standard time. In addition, the gap in item completion rates between extended- and standard-time test-takers closes slightly on the late test section (see Figures 3 versus 4 and Figures 5 versus 6). Since double-time test-takers are testing over two days, it is less likely that the length of the testing session contributed to the lower item completion rates on late test sections.

The one exception to the finding of higher item completion rates for extended-time test-takers is displayed in Figures 7 versus 8 (Form 2 critical reading), where students with learning disabilities who received the



Note: For all subgroups, this was section 3.

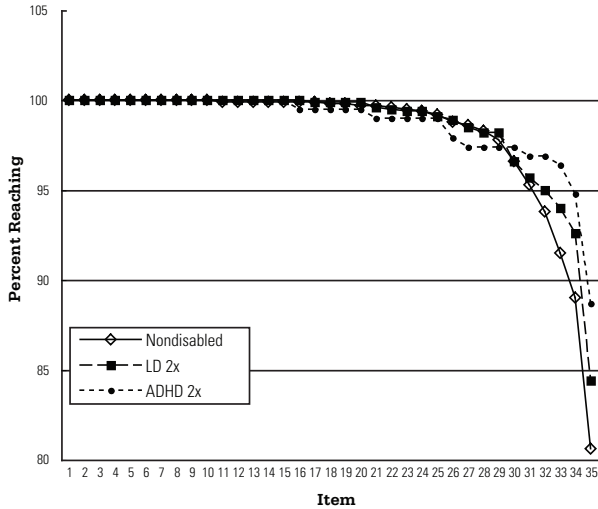
Figure 1. Percent reaching rates for Form 1 early writing section (standard and time and a half).



Note: For the Nondisabled subgroup, this section was section 10 of 10, and for the extended-time subgroups, this was section 9 of 9.

Figure 2. Percent reaching rates for Form 1 late writing section (standard and time and a half).

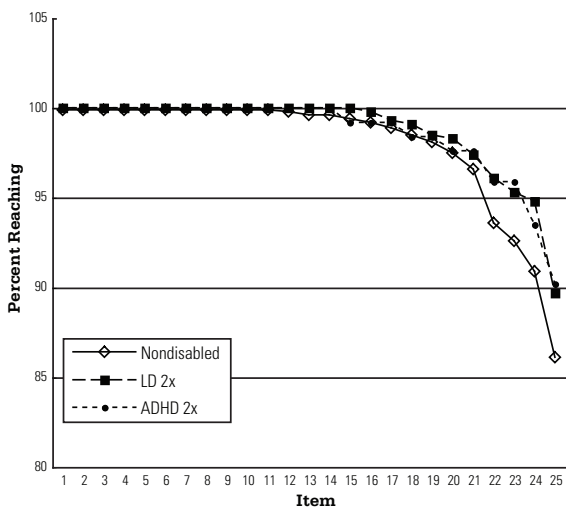
⁷This gradual decrease is somewhat masked in the ADHD sample, due to eight test-takers who did not attempt any items in this section.



Note: For all subgroups, this was section 3.

Figure 3. Percent reaching rates for Form 1 early writing section (standard and double time).

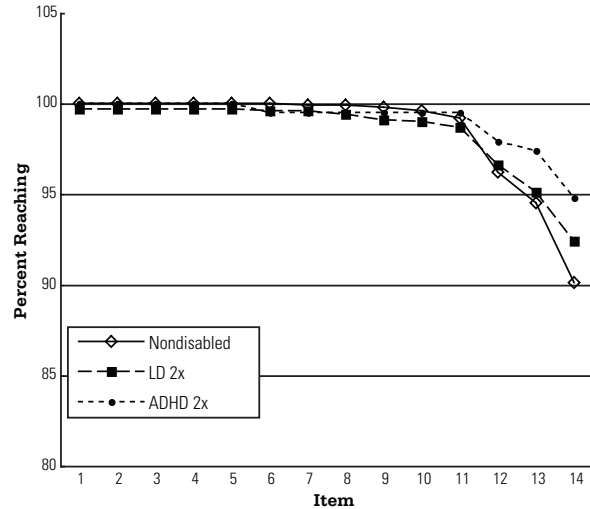
time-and-a-half accommodation (LD 1.5x) had item completion rates comparable to standard-time test-takers on the early section and slightly lower item completion rates on the late test section. Although this finding may indicate a slight fatigue effect from the extended-time conditions, the similar findings from the double-time condition give pause to this conclusion and indicate that some other factors, i.e., difficulty concentrating for long periods of time or poor performance on the early section resulting in decreased effort on later sections, may be



