CollegeBoard
connect to college success™

The Impact of Extended Time on SAT® Test Performance

**Ellen B. Mandinach, Brent Bridgeman,
Cara Cahalan-Laitusis, and
Catherine Trapani**

www.collegeboard.com

The Impact of Extended Time on SAT® Test Performance

Ellen B. Mandinach, Brent Bridgeman,
Cara Cahalan-Laitusis, and Catherine Trapani

Acknowledgments

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 5,000 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 #050481688) 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.

Printed in the United States of America.

Contents

Abstract

The effects of extended time on SAT Reasoning Test™ performance are examined. The study explored the impact of providing standard time, time and a half (1.5 time) with and without specified section breaks, and double time without specified section breaks on the verbal and mathematics sections of the SAT®. Differences among ability, disability, and gender groups were examined. Results indicated that time and a half with separately timed sections benefits students with and without disabilities. Some extra time improves performance, but too much may be detrimental. Extra time benefits medium- and high-ability students but provides little or no advantage to low-ability students. The effects of extended time are more pronounced for the mathematics sections of the SAT. The implications for potential changes to the SAT and the need for future research are discussed.

Keywords: Extended time, Accommodations, Speededness, SAT Reasoning Test, Learning Disabilities

Introduction

The Issue

One of the most contentious debates in the field of testing currently is the potential impact of extended time and speededness on performance on standardized tests. The debate has taken place in the professional literature (Camara, Copeland, and Rothschild, 1998; Huesman and Frisbie, 2000; Munger and Loyd, 1991; Zuriff, 2000) and the public media (Heyboer and McCarron, 1999; Weiss, 2000a), and is closely linked to issues concerning the rights of test-takers with disabilities to reasonable and appropriate test accommodations (Mandinach, Cahalan, and Camara, 2002; Pitoniak and Royer, 2001). The primary issue is the extent to which the provision of extra time on standardized admissions tests may or may not impact performance.

According to Evans (1980), admissions tests are not intended to be speeded. Time limits are imposed for practical and efficiency reasons. Donlon (1984) notes that speed of response on the SAT should play only a minor role. The guidelines specified by Educational Testing Service (ETS) consider a test unspeeded if virtually all tests-takers answer at least 75 percent of the items, and 80 percent of the test-takers reach the last question (Swineford, 1974). As Nunnally (1978) notes, these rules do not address rushing and other suboptimal behavior in response to time pressure. Recent analyses (College Board, 1997, 2002) indicate that completion rates and response patterns fall well within these parameters. By this definition, the SAT Reasoning Test is an unspeeded test. Yet, it is still possible for speededness to affect performance differentially among groups of test-takers. As Munger and Loyd (1991) note, "When time limitations affect examinees' scores, a test is considered to be a partially speeded instrument, and scores are determined by the number of items attempted as well as the accuracy of responses" (p. 54). The research reported here, in conjunction with another study (Bridgeman, Trapani, and Curley, 2004), attempts to explicate some of the highly complex issues related to timing, speed, and differential test performance.

The impact of timing generally is studied two ways. The first method is what we refer to as the examination of speededness. A test has a specific number of items and a time limit. Decreasing the number of items but keeping constant the testing time allows researchers to determine the impact of speed of response. Respondents may reach more items, a higher percent may complete the test, or test scores may be affected either positively or negatively as indications of the effects of speededness. The second method is to examine the effects of extended time. The test remains constant, but students are given additional time in which to work through the test. There also are hybrid models of these two methodologies. In real and operational settings, it is often more practical to invoke the first method if decreasing the number of items does not interfere with the technical properties of the tests. The second method presents a myriad of practical issues and constraints, given that an extended time SAT requires students to test for 4.5 to 6 hours at one sitting. Additionally, simply because extended time may be available does not mean that examinees use any or all allotted extra time. In operational settings, many test-takers who have been granted extended time leave at the same time as the standard test-takers (J. Zumoff, personal communication, April, 1999).

Extended Time, Speededness, and Accommodations

The diversity of the literature on the effects of timing on test performance makes generalizations across testing situations difficult. Studies differ across age of respondents, methodology, type of test manipulation (e.g., speededness versus extended time), content areas, and type of test (i.e., admissions, school classification, achievement test, etc.). Many of the studies compare the performances of students with and without disabilities. Further, the operationalization of disability differs substantially across studies.

Few studies have focused solely on the impact of time on standard administrations (Bridgeman et al., 2004; Wild and Durso, 1979). Bridgeman et al. report in a recent study on the impact of speededness on the SAT

that the verbal section was only slightly speeded, affecting performance by less than 10 points on the SAT scale. The mathematics section was slightly more speeded, impacting performance by approximately 20 points. This effect was more pronounced for high-ability students. Decreasing speededness provided absolutely no benefit to low-ability students. The provision of extra time to students who do not have effective solution strategies will not prove beneficial. Further, decreasing the speededness of the test had no differential impact for gender or ethnic groups. Wild and Durso (1979), using data from the Graduate Record Examination (GRE), examined the impact of extra time on the performance of gender, ethnic, and age groups. Small score gains were found for all groups, but the provision of extra time did not differentially help any particular group of test-takers.

The most pervasive trend in research on differential performance and the impact of extended time can be found in the literature on disabilities, particularly learning disabilities. Test-takers with disabilities are entitled to test accommodations that provide equal access, remove irrelevant sources of difficulty, or in common parlance, attempt to level the playing field. According to the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999), the goal of an accommodation is "to minimize the impact of the test-taker attributes that are not relevant to the construct that is the primary focus of the assessment" (p. 101). Requests for accommodations, by law, are to be examined on a case-by-case basis. This is a daunting task given that every test-taker's disability differs in form and severity and may require quite different accommodations, creating almost infinite disability by accommodation combinations.

One of the most widely used accommodations is the provision of extended time, particularly for test-takers with learning disabilities. Extended time is intended to compensate for the slower processing speed of test-takers with disabilities. The objective of extended time as an accommodation is to have the same proportion of students with disabilities finish a standardized test as students without disabilities (Ragosta and Wendler, 1992). This way, neither group of test-takers attains an undue advantage over the other group. The appropriate amount of extra time has been estimated to be between 1.5 time (4.5 hours) and double time (6 hours) for the 2001 version[1] of the SAT. Examinees taking a standard administration were provided 3 hours of testing time for the seven sections with specified breaks. Each of the seven sections were separately timed in a standard administration. (In 2001 the SAT consisted of two 30-minute math sections,

two 30-minute verbal sections, one 15-minute math section, one 15-minute verbal section, and one 30-minute equating section [either math or verbal], totaling 3 hours of standard testing time.) In contrast, examinees taking an extended-time administration were given an open block of time, typically 4.5 hours, but no breaks between sections were given. The rationale for having no set section breaks was to allow the student to work through the test, using the extra time provided on the sections most affected by the person's disability, which may be more severe on the math than the verbal, or vice versa.

There is no doubt that some students with disabilities require more time on tests such as the SAT. Packer (1987) attempted to provide baseline figures for the amounts of time required by specific groups of test-takers with disabilities and specific accommodations. Students taking a Braille test took the most time, followed by visually impaired students using cassettes, students with physical or learning disabilities and, finally, hearing-impaired students who took a regular test. Packer also examined the relationship between time and test performance. Results indicated that students who took more time, regardless of their disability, tended to obtain higher scores on the math sections. In contrast, on the verbal section, visually impaired and physically disabled students who devoted more time to the test received lower scores, whereas for students with hearing impairments or learning disabilities, the more time spent on the test, the better their performance. It is important to note that these findings are correlational, not causal, and that these analyses were conducted on an old version of the SAT that contained five sections to be completed in 2.5 hours. Wright and Wendler (1994) conducted a field test on the then new SAT to try to establish timing estimates. Some accommodations of the test required more time than other formats. Results for the verbal section indicated that accommodations using cassettes were the least speeded, whereas large type was the most speeded. The math section, regardless of format, was more speeded than the verbal section.

Phillips (1994) proposes five criteria that determine the technical appropriateness of an accommodation. The criterion that relates most closely to extended-time policies and is most hotly debated, is that nondisabled examinees should not benefit from the same accommodation that assists examinees with disabilities (Elliott and McKevitt, 2000; Mandinach et al., 2002). There would be no disagreement, of course, that Braille is a necessary accommodation for some visually impaired examinees but that a sighted examinee would not benefit from a Braille test. The rationale for providing extended time is different: The extra time is intended to compensate test-takers with specific time-related disabilities for their particular type

---

[1]A revised version of the SAT, with different time limits, was introduced in spring 2005. Data collection for this study occurred in 2001 and was designed around what were then the standard specifications for the SAT. For the purposes of this report, we are referring to the SAT I: Reasoning Test (which is now known as the SAT Reasoning Test).

of disability. If extra time also enhances the performance of nondisabled test-takers to the same extent, there is no relative compensation to the test-taker with a disability according to the Phillips criterion. It would go against the principle of providing equal access and leveling the playing field.