Note: For all subgroups, this section was section 2.

Figure 5. Percent reaching rates for Form 3 early critical reading section (standard and double time).

⁸The total of 370 comparisons includes 196 comparisons in Form 1 (49 writing items x 4 reference-focal comparisons), 86 comparisons in Form 2 (43 critical reading items x 2 reference-focal comparisons), and 88 comparisons in Form 3 (44 critical reading items x 2 reference-focal comparisons).



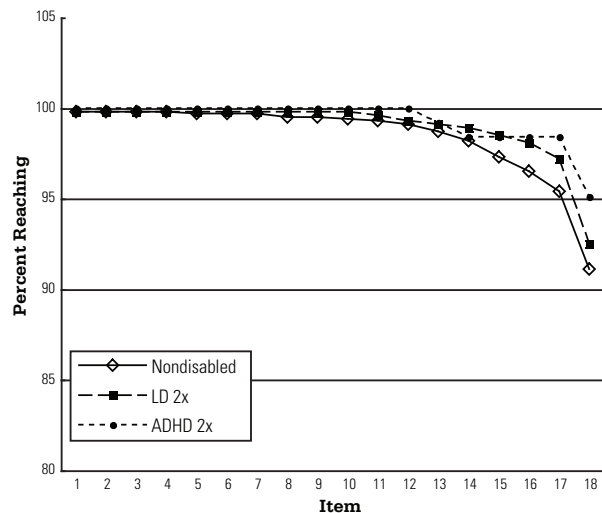
Note: For the Nondisabled subgroup, this section was section 10 of 10, and for the extended-time subgroups, this was section 9 of 9.

Figure 4. Percent reaching rates for Form 1 late writing section (standard and double time).

impacting the item completion rates from test-takers with disabilities.

Differential Item Functioning

Table 6 shows the degree and direction of DIF by item placement (early or late), test form and content, and reference-focal comparison. For more detailed DIF information on an item-by-item basis, see the Appendix. Across the 370 reference-focal group comparisons,⁸



Note: For the Nondisabled subgroup, this was section 8 of 10, and for the extended-time subgroups, this was section 7 of 9.

Figure 6. Percent reaching rates for Form 3 late critical reading section (standard and double time).

Table 6

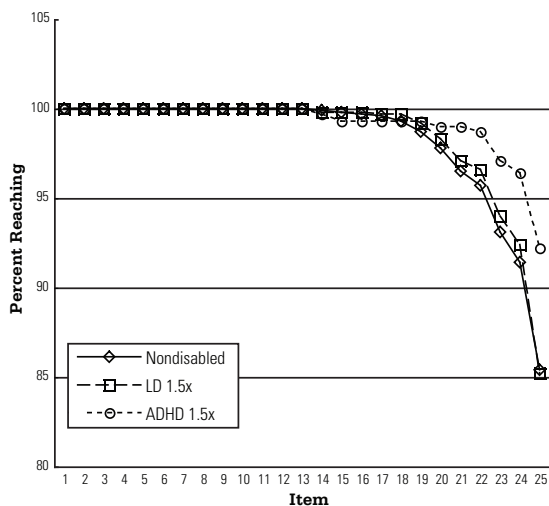
Summary of DIF Results by Test, Reference–Focal Comparison Groups, and Section Order