Elliott and McKevitt (2000) argue that the Phillips criterion is wrong. They assert that an accommodation such as extended time can benefit both disabled and nondisabled examinees without sacrificing the validity of the accommodation or test. They recommend the use of single-subject designs that go beyond simple comparisons between nondisabled and disabled groups due to the substantial heterogeneity of the diagnoses, disabilities, and accommodations.

Zuriff (2000) examines the Phillips criterion with reference to a perspective he refers to as the Maximum Potential Thesis. According to Zuriff, students who do not have a learning disability perform at their maximum potential under timed conditions, and therefore they should not benefit from additional time. He believes that nondisabled test-takers do not suffer from a lack of time. In contrast, students with learning disabilities should benefit from extra time because the timing constraints do not allow them to perform up to their potential. Zuriff conducted a meta-analysis of a small number of studies and found that results were inconsistent and failed to confirm Phillips's position. Although students with learning disabilities often improved performance with extended time, students without disabilities also improved.

More important, other studies fail to provide evidence that tests are nonspeeded for nondisabled students. For example, Mehrens, Millman, and Sackett (1994) obtained results that indicated that nondisabled test-takers on the Multistate Bar Examination significantly improved their scores with extended time. Hill (1984) compared examinees with and without learning disabilities on the ACT under timed and untimed conditions. Results indicated that under timed conditions, students with learning disabilities scored significantly lower than those without disabilities. There were no differences found between the groups in the untimed condition. The study also examined performance on the Nelson-Denny Reading Test and found that nondisabled students improved test scores under untimed conditions, also contradicting Zuriff's MPT assumptions. Halla (1988) found that performance of students both with and without disabilities improved under untimed conditions on the GRE and the Nelson-Denny Reading Test.

Huesman and Frisbie (2000) found that the amount of extra time needed by students with learning disabilities varied greatly, from those needing no additional time to others who required more than twice the standard time. They noted that many of the students with disabilities finished within standard time limits, a finding that is often reported by ETS test center supervisors (Mandinach et al., 2002). Results differed for the nondisabled students dependent upon the instructions they were given. It is clear that the instructions given to examinees can impact their test-taking strategies and ultimately their performance. When told to work carefully and take their time, the test performance of the nondisabled students improved with extra time. No differences were found between standard and extended-time conditions when the nondisabled students were told to work at their normal rate. Munger and Loyd (1991) found no differences in test performance between students with and without disabilities on timed and extended-time administrations of the Iowa Test of Basic Skills. Small but equivalent score increases were found for both groups in the untimed testing conditions on the Language Usage and Expression and the Mathematics Concepts tests. Discriminant analysis failed to differentiate the groups based on the number of items attempted or completed.

Camara et al. (1998) examined gains from retests of the SAT in which the first test was administered with standard time and the second test was administered with extended time. Students with learning disabilities were found to score half a standard deviation below those without disabilities on the SAT. The gains achieved by students with disabilities were three times as large as the gains attained for nondisabled students taking two standard administrations or those with learning disabilities taking two tests with extended time. Providing retests for examinees with disabilities improved performance an estimated 32 and 26 points respectively on the verbal and math sections of the SAT. The amount of gain was found to be proportional to the amount of extra time used by the student.

Runyan (1991) examined the impact of extended time for university students on the Nelson-Denny Reading Comprehension Test. Under timed conditions, students without learning disabilities performed better than examinees with learning disabilities. No significant differences were found between the groups when given extended time. Scores of the students without disabilities did not improve significantly with extended time. Runyan noted that the amount of extended time used by students with disabilities varied substantially. The conclusion drawn from the results was that because students with disabilities are so heterogeneous, providing fixed amounts of extra time may not be an appropriate method by which to assign an accommodation.

Results have been found to differ for the same students dependent upon the content of the test. Weaver (1993) examined the performance of college students on vocabulary and reading comprehension tests and found that performance for students with learning disabilities improved in the extended-time and untimed conditions.

According to Weaver, the extra time allowed them the opportunity to process the needed information and demonstrate their knowledge. Weaver concludes that timing constraints masked their ability to perform on both tests. Students without disabilities, however, improved only marginally with extended time on the vocabulary test but improved in the extended-time or untimed conditions on the comprehension test.

There are some logical explanations for the findings on vocabulary and reading comprehension tests that can be generalized to the current study. Extra time does not always produce improved performance. For example, on a vocabulary test, if students do not know the material, such as a particular word, extra time probably will not help. On a reading comprehension test, in contrast, the problem may be that test-takers simply cannot get to the material due to lack of time. If given extra time to reflect on a reading passage, there may indeed be a positive effect on test scores. Hence, for the SAT, extended time may differentially impact performance on specific item types.

Marquart (2000) examined not only the impact of extended time on outcomes, but also on test-taking strategies for students with disabilities, students functioning at or above grade level, and students at risk academically. Although the results showed no significant differences among three groups of students in terms of scores on extended-time versus standard-time administrations, Marquart found that extended time changed the way students approached the test. Students reported that they were more positive, felt more relaxed, were more likely to answer every question, and were more motivated under extended-time conditions. Marquart's conclusion was that extended time stimulated students to use better test-taking strategies, and that these in turn may reduce test anxiety. A ceiling effect noted in this study makes it difficult to discern the effects of extended time on test performance. Students finished the standard-time and extended-time conditions with time to spare, and therefore were not likely to experience the same sort of time constraints as in time-pressured situations.

In sum, claims about the impact of extended time are varied and controversial. It is clear, however, that more students with disabilities are applying to and attending college than ever before (Greenberg, 2000; Henderson, 1999; Lee, 2000). School counselors and admissions officers report large increases in applicants with learning disabilities (Mandinach et al., 2002). Applicants with disabilities are entitled to appropriate and reasonable accommodations on standardized admissions tests, assuming that they have provided the required documentation (College Board, 2001). The accommodation most often requested and granted is extended time. There is, however, substantial concern about the demographic disparities in requests for extended time for the SAT (Leatherman, 2000; Mandinach et al., 2002; Weiss, 2000b). Students from certain ethnic groups (black, Hispanic, Asian, and Native American), geographical regions (the South and Midwest, as well as rural and urban areas), and from low SES families are less likely to seek accommodations than are test-takers from the Northeast Corridor and California, suburban schools, and from educated families with higher incomes. In response to the recent media attention about the potential abuse of extended-time accommodations implied by these disparities, the president of the College Board (Caperton, 2000) wrote an impassioned letter to the *Los Angeles Times*, indicating the position of his organization:

> The lesson is not that new, more-rigid barriers (tougher medical documentation and the like) should be raised against disabled students who require extra time to take the SAT. Instead, even while we stop abuse by a relative few, we must also take steps to raise the awareness of students, parents, counselors, and administrators to the accommodations that are available to learning-disabled youngsters. Let's throw open the door of college admission to all students who truly qualify for a higher education.

Many of the admissions officers and school counselors interviewed and surveyed in Mandinach et al. (2002) expressed a pressing need for research on the impact of extended time, not just for students with disabilities, but also for those without disabilities, in an effort to ascertain the validity of tests when extra time is provided. This study is, in part, a first step toward responding to that need.

Design

The study design permits the comparison of standard timing for the 2001 version of the SAT with longer tests that allow 1.5 and 2.0 times the standard time. The impact of providing specified section breaks also is examined. Section breaks (i.e., separately timed sections) were used in standard administrations but were not provided in extended-time administrations in 2001. Effects of extended time on both verbal and mathematical scores are evaluated in both disabled and nondisabled groups.

The disabled group consisted of students identified as having a learning disability (LD) and/or Attention-Deficit/Hyperactivity Disorder (ADHD). Differential performance trends for ability and gender groups also are examined.

Participants

Participants in the study were recruited from high school juniors who had taken the 2000 Preliminary SAT/National Merit Scholarship Qualifying Test (PSAT/NMSQT®) and planned to take the May or June 2001 SAT. We identified about 500 high schools that had a high incidence of students who had applied for testing

accommodations with diagnoses of LD and/or ADHD. This resulted in a disproportionate number of schools in the Northeast and Mid-Atlantic regions because of the concentration of LD and ADHD students in those states. The high schools represent 19 states (categorized by the College Board regions: 13 in the Northeast, 32 in the Mid-Atlantic, 18 in the South, 3 in the Midwest, 1 in the Southwest, and 5 in the West). Invitations to participate were sent to the 500 high schools in an attempt to obtain their participation, and 72 volunteered to participate in the study. Each participating high school was then sent a packet of invitations that were distributed to students with and without disabilities. In the sample, there were 47 public, 13 independent, and 12 religious schools. The schools indicated that between 2,500 and 3,200 students were interested in participating in the study.

Methods

Students were given three incentives to participate in the study: (a) a free practice test; (b) constructive feedback on correct and incorrect answers; and (c) the opportunity to win a cash award. Students were told that if their performance on the experimental test was as high as or higher than predicted from their PSAT/NMSQT scores, then they would be eligible for a drawing for a $250 prize. Ten prizes were awarded. Honoraria were given to the test center supervisors of the participating high schools. A letter informing parents about the study was sent home, and signed parental consent forms were filed with the school and with ETS.

The window for the test administrations was one week in April 2001. Each school determined when to give the test. The schools needed at least three testing rooms, given the different timing conditions (see below—Group 1, Group 2, and Groups 3 and 4). Color-coded testing books were produced for the three rooms. Different sets of instructions were generated for each of the conditions.

Following the administrations, the test booklets were left at the schools, and answer keys were sent to the schools so students could compare their responses to the correct answers. Letters of explanation were sent to the participants along with a breakdown of their responses in terms of rights, wrongs, and omits.

The Tests and Procedures

Test sections were taken from an SAT form that was disclosed but that was not widely distributed in test-preparation books or software. A standard SAT answer sheet was used. Modifications were made to the administrators' manual to explain the different timing conditions.

All experimental tests were created by taking sections from a single base form. The base form contained all of the sections of a standard 2001 version of the SAT except for the equating section: three verbal sections (V1, V2, and V3) and three math sections (M1, M2, and M3). In a standard administration, the time limit for each section is 30 minutes except for V3 and M3, which are 15-minute sections. V1 consisted of 10 sentence completion questions, 13 analogy questions, and 12 questions based on a single reading passage. V2 consisted of 9 sentence completion questions, 6 analogy questions, and 15 questions based on two reading passages, and V3 consisted of 13 questions based on a pair of reading passages. M1 contained 25 standard five-choice math questions, and M2 contained 15 quantitative comparison questions (comparing which of two quantities is larger, whether they are equal, or whether the relationship cannot be determined) and 10 student-produced response questions in which the examinee enters a numerical answer on a special grid rather than choosing a correct answer. M3 was 10 standard five-choice questions.

Four test books were produced with sections as follows:

1) V1, M1, V2, M2, V3, M3;
2) M1, V1, M2,V2, M3, V3;
3) M1, V1, M2, V2; and
4) V1, M2, V2.

Two versions of each book were produced to randomize the order of the specific sections (see Table 1). Group 1 received a standard administration for all but the fourth

Table 1

Timing Conditions (in Minutes) by Experimental Group and Test Order Spiral

Group	V1	M1	V2	M2	V3	M3	Total
Test Order and Time for Spiral A							
1. Standard time (for first four and last sections)	30	30	30	30	30	15	165
2. 1.5 time, separately timed sections	45	45	45	45	NA	NA	180
3. 1.5 time, no separate section times	180				NA	NA	180
4. Double time, no separate section times	180			NA	NA	NA	180
Test Order and Time for Spiral B							
1. Standard time (for first four and last sections)	30	30	30	30	30	15	165
2. 1.5 time, separately timed sections	45	45	45	45	NA	NA	180
3. 1.5 time, no separate section times	180				NA	NA	180
4. Double time, no separate section times	180			NA	NA	NA	180

Note: Spirals A and B are identical in content. The order of presentation for the sections was randomized.

section which represented double time. They were tested in one room for 2 hours and 45 minutes. Group 2, in the second room, received four sections and was allowed 3 hours of test time. This was the equivalent of a 1.5-time condition. The test that Group 2 was given contained separately timed sections. Group 3 also received four sections and 3 hours of test time, or the equivalent of 1.5 time, but they received a test without separately timed sections. This was the same test that was given to Group 2 but without specified section breaks. Group 4 was given three sections without section timing, the equivalent of a double-time condition. Groups 3 and 4 were either combined into a third room or were tested in separate rooms. Students who were assigned to Groups 3 and 4 also were asked to keep a record of the time spent on each of the sections by entering start times and stop times on a log that we provided. All students were requested to fill out a short questionnaire at the conclusion of the test. The primary intent of the timing conditions was to compare the effects of 1.5 time or double time on performance, compared to standard time. The second goal was to examine the impact of section breaks, given that extended-time accommodations do not include specified section breaks.

Participants were randomly assigned to the four testing groups by staff at ETS prior to the testing dates. Students were distributed across the two test books containing different section orders within each timing condition, accounting for eight total testing spirals. School counselors were asked to identify students with learning disabilities, if possible. These students were also randomly assigned to one of the eight spirals. Room and testing assignments were sent to each school prior to the testing day with specific instructions about which students should receive which test books in which room.

Results

Motivation Screening

Scores on the test used in the study were not reported to colleges; thus, it was possible that some students would not take the experimental test seriously. We used two methods to screen out such students. From the total sample of 2,088 students who had completed the test, had PSAT/NMSQT scores, and had no physical handicap that might interfere with test performance, we first compared results on the experimental test with scores from the operational administration of the PSAT/NMSQT that the students had taken six months prior to the experimental test.

Scores on the PSAT/NMSQT and the experimental test were converted to z-scores in the sample of students taking both tests. If a student's z-score on *either* the math or the

verbal sections of the experimental test was 1.5 or more points lower than the z-score on the corresponding math and verbal sections on the PSAT/NMSQT, the student was removed from the analysis. Also, if the scores on *both* the verbal and math sections of the experimental test were 1.0 points lower than PSAT/NMSQT scores, the student was removed. These cut scores were selected to remove only the most extreme score declines. In a national sample of students who took the PSAT/NMSQT and the SAT, declines of this size would be expected in a fraction of 1 percent of the sample. This resulted in the removal of fewer than 3 percent of the tested students.

An additional 113 students were removed based on their responses to the question on the exit questionnaire that asked, "How much effort did you put into doing well on this test?" Answer choices were: (A) None, did not try to do well; (B) Some effort; (C) Moderate effort; (D) Strong effort; or (E) Very strong effort. The students who selected A or B were removed from the sample, resulting in a final sample of 1,929.

Of the 1,929 students in the final sample, 264 were classified as learning disabled and/or ADHD based on having formally requested testing accommodations, or on a report from the school counselor. The composition of the sample is presented in Table 2. Because work on a prior study suggested that effects of extra time differed across ability groups, examinees were divided into three ability groups based on PSAT/NMSQT scores. In the nondisabled group, the low group was defined as students with PSAT/NMSQT math scores of 45 and below, the middle-ability group had scores ranging from 46 to 55, and the high-ability group had scores greater than 55. These cuts would divide the national sample roughly into thirds. Because the current sample was generally above average in ability, there were about twice as many students in the high-ability group as in the low-ability group. Because only 36 students in the LD sample scored in the highest ability group (PSAT/NMSQT score above 55), the two highest categories were collapsed into a single category. Thus, the lower-ability group was 45 and below, and the higher group was 46 and above. Sample sizes for

Table 2

Composition of the Sample by Timing Condition, Disability Status, and Ability Group

	Disability Status				
	LD		Non-LD		
Timing Condition	Lower Ability	Higher Ability	Low Ability	Medium Ability	High Ability
Regular	38	28	92	146	187
1.5 With Sections	31	30	87	140	196
1.5 No Sections	42	26	87	154	194
Double V1, M1, V2	23	14	49	85	69
Double M1, V1, M2	22	10	40	64	75

analyses that examined effects on scores from both math sections (M1 and M2) or both verbal sections (V1 and V2) were smaller because students in the double-time conditions took either two math or two verbal sections, but not two sections of each.

Mathematics Results for the Nondisabled Sample

Preliminary analyses indicated no significant interactions of gender groups with experimental conditions (Fs less than one), so gender groups were combined to simplify data presentation. Scores from M1 and M2 were added together to form a total math score. Consistent with scoring procedures on operational SAT tests, a correction for guessing was applied that penalizes incorrect responses more than omitted responses. Specifically, a formula score was created as follows: formula score = $R - W/(k-1)$ where R is the number of questions answered correctly, W is the number of questions answered with the wrong response, and k is the number of multiple-choice response options.