Reference	Focal	Early					Late				
		C-	B-	A	B+	C+	C-	B-	A	B+	C+
Form 1: Writing											
Nondisabled	LD 1.5x	0	0	35	0	0	0	0	14	0	0
Nondisabled	ADHD 1.5x	0	0	33	2	0	0	0	14	0	0
Nondisabled	LD 2x	0	1	33	1	0	0	0	14	0	0
Nondisabled	ADHD 2x	0	0	34	1	0	0	0	13	1	0
Form 2: Critical Reading											
Nondisabled	LD 1.5x	0	0	25	0	0	0	2	17	0	0
Nondisabled	ADHD 1.5x	0	0	25	0	0	0	2	17	0	0
Form 3: Critical Reading											
Nondisabled	LD 2x	0	0	24	1	0	0	0	18	0	0
Nondisabled	ADHD 2x	1	0	24	0	0	0	1	15	2	0

nearly all items (95 percent) showed no differential item difficulty between students with disabilities who received timing accommodations and students without disabilities who tested under standard conditions. Of the remaining items, six showed moderate DIF favoring the reference group, eight showed moderate DIF favoring the focal group, and one showed strong DIF favoring the reference group. Two items exhibited DIF for both the time-and-a-half (1.5x) accommodation for students with learning disabilities and for the time-and-a-half (1.5x) accommodated students with ADHD. Therefore, of the 15 cases of DIF identified, only 13 unique items

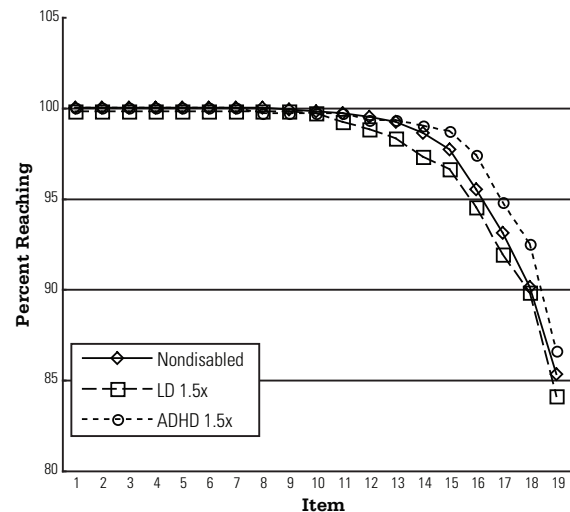
were involved. Of these items 7 of the 13 unique items exhibiting DIF appeared early in the test (section 2 or 3) rather than later (section 8, 9, or 10). The items that showed DIF were examined to attempt to identify any characteristics that may be contributing to the DIF for that group of examinees. Careful inspection of each item exhibiting DIF revealed no obvious rationale for the existence of DIF on that item. Items flagged for DIF were both of easy and more challenging difficulty and occurred across item types.⁹

Since the existence of DIF does not by itself indicate a fatigue effect, we also examined the direction of the



Note: For all subgroups, this was section 2.

Figure 7. Percent reaching rates for Form 2 early critical reading section (standard and time and a half).



Note: For the Nondisabled subgroup, this was section 8 of 10, and for the extended-time subgroups, this was section 7 of 9.

Figure 8. Percent reaching rates for Form 2 late critical reading section (standard and time and a half).

⁹All items were multiple choice, but the item types included a variety of content. The critical reading sections included sentence completion items as well passage-based reading comprehension items, while the writing sections included items focused on improving sentences, identifying sentence errors, and improving paragraphs.