A 3 (ability) × 4 (experimental condition) ANOVA indicated the expected significant effect for ability, $F(2, 1450) = 1064.9$, $p < .0001$, a significant condition effect, $F(3, 1450) = 7.1$, $p < .0001$, and a significant ability by condition interaction, $F(6, 1450) = 2.5$, $p < .03$. The mean formula scores in each group and their associated 95 percent confidence intervals are summarized in Figure 1. Overlapping bars indicate differences that are not statistically significant. More conservative Scheffé tests within each ability group confirmed significant differences only in the middle-ability group; 1.5 time with section

timing was significantly higher than regular time. The difference of 4.2 formula score points for this comparison is equivalent to approximately 43 points on the SAT scale, estimated by dividing the formula score difference by the standard deviation in this sample (10.7) and multiplying by the national SAT-M standard deviation (110). In the lowest ability group, the means for regular time and 1.5 time with section timing were virtually identical (15.2 and 14.9, respectively). Although the mean in this low-ability group appeared to be somewhat higher for double time (17.1), the 95 percent confidence intervals for regular time and double time were overlapping. In the two higher-ability groups, the means for double time were actually less than the means for 1.5 time with section timing. M1 and M2 were also analyzed separately; effects in each section were consistent with the overall results.

Verbal Results for the Nondisabled Sample

For the verbal score (V1 + V2), gender once again did not interact with condition, so results were combined across genders. The 3 × 4 ANOVA indicated the expected ability effect, $F(2, 1474) = 999.2$, $p < .0001$ and a condition effect, $F(3, 1474) = 3.1$, $p < .03$, but the interaction was not significant, $F(6, 1474) = 1.4$, $p = .23$. Results are summarized in Figure 2. Although the interaction was not statistically significant, the pattern of the means was comparable to that for math scores, with virtually no effects in the lowest ability group and the largest difference in the middle-ability group. The difference of 3.3 formula score points in that group corresponds to a difference of about 29 points on the SAT
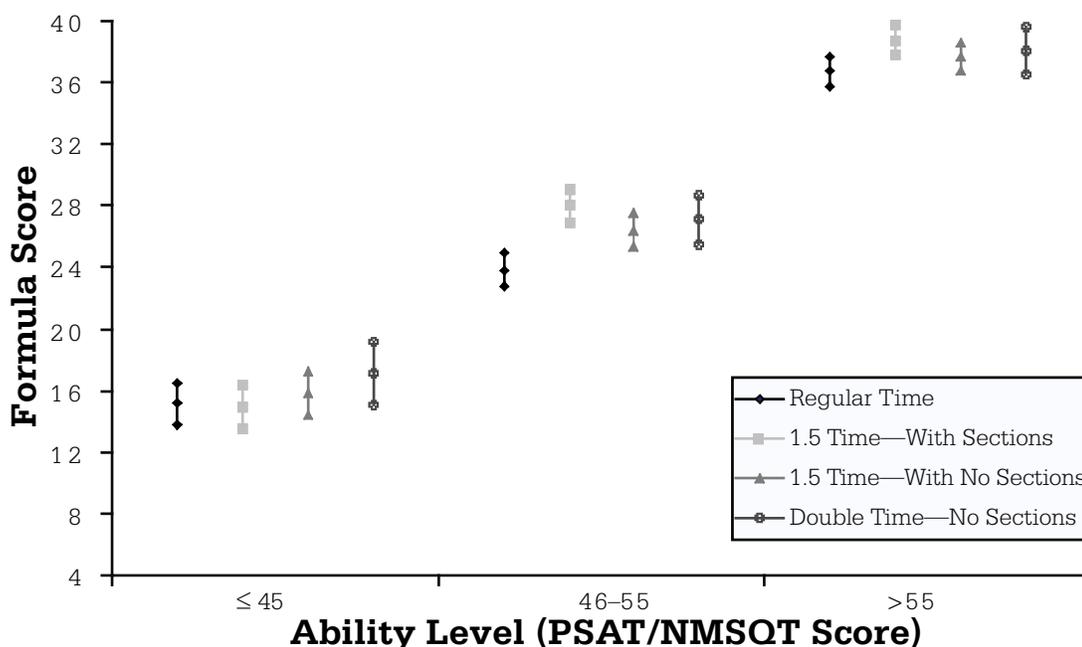
Figure 1. Means and 95 percent confidence intervals for math scores (M1 + M2) in the non-LD sample.
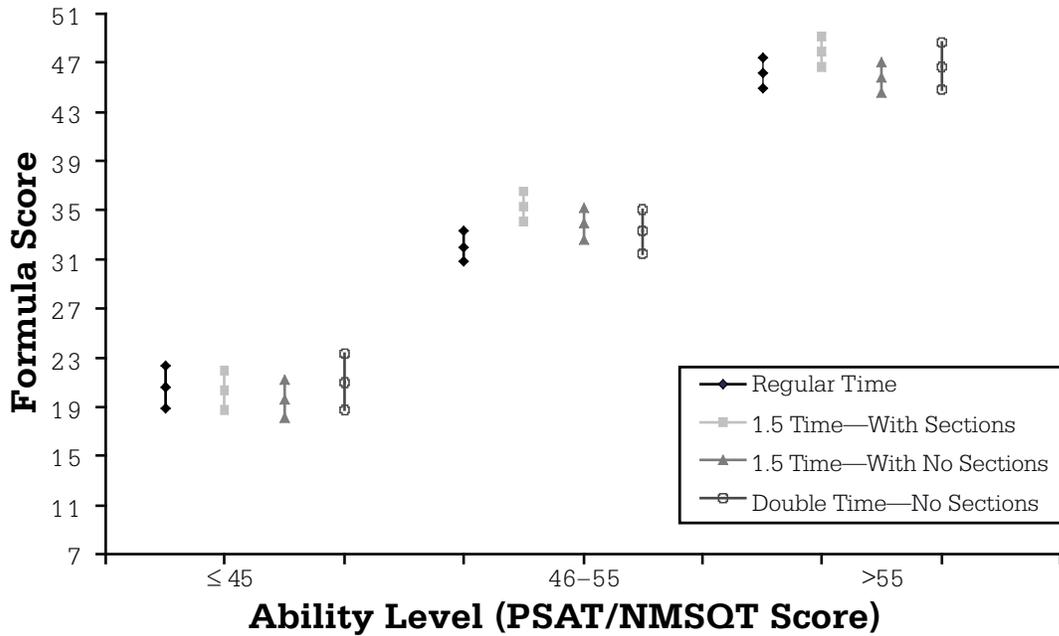
Figure 2. Means and 95 percent confidence intervals for verbal scores (V1 + V2) in the non-LD sample.

scale, while the difference in the high-ability group is about 18 points on the SAT scale. As was noted in the math results, the double-time group mean was lower than the mean for 1.5 time with section timing.

Mathematics Results for the Learning-Disabled Sample

Again, gender interactions were not significant, so gender groups were combined. Neither the condition nor the condition-by-ability interactions were significant, F (3, 219) = 1.0, p = .39; F (3, 219) = 1.7, p = .17. The means are summarized in Figure 3. Although the differences were not statistically significant in these relatively small samples, the 5 formula score point difference between standard time and 1.5 time with section timing in the higher-ability group is larger than the statistically significant difference noted in the non-LD group. Furthermore, the trivial differences in the lower-ability group are consistent with the findings for the non-LD group. Because over half of
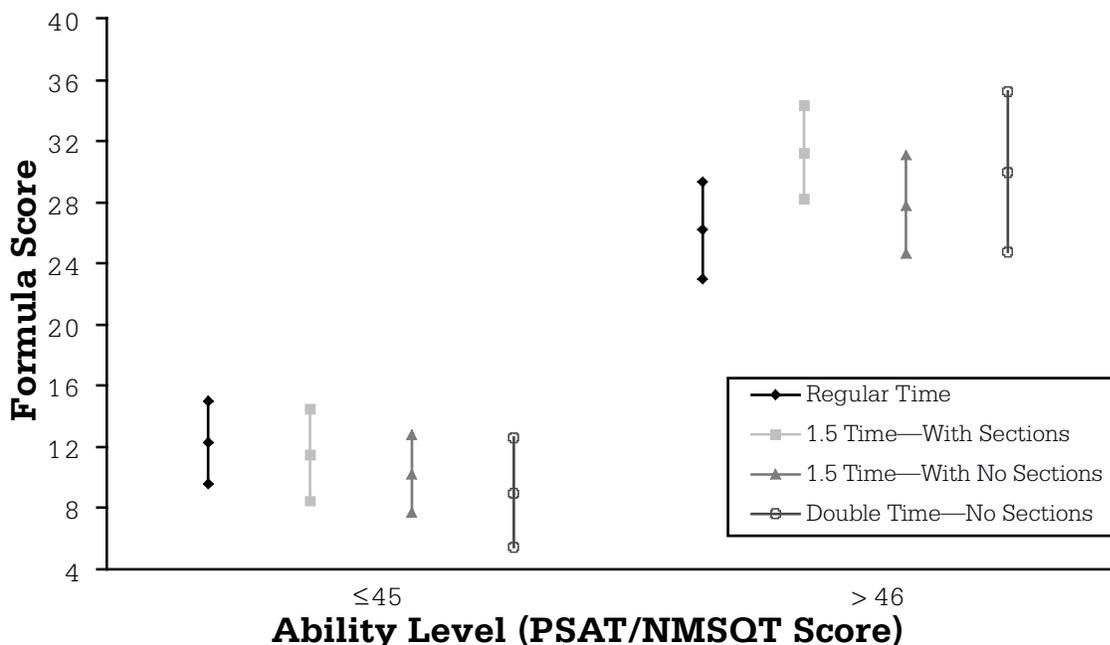
Figure 3. Means and 95 percent confidence intervals for math scores (M1 + M2) in the LD sample.