DIF relative to the placement of the item (early or late test section) and the extended-time condition (1.5x or 2x). When comparisons between the standard-time and time-and-a-half conditions were examined, results indicated that the items exhibiting DIF on the early sections favored the reference group (two items with positive B-DIF and no items with negative DIF) and items exhibiting DIF on the late sections favored the focal groups (four items with negative B-DIF and no items with positive DIF). This finding may indicate that fatigue was slightly more of a factor for the students in the time-and-a-half condition compared to the standard-time condition, but other contributing factors (e.g., impact of disability or changes in items) can not be ruled out. The comparisons for standard time and double time were nearly identical for early and late sections (i.e., three items with positive DIF on both early and late sections, two items with negative DIF on early sections, and one item with negative DIF on a late section). While there appears to be a slight change in item difficulty for the time-and-a-half condition compared to standard time, it is important to point out that all of the observed DIF was slight to moderate (B-DIF), and 97 percent of the comparisons between time-and-a-half and standard-time subgroups exhibited no DIF.

Conclusions

This study represents a first step in examining fatigue effects from the extended-time testing accommodation. Given the limitations of this research identified earlier in this report, the results should be interpreted with caution, and additional research is still required. That said, the results of this research did not provide evidence of a significant fatigue effect (determined by either item completion or differential item difficulty) between standard- and extended-time testing conditions on the SAT writing sections studied. However, some slight discrepancies in the direction of differential item functioning on early and late sections were noted between the standard-time and extended-time conditions on the critical reading sections of Form 2. The same test form and section also showed slightly decreased item completion rates on the late section for the LD 1.5x group compared to the students without disabilities who tested under standard time. However, the differences observed were extremely small and may be of no practical significance.

Overall, the item completion rates for students who tested under extended-time conditions, both time and a half and double time, were generally higher than for the sample without disabilities on early sections. This provides some indication that the extended time may be allowing relatively more students with disabilities, particularly students with ADHD, the opportunity to complete all

items in the test section. This finding is consistent with prior research indicating that, for the majority of students with learning disabilities and ADHD, the time-and-a-half accommodation is sufficient to complete the entire SAT critical reading and writing test sections (Cahalan Laitusis, King, Cline, and Bridgeman, 2006).

Although this study provides some evidence that students who receive extended-time accommodations are not significantly more fatigued than students who test under standard-time conditions (particularly on the writing section), there are several limitations to this research. The most obvious limitation is that by using operational data, we are unable to disentangle the interaction between disability and accommodation. Another limitation is that while we compared sections given in the same position, these sections were not administered at the same time, so the exact time that the early section began was between 13 and 50 minutes later for the extended-time test-takers, and the later sections were actually given 25 to 30 minutes earlier in the test day for students in the double-time condition (all of whom tested over a two-day period) than students who tested under standard-time conditions. Finally, it is possible that the item completion rates are influenced by factors other than testing time. For example, students without disabilities may be more likely than students with ADHD to leave items at the end of a test section blank (to reduce the point deductions incurred from wrong answers on a formula scored test such as the SAT). Even with these limitations, the research does provide information on how test-takers who receive extended-time accommodations are performing on a specific section of the operationally administered SAT Reasoning Test relative to students who test under standard-time conditions. More research is needed to confirm the present findings and to expand these analyses to the mathematics sections.

Future research should attempt to replicate these analyses using additional test forms and administrations. It would be particularly important to see if the findings observed when comparing students without disabilities to students with learning disabilities testing under time-and-a-half conditions on the critical reading sections, where some slight changes in item difficulty and item completion rates between early and late sections were seen, occurs with other test forms. In addition, this research should be expanded to examine item completion rates and differential item functioning on early and late mathematics sections of the SAT. Finally, research should attempt to compare changes in item completion rates and difficulty level on the same items administered in different positions (early and late) on two spirals of the same test form that are randomly assigned to test-takers who test under standard- and extended-time conditions.

References

- Braun, H., Ragosta, M., & Kaplan, B. (1986). Predictive validity. In W. W. Willingham, M. Ragosta, R. E. Bennett, H. Braun, D. A. Rock, & D. E. Powers (Eds.), *Testing handicapped people* (109–32). Boston: Allyn and Bacon.
- Camara, W. J., Copeland, T., & Rothschild, B. (1998). *Effects of extended time on the SAT I: Reasoning Test score growth for students with learning disabilities*. (College Board Research Report No. 1998-7). New York: The College Board.
- Cahalan Laitusis, C., King, J., Cline, F., & Bridgeman, B. (2006). *Observational timing study on the SAT Reasoning Test for test-takers with learning disabilities and/or ADHD*. (College Board Research Report No. 2006-4). New York: The College Board.
- Cahalan, C., Mandinach, E. B., & Camara, W. J. (2002). *Predictive validity of SAT I: Reasoning Test for test-takers with learning disabilities and extended time accommodations*. (College Board Research Report No. 2002-5). New York: The College Board.
- Cahn, M., & Kelly, R. (2006). *Test analysis: College Board SAT Reasoning Test—fall 2005 administrations, 4BSA08-4BSA10*. Princeton, NJ: Educational Testing Service.
- Holland, P. H., & Thayer, D. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer and H. Braun (Eds.), *Test Validity* (129-45). Hillsdale, NJ: Erlbaum.
- Liu, J., Allspach, J. R., Feigenbaum, M., Oh, H-J., & Burton, N. (2004). *A study of fatigue effects from the new SAT*. (College Board Research Report No. 2004-5). New York: The College Board.
- Mandinach, E. B., Bridgeman, B., Cahalan Laitusis, C., and Trapani, C. (2005). *The impact of extended time on SAT test performance*. (College Board Research Report No. 2005-8). New York: The College Board.
- Wang, X. B. (March, 2006). *Investigating the effect of new SAT test lengths on the performance of regular SAT examinees*. Unpublished research at the College Board, New York.
- Willingham, W. W., Ragosta, M., Bennett, R. E., Braun, H., Rock, D. A., and Powers, D. E. (Eds.). (1988). *Testing handicapped people*. Boston: Allyn and Bacon.

Appendix: Tables

Table A1

DIF Results by Item for Form 1 Writing (Standard and Time and a Half)

Item	Section	Nondisabled	LD 1.5x			ADHD 1.5x		
		AIS	AIS	MH D-DIF	Category	AIS	MH D-DIF	Category
1	3	0.83	0.8	0.18		0.87	0.46	
2	3	0.87	0.84	0.19		0.84	-0.89	
3	3	0.85	0.84	0.51		0.88	0.42	
4	3	0.73	0.63	-0.24		0.79	0.63	
5	3	0.73	0.63	-0.47		0.71	-0.44	
6	3	0.64	0.51	-0.72		0.68	0.13	
7	3	0.73	0.69	0.32		0.76	0.15	
8	3	0.54	0.49	0.05		0.54	-0.17	
9	3	0.53	0.42	-0.63		0.5	-0.59	
10	3	0.29	0.23	-0.06		0.25	-0.9	
11	3	0.31	0.23	-0.39		0.29	-0.6	
12	3	0.93	0.9	-0.04		0.96	1	
13	3	0.57	0.54	0.35		0.65	0.61	
14	3	0.92	0.9	0.47		0.93	0.17	
15	3	0.89	0.83	-0.45		0.89	-0.33	
16	3	0.68	0.63	0.24		0.77	1.03	B+
17	3	0.74	0.66	-0.13		0.79	0.34	
18	3	0.74	0.74	0.99		0.84	1.36	B+
19	3	0.79	0.71	-0.41		0.77	-0.46	
20	3	0.25	0.19	-0.25		0.28	0.13	
21	3	0.62	0.55	-0.03		0.62	-0.25	
22	3	0.55	0.5	0.1		0.54	-0.39	
23	3	0.5	0.42	-0.08		0.51	-0.18	
24	3	0.89	0.85	-0.15		0.93	0.83	
25	3	0.16	0.13	0.16		0.19	0.34	
26	3	0.26	0.23	0.64		0.32	0.83	
27	3	0.19	0.16	0.15		0.2	-0.13	
28	3	0.61	0.54	0.37		0.68	0.58	
29	3	0.57	0.49	-0.28		0.53	-0.64	
30	3	0.66	0.52	-0.65		0.67	-0.17	
31	3	0.73	0.71	0.69		0.8	0.75	
32	3	0.64	0.59	-0.04		0.63	-0.22	
33	3	0.75	0.69	0.03		0.8	0.57	
34	3	0.81	0.71	-0.4		0.83	0.24	
35	3	0.85	0.79	-0.16		0.79	-0.95	
1	10	0.96	0.95	-0.16		0.97	-0.2	
2	10	0.9	0.85	-0.19		0.92	0.27	
3	10	0.91	0.86	-0.16		0.91	-0.37	
4	10	0.89	0.86	0.38		0.93	0.92	
5	10	0.75	0.66	-0.36		0.76	-0.22	
6	10	0.88	0.82	-0.54		0.92	0.78	
7	10	0.84	0.77	-0.18		0.87	0.21	
8	10	0.71	0.64	-0.05		0.77	0.54	
9	10	0.67	0.55	-0.78		0.67	-0.32	
10	10	0.72	0.63	-0.42		0.74	-0.02	
11	10	0.55	0.49	0.17		0.57	-0.16	
12	10	0.65	0.63	0.53		0.66	-0.05	
13	10	0.46	0.4	0		0.45	-0.41	
14	10	0.34	0.3	0.16		0.34	-0.27	