the LD sample was in the lower-ability group, in contrast to less than a quarter of the sample in the non-LD group, the overall impact of extra time for the entire sample of LD students was reduced.

Verbal Results for the Learning-Disabled Sample

As indicated in Figure 4, results for the verbal test were consistent with the results found in the non-LD sample, though there were no statistically significant differences in this smaller group. Again, extra time appears to be of little or no benefit for lower-ability students, and double time with no section timing is no better than 1.5 time with or without section timing. The difference of 5.7 formula score points (about 50 points on the SAT scale) between standard time and 1.5 time in the higher-ability group is larger than the differences noted in the non-LD sample.

Timing Records

Students who received the test with extended time (1.5 time or double time) but without breaks between sections were asked to complete timing records during testing. On the timing record, students recorded the time they started and stopped each section as well as the amount of time spent during the break between the second and third sections. Although all test-takers in the 1.5-time and double-time test conditions (without section breaks) were asked to complete timing records, 26 students in the 1.5-time condition and 20 students in the double-time condition did not fully complete timing records. In both

timing conditions, two of the students without timing records had learning disabilities or ADHD.

Use of Time for Nondisabled Students

Analyses of variance were computed to investigate average time used by ability (low, medium, and high) and testing condition (1.5 time without section breaks and double time without section breaks). Analyses were computed separately for verbal and math tests. A significant interaction (condition × ability) was found for M2, $F (2, 584) = 6.74$, $p < .001$, but not V1, V2, or M1. The interaction was primarily due to the changes in the amount of time used by high-ability students in the two timing conditions. High-ability students in the 1.5-time condition spent the most time (35.0 minutes), followed by medium- (34.4), and low-ability (31.9) students. In the double-time condition, high-ability students spent the least time (38.0), while medium- and low-ability students spent 44.0 and 39.9 minutes, respectively. On average, the high-ability students were more consistent between the two testing conditions (35 to 38 minutes) than the medium- and low-ability students (approximately 34 to 44 and 32 to 40).

Timing Condition

A significant effect by timing condition was found on all four subtests (M1, V1, M2, and V2) and indicated that students spent more time per subtest in the double-time condition than in the 1.5-time condition. As Table 3 indicates, students in the double-time condition spent an average of:
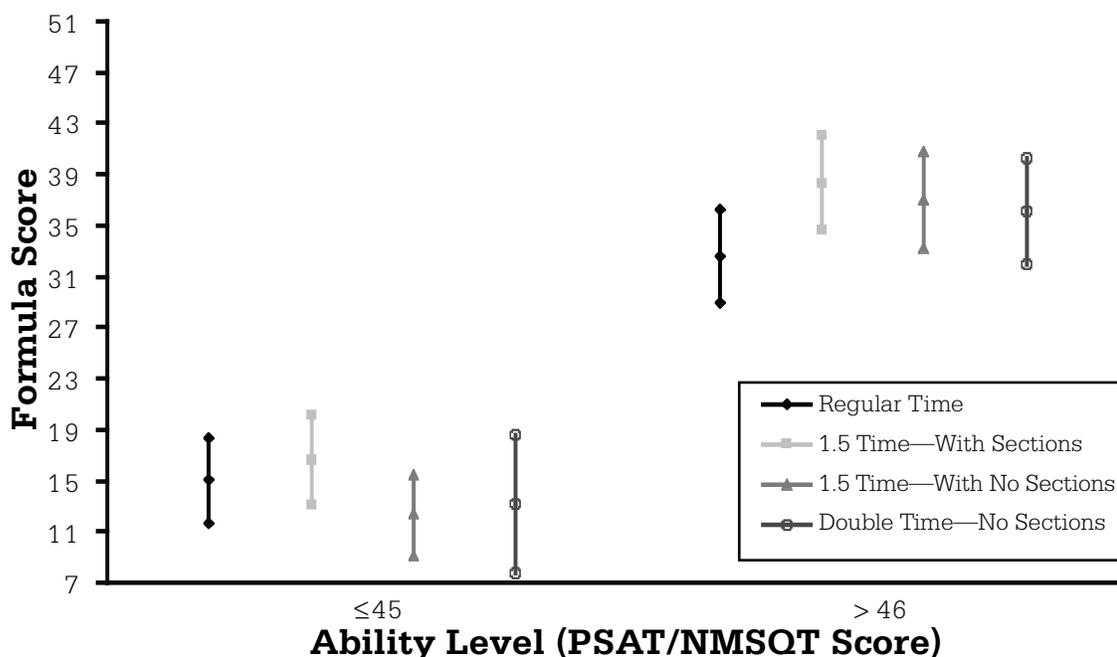
Figure 4. Means and 95 percent confidence intervals for verbal scores (V1 + V2) in the LD sample.

Table 3

Average Time Reported Used Per Subtest by Timing Condition and Disability Status

Subtest	1.5 Time (no breaks)			Double Time (no breaks)			Difference
	M	N	SD	M	N	SD	Double–1.5
Nondisabled							
M1	35.10	418	7.32	37.49	371	10.13	2.4
V1	33.15	418	7.93	34.99	371	9.16	1.8
M2	34.54	418	10.01	40.57	174	11.24	6.0
V2	30.47	418	9.11	34.19	197	10.15	3.7
Total	133.26	418	20.48	108.91	371	22.52	
LD and/or ADHD							
M1	35.15	66	9.73	39.04	67	13.92	3.9
V1	35.52	66	10.06	37.18	67	12.61	1.7
M2	28.97	66	9.70	35.00	31	12.16	6.0
V2	30.52	66	9.47	37.42	36	15.39	6.9
Total	130.16	66	26.86	111.73	67	28.90	

Note: Total time allowed was 180 minutes. Total time for the double-time group is lower than total time used by the 1.5-time group because fewer subtests were taken.

2.4 minutes more on M1 than students with 1.5 time, $F(1, 787) = 15.02$, $p \leq .001$;

1.8 minutes more on V1 than students with 1.5 time, $F(1, 787) = 8.52$, $p \leq .001$;

6.0 minutes more on M2 than students with 1.5 time, $F(1, 584) = 45.56$, $p \leq .001$; and

3.7 minutes more on V2 than students with 1.5 time, $F(1, 612) = 20.21$, $p \leq .001$.