Note: AIS (Average Item Score) is an indicator of item difficulty, calculated as the mean of the scored responses to the item.

Table A2

DIF Results by Item for Form 1 Writing (Standard and Double Time)

Item	Section	Nondisabled	LD 2x			ADHD 2x		
		AIS	AIS	MH D-DIF	Category	AIS	MH D-DIF	Category
1	3	0.83	0.76	-0.18		0.82	-0.13	
2	3	0.87	0.79	-0.39		0.82	-0.86	
3	3	0.85	0.84	0.68		0.86	0.22	
4	3	0.73	0.59	-0.55		0.74	0.16	
5	3	0.73	0.64	-0.19		0.7	-0.32	
6	3	0.64	0.5	-0.64		0.57	-0.9	
7	3	0.73	0.63	-0.24		0.68	-0.69	
8	3	0.54	0.5	0.24		0.52	-0.22	
9	3	0.53	0.44	-0.28		0.51	-0.22	
10	3	0.29	0.22	-0.26		0.23	-0.83	
11	3	0.31	0.22	-0.52		0.32	0.08	
12	3	0.93	0.9	0.25		0.94	0.68	
13	3	0.57	0.52	0.25		0.58	0.05	
14	3	0.92	0.84	-0.55		0.93	0.64	
15	3	0.89	0.82	-0.28		0.88	-0.17	
16	3	0.68	0.61	0.1		0.73	0.65	
17	3	0.74	0.66	0.08		0.78	0.6	
18	3	0.74	0.69	0.7		0.73	-0.18	
19	3	0.79	0.71	-0.28		0.78	-0.04	
20	3	0.25	0.21	0.11		0.21	-0.74	
21	3	0.62	0.55	0.04		0.56	-0.6	
22	3	0.55	0.47	-0.11		0.57	0.19	
23	3	0.5	0.38	-0.56		0.49	-0.14	
24	3	0.89	0.83	-0.3		0.85	-0.83	
25	3	0.16	0.12	-0.07		0.19	0.67	
26	3	0.26	0.24	0.75		0.34	1.29	B+
27	3	0.19	0.22	1.14	B+	0.23	0.63	
28	3	0.61	0.51	0.15		0.67	0.76	
29	3	0.57	0.54	0.33		0.56	-0.13	
30	3	0.66	0.47	-1.17	B-	0.6	-0.79	
31	3	0.73	0.69	0.57		0.74	0.26	
32	3	0.64	0.63	0.36		0.58	-0.64	
33	3	0.75	0.72	0.56		0.73	-0.09	
34	3	0.81	0.69	-0.47		0.79	-0.08	
35	3	0.85	0.77	-0.37		0.83	-0.03	
1	10	0.96	0.93	-0.34		0.96	-0.18	
2	10	0.9	0.84	-0.05		0.91	0.56	
3	10	0.91	0.85	-0.11		0.94	1.71	B+
4	10	0.89	0.83	0		0.89	0.15	
5	10	0.75	0.7	0.4		0.74	-0.25	
6	10	0.88	0.81	-0.39		0.85	-0.77	
7	10	0.84	0.73	-0.46		0.83	0.01	
8	10	0.71	0.63	0.04		0.74	0.31	
9	10	0.67	0.55	-0.63		0.65	-0.25	
10	10	0.72	0.63	-0.2		0.72	-0.08	
11	10	0.55	0.48	0.25		0.58	0.28	
12	10	0.65	0.58	0.09		0.69	0.61	
13	10	0.46	0.41	0.23		0.52	0.75	
14	10	0.34	0.32	0.5		0.36	0.23	

Note: AIS (Average Item Score) is an indicator of item difficulty, calculated as the mean of the scored responses to the item.

Table A3

DIF Results by Item for Form 2 Critical Reading (Standard and Time and a Half)

Item	Section	Nondisabled	LD 1.5x			ADHD 1.5x		
		AIS	AIS	MH D-DIF	Category	AIS	MH D-DIF	Category
1	2	0.91	0.82	-0.79		0.91	-0.42	
2	2	0.54	0.4	-0.7		0.57	0.06	
3	2	0.46	0.38	0.25		0.49	0.02	
4	2	0.55	0.48	0.09		0.58	0.05	
5	2	0.45	0.38	0.17		0.48	0.1	
6	2	0.32	0.23	-0.53		0.32	0.11	
7	2	0.37	0.29	0.1		0.39	0.23	
8	2	0.26	0.2	-0.39		0.25	-0.06	
9	2	0.81	0.78	0.32		0.83	0.04	
10	2	0.83	0.79	0.35		0.87	0.29	
11	2	0.78	0.66	-0.47		0.8	-0.17	
12	2	0.47	0.37	-0.13		0.53	0.43	
13	2	0.43	0.35	-0.02		0.46	0.24	
14	2	0.57	0.51	0.21		0.59	0.07	
15	2	0.57	0.51	0.24		0.62	0.27	
16	2	0.85	0.82	0.48		0.91	0.94	
17	2	0.88	0.82	0.19		0.89	-0.21	
18	2	0.84	0.79	0.31		0.87	0.15	
19	2	0.57	0.47	-0.37		0.57	-0.1	
20	2	0.24	0.24	0.62		0.28	0.72	
21	2	0.53	0.49	0.27		0.54	-0.07	
22	2	0.74	0.68	0.37		0.8	0.77	
23	2	0.82	0.75	0.38		0.86	0.5	
24	2	0.62	0.53	-0.23		0.6	-0.28	
25	2	0.65	0.59	0.39		0.63	-0.33	
1	8	0.94	0.87	-1.13	B	0.93	-1.26	B
2	8	0.81	0.75	0.11		0.84	0.13	
3	8	0.64	0.46	-1.2	B	0.56	-1.04	B
4	8	0.54	0.41	-0.61		0.5	-0.76	
5	8	0.51	0.44	0.32		0.47	-0.65	
6	8	0.31	0.22	-0.47		0.28	-0.48	
7	8	0.74	0.62	-0.35		0.76	-0.03	
8	8	0.82	0.78	0.38		0.84	0	
9	8	0.76	0.69	0.17		0.81	0.34	
10	8	0.44	0.39	0.02		0.46	0.19	
11	8	0.7	0.58	-0.64		0.67	-0.58	
12	8	0.45	0.39	0.08		0.45	-0.13	
13	8	0.59	0.51	0.08		0.6	-0.16	
14	8	0.27	0.21	-0.36		0.28	0.14	
15	8	0.66	0.59	0.07		0.64	-0.57	
16	8	0.6	0.52	-0.25		0.61	-0.07	
17	8	0.67	0.62	0.34		0.69	-0.01	
18	8	0.76	0.66	-0.19		0.76	-0.42	
19	8	0.66	0.62	0.45		0.69	0.13	