It is possible that the differences in time spent on M2 and V2 are due to changes in content from M1 and V1. Although both M1 and M2 contain standard multiple-choice questions, M2 had fewer multiple-choice questions and additional questions with quantitative comparisons (QC) and student-produced responses (SPR). The QC and SPR items take longer than standard multiple-choice questions. Differences between V1 and V2 also exist; on V1 students received one set of reading comprehension questions, and on V2 students received two sets of reading comprehension questions. Since test-takers are less likely to complete verbal sections with two reading passages than one reading passage during operational administration, it is possible that the additional reading passage contributed to the increased time used on V2 compared to V1. The increased differences on the final two tests (V2 and M2) may also have been a result of students recording the time used to review responses on the entire test as the completion time for the last subtest. Because students in the double-time condition completed only three sections, V2 or M2 was the final subtest for the entire double-time sample but was the final subtest for only half of the sample in the 1.5-time condition. Figure 5 displays the percentage of students (by timing condition and subtest order) plotted against minutes remaining at the start of the final subtest. Although students in the double-time condition had substantially more time remaining for the final subtest, all students in both timing conditions had at least standard time (30 minutes), and a vast majority (98 to 100 percent)
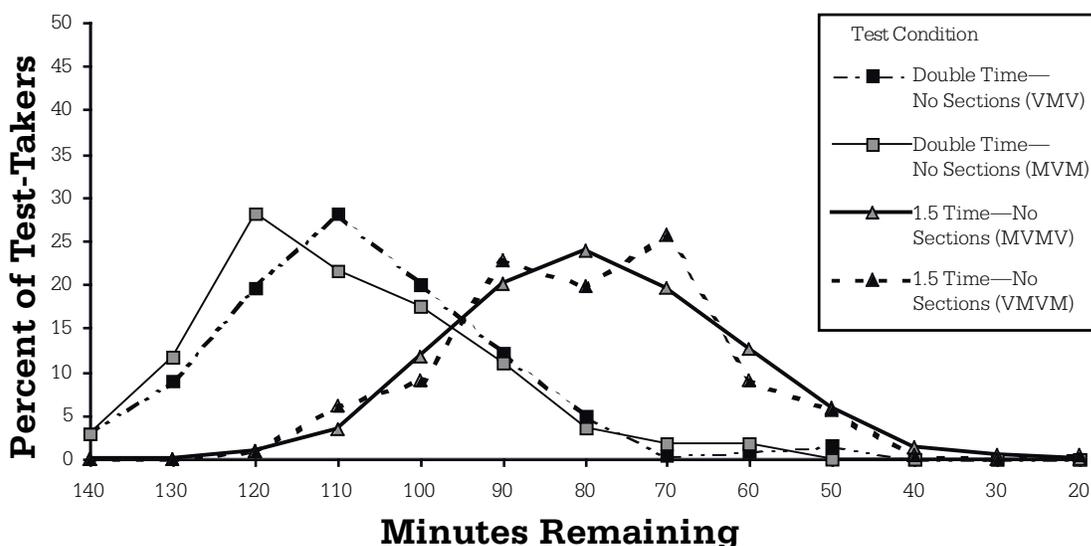
Figure 5. Percent of test-takers (without a disability) reporting "x" minutes remaining for last subtest.

Note: For the students in the double-time condition, the remaining time was calculated by subtracting the time used on M1 and V1 from the total time allowed (180 minutes), while remaining time for the 1.5-time condition was calculated by subtracting time used on M1, V1, and either M2 or V2 (depending upon the test order) from total time allowed. For ease of viewing, the amount of time remaining was grouped into 10-minute intervals.

Table 4

Average Time (in Minutes) Used Per Subtest by Ability Level and Disability Status

Subtest	High/Higher			Medium			Low/Lower		
	M	N	SD	M	N	SD	M	N	SD
Nondisabled									
M1	34.36	330	7.61	37.71	291	9.30	37.32	167	9.57
V1	32.51	302	7.58	35.06	295	9.18	34.77	192	8.79
M2	36.42	255	9.69	37.12	210	10.84	34.46	120	12.36
V2	30.19	230	8.77	32.37	229	9.47	32.84	154	10.74
LD and/or ADHD									
M1	41.10	51	12.88				34.63	82	11.02
V1	37.91	58	12.65				35.15	75	10.26
M2	34.70	33	12.20				28.94	64	9.62
V2	34.14	49	13.47				31.85	53	11.09

Note: Ability level is based upon PSAT/NMSQT score for the same subject (math or verbal).

of the students had at least 1.5 time (45 minutes) remaining for the final subtest.

Ability Level

Analyses of variance indicated a significant difference between the average time used by students with varying PSAT/NMSQT scores (high, medium, and low) on all four subtests. See Table 4 for descriptive statistics and Table 5 for subgroup differences and tests of significance. Scheffé tests indicated that on three of the four subtests (M1, V1, and V2), students with low PSAT/NMSQT scores used significantly more time than at least one of the other ability levels (medium or high). Low- and medium-ability students used approximately 3 minutes more than high-ability students for M1 and V1. On V2, low-ability students used approximately 3 minutes more than high-ability students but were not

significantly different from medium-ability students. The Scheffé test indicated a reversed pattern was found for M2, where medium-ability students spent about 3 minutes more than low-ability students, but were not significantly different from high-ability students. Figures 6 and 7 display the percentage of students (by ability level) plotted against minutes remaining at the start of the final subtest for the 1.5-time and double-time conditions (without section breaks), respectively. From these graphs it appears that ability level plays a smaller role than timing condition in determining how much time students use. In both conditions, however, the proportion of high- and medium-ability students with less than double time (60 minutes) remaining for their final subtest was lower than the proportion of low-ability students with less than double time remaining for the final subtest.

Table 5

Significance and Difference in Time (in Minutes) Used Per Subtest by Ability Level and Disability Status

Subtest	N	Ability			F	Scheffé
		L–H	M–H	L–M		
Nondisabled						
M1	789	2.96	3.35	-0.39	13.15***	L & M > H
V1	789	2.26	2.55	-0.29	7.65***	L & M > H
M2	585	2.65	2.18	0.47	3.91*	L > H
V2	613	-1.96	0.70	-2.66	3.80*	M > L
LD and/or ADHD						
M1	133	-6.46			9.27**	H > L
V1	133	-2.76			1.82	Not Significant
M2	97	-5.76			8.41**	H > L
V2	102	-2.29			0.05	Not Significant

Note: ***p <.001, **p <.01, *p <.05. L = Low/Lower-Ability, H = High/Higher-Ability, M = Medium-Ability.

Figure 6. Percent of test-takers (without a disability) in 1.5-time condition without breaks reporting "x" minutes remaining for last subtest, by verbal ability.

Figure 7. Percent of test-takers (without a disability) in double-time condition without breaks reporting "x" minutes remaining for last subtest, by verbal ability.
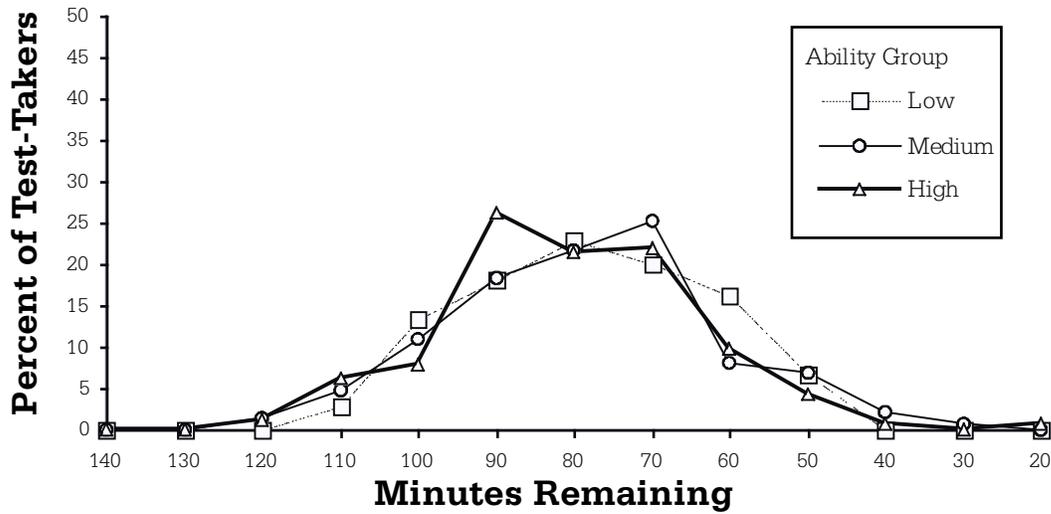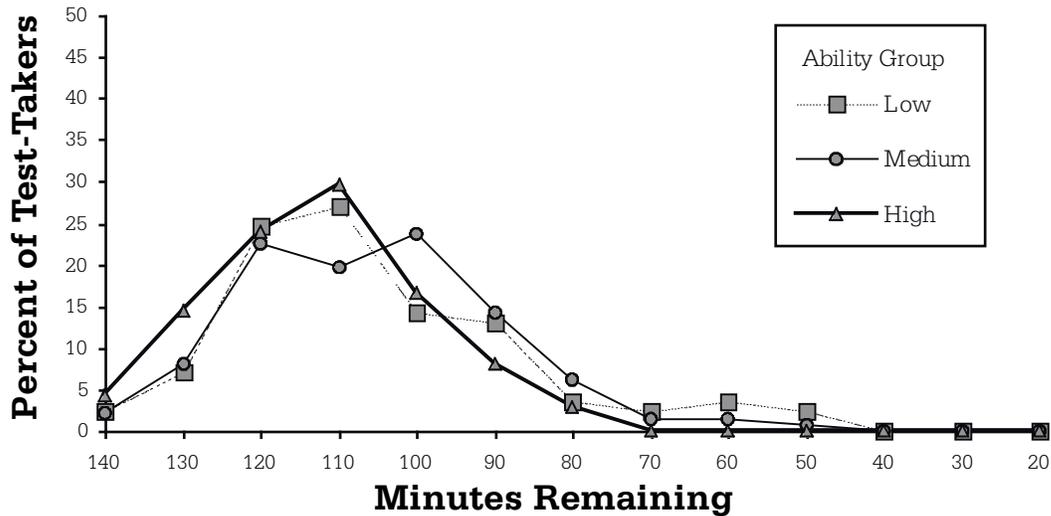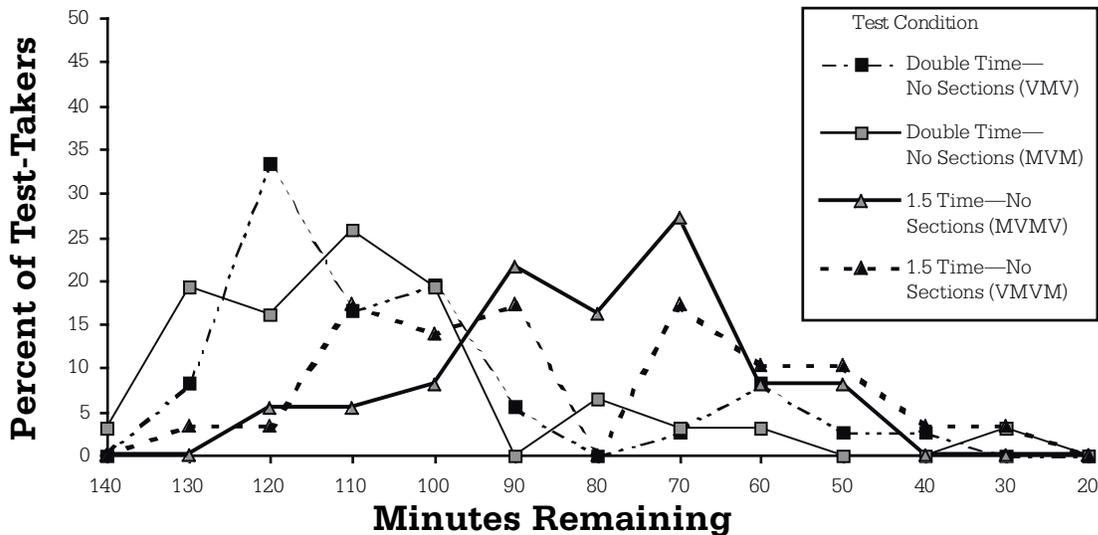