Note: AIS (Average Item Score) is an indicator of item difficulty, calculated as the mean of the scored responses to the item.

Table A4

DIF Results by Item for Form 3 Critical Reading (Standard and Double Time)

Item	Section	Nondisabled	LD 2x			ADHD 2x		
		AIS	AIS	MH D-DIF	Category	AIS	MH D-DIF	Category
1	2	0.92	0.93	1.5	B	0.94	0.55	
2	2	0.67	0.54	-0.56		0.56	-1.96	C
3	2	0.81	0.73	0.16		0.87	1.04	
4	2	0.46	0.36	-0.1		0.5	0.07	
5	2	0.3	0.21	-0.57		0.28	-0.52	
6	2	0.27	0.27	0.46		0.25	-0.58	
7	2	0.23	0.21	0.56		0.27	0.16	
8	2	0.2	0.18	0.13		0.25	0.55	
9	2	0.88	0.8	-0.43		0.86	-1	
10	2	0.83	0.71	-0.75		0.85	0.19	
11	2	0.73	0.62	-0.51		0.73	-0.24	
12	2	0.39	0.38	0.41		0.35	-0.77	
13	2	0.69	0.63	0.14		0.72	0.19	
14	2	0.8	0.74	0.14		0.78	-0.51	
15	2	0.33	0.34	0.75		0.36	0.08	
16	2	0.64	0.5	-0.49		0.66	-0.14	
17	2	0.52	0.47	0.27		0.61	0.65	
18	2	0.52	0.42	-0.44		0.5	-0.49	
19	2	0.63	0.59	0.27		0.68	0.45	
20	2	0.77	0.71	0.27		0.78	-0.32	
21	2	0.73	0.59	-0.48		0.73	-0.29	
22	2	0.64	0.54	0.24		0.72	0.87	
23	2	0.75	0.63	-0.16		0.75	-0.43	
24	2	0.77	0.61	-0.64		0.78	-0.12	
25	2	0.68	0.59	0.36		0.66	-0.69	
1	8	0.94	0.91	0.17		0.96	0.89	
2	8	0.82	0.75	0.02		0.87	0.71	
3	8	0.87	0.79	0.02		0.93	2.08	B
4	8	0.61	0.5	-0.32		0.6	-0.45	
5	8	0.52	0.42	0.26		0.52	-0.71	
6	8	0.22	0.19	0.03		0.22	-0.35	
7	8	0.66	0.61	0.22		0.62	-0.71	
8	8	0.54	0.44	-0.24		0.5	-0.85	
9	8	0.51	0.43	-0.12		0.53	-0.13	
10	8	0.41	0.32	-0.42		0.41	-0.28	
11	8	0.53	0.42	-0.37		0.61	0.64	
12	8	0.37	0.3	-0.27		0.31	-1.07	B
13	8	0.48	0.45	0.31		0.6	1.11	B
14	8	0.22	0.17	-0.11		0.23	-0.26	
15	8	0.78	0.63	-0.73		0.75	-0.86	
16	8	0.74	0.61	-0.62		0.71	-0.77	
17	8	0.68	0.56	-0.06		0.74	0.84	
18	8	0.52	0.44	0.02		0.56	0.18	

Note: AIS (Average Item Score) is an indicator of item difficulty, calculated as the mean of the scored responses to the item.