Figure 8. Percent of test-takers (with a learning disability and/or ADHD) reporting "x" minutes remaining for last subtest.

Use of Time for Students with Learning Disabilities and/or ADHD

Separate analyses were computed to investigate average time used by ability and by testing condition for students with a learning disability and/or ADHD. For these analyses the high- and medium-ability categories were collapsed into a single category because there were so few students that scored at the highest ability level (n = 20 for verbal ability and n = 23 for math ability). As with the nondisabled sample, analyses were computed separately for verbal and math tests. No significant interaction (condition by ability) was found for any of the subtests; F's were consistently under 1.0, ranging from .07 to .80.

Timing Condition

Analyses of variance in the average minutes per subtest for students with a learning disability and/or ADHD indicated that students in the double-time condition reported spending significantly more time (an average of 5.0 minutes more) per subtest than students in the 1.5-time condition, F (1, 132) = 10.56, $p < .001$. The difference between reported minutes used by students in each timing condition was only statistically different for one of the four subtests (V2) in these relatively small sample sizes. The reported differences, however, are similar to the statistically significant differences found in the larger sample of students without disabilities; on all four subtests, students in the 1.5-time condition spent less time than students in double-time condition. Students in the double-time condition used an average of:

3.9 minutes more on M1 than students with 1.5 time, F (1, 132) = 1.44, $p = .23$;

1.7 minutes more on V1 than students with 1.5 time, F (1, 132) = 0.02, $p = .90$;

6.0 minutes more on M2 than students with 1.5 time, F (1, 96) = 3.19, $p = .08$; and

6.9 minutes more on V2 than students with 1.5 time, F (1, 101) = 5.56, $p \le .05$.

Similar to the nondisabled sample, the disparity between minutes used in the two timing conditions was greatest on the last two subtests (M2 and V2). As with the nondisabled sample, this disparity may be due to a combination of three factors: changes in test content (e.g., additional reading passage, QC items, and SPR items) on M2 and V2, students reviewing the entire test during the final timing period, and

students in the double-time condition taking only three subtests rather than four (M1 and V1, followed by M2 *or* V2). Figure 8 displays the percentage of students plotted against minutes remaining at the start of the final subtest. Because of the smaller sample size, this graph is not as smooth as the graph of time used by nondisabled students; the patterns are similar, however. None of the students reported having less than standard time (less than 30 minutes) remaining for the final subtest. A vast majority of students in the double-time condition (97–100 percent) reported having at least 1.5 time (45 minutes or more) remaining for the final subtest, but the 1.5-time condition varied based on test order. The percent of students with double time or more remaining for the final subtest ranged from a high of 97 percent (double-time condition MVM) to a low of 79 percent (1.5-time condition VMVM).

Ability Level

Analysis of timing data by ability level compared two ability levels, "higher" and "lower," based on the same criteria described earlier for analyses of test performance by ability for students with disabilities. Table 4 displays descriptive statistics on time use for each subtest by ability level and disability status, and Table 5 displays significance statistics. Higher-ability students used significantly more time than lower-ability students for both subtests, M1 and M2. In both cases higher-ability students spent an average of 6 more minutes on each subtest. The differences between the minutes used by higher- and lower-ability students on the verbal tests were not significantly different from one another. However, the actual difference followed the same pattern as the math subtests. On V1 and V2, higher-ability students spent approximately 2.5 more minutes than lower-ability students. Figures 9 and 10 display the percentage of students (by verbal ability level) plotted against minutes remaining at the start of the final subtest for test-takers in the 1.5-time and double-time conditions, respectively. Ability level and timing condition appear to play a role in the amount of time used. It is clear from the graph that the amount of time used by lower-performing students is more varied than the amount of time used by higher-performing students. This relationship is opposite to the pattern found for students without disabilities, where high-, medium-, and low-performing students reported to be using relatively similar amounts of time.

Perceived Use of Time

Students in the standard time condition and the 1.5-time condition (with section breaks) were asked, "On average, did you have enough time to complete the verbal sections?" and "On average, did you have enough time
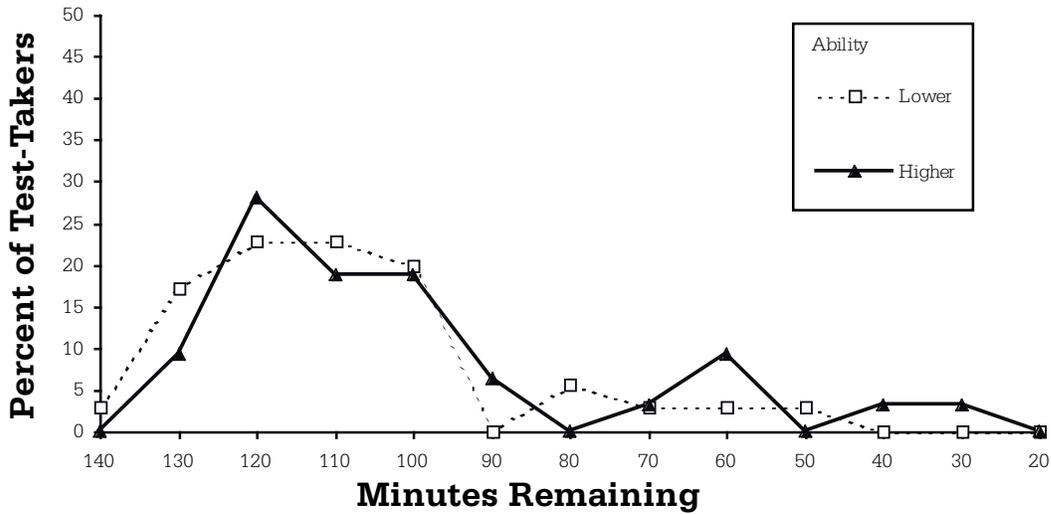
Figure 9. Percent of test-takers (with a learning disability and/or ADHD) in 1.5-time condition without breaks reporting "x" minutes remaining for last subtest, by verbal ability.
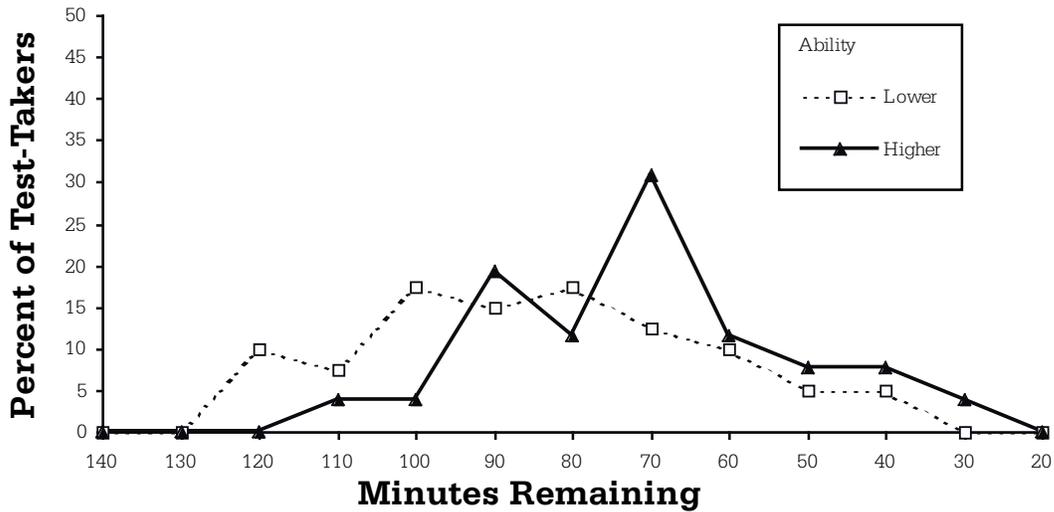
Figure 10. Percent of test-takers (with a learning disability and/or ADHD) in double-time condition without breaks reporting "x" minutes remaining for last subtest, by verbal ability.
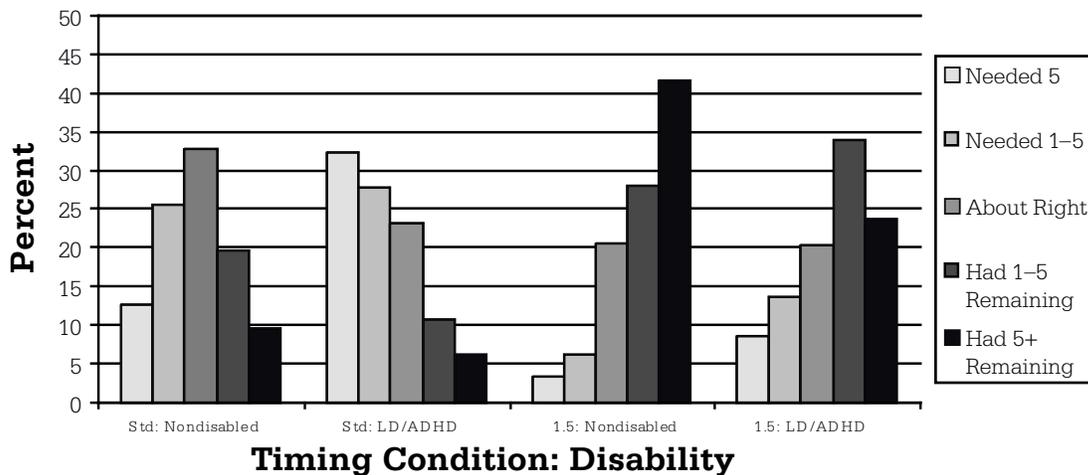
Figure 11. Percent of test-takers (by timing condition and disability status) that reported needing more time or having extra time (on average) on the math subtests.
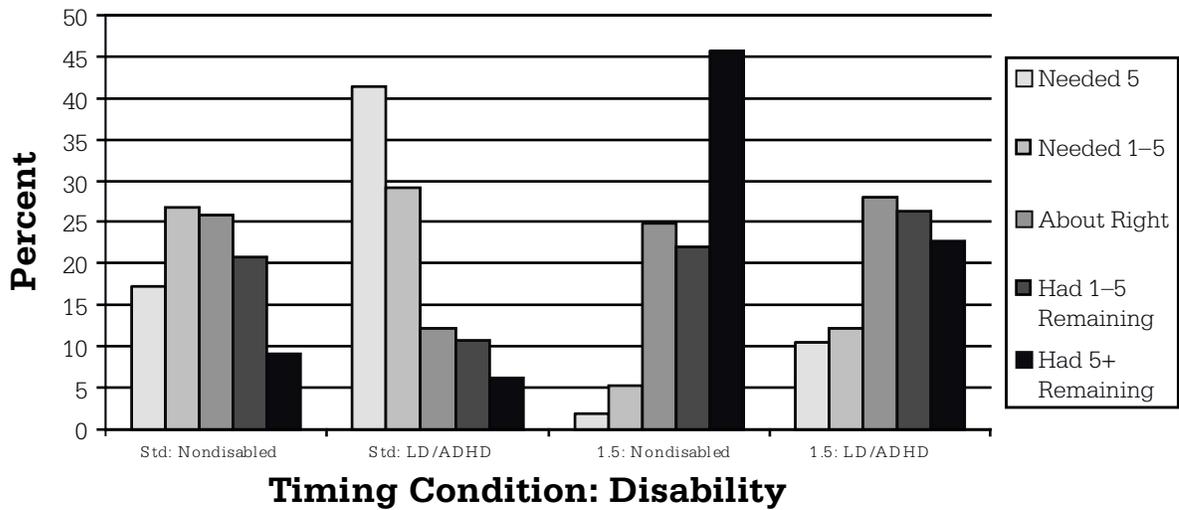
Figure 12. Percent of test-takers (by timing condition and disability status) that reported needing more time or having extra time (on average) on the verbal subtests.

to complete the math sections?" Answer choices for both questions were: (A) No, I needed more than 5 extra minutes per section; (B) No, I needed 1–5 extra minutes per section; (C) The time was just about right; (D) Yes, I finished each section 1–5 minutes before time expired; or (E) Yes, I finished each section more than 5 minutes before time expired. Chi-square analyses comparing responses between nondisabled students and students with a learning disability and/or ADHD indicated significant differences in responses with regard to verbal and math subtests:

math subtests with standard time and section breaks, $\chi 2$ (4, N = 471) = 17.81, p = .001;

math subtests with 1.5 time and section breaks, $\chi 2$ (4, N = 459) = 12.22, p = .016;

verbal subtests with standard time and section breaks, $\chi 2$ (4, N = 471) = 24.00, p = .001; and

verbal subtests with 1.5 time and section breaks, $\chi 2$ (4, N = 459) = 22.72, p = .001.

Figures 11 and 12 display response percentages by timing condition and disability status. Under standard-time conditions, students with a disability were twice as likely to report that they needed more than 5 additional minutes on the math subtests (observed = 21, expected = 10.2); similar results were found on the verbal subtest (observed n = 27, expected n = 13.4). Although fewer students needed more than 5 additional minutes per subtest in the 1.5-time condition, students with LD and/or ADHD were still twice as likely as nondisabled students to report that they did (for math subtests the observed n = 5 and expected n = 2.4; for verbal subtest observed n = 6, expected n = 1.7).

Item Trends for Correctness, Omits, and Items Not Reached

Subtests were examined for indications of speededness in terms of the proportion of students that omitted or did not reach each item and the proportion of students that correctly answered each item. It is difficult to distinguish between items not reached and those omitted. It is generally acknowledged that items are labeled as "not reached" if a test-taker fails to attempt any subsequent questions. The data by subtest, disability status, and timing condition are presented in the Appendix. The data are presented in three forms for each subtest and disability group: (a) a graphical display of the proportion of omits and not-reached items; (b) a tabular display that distinguishes between omits and not-reached items; and (c) a graphical display of the proportion correct for each item.

Results indicated that the proportions of correct, omitted, or not-reached responses are reflective of item difficulty. In most instances, the patterns of item difficulty are evident across the four timing conditions. The graph lines parallel one another for the most part, with minor deviations. The graphs show, however, that students in the standard time were more likely to omit or not reach a higher proportion of items than the other timing condition groups and got fewer items correct. These trends become more evident for the final five items on V1 for the nondisabled and disabled groups, V2 and M1 for students with disabilities, and M2 for the nondisabled students. The results also indicated that across all four subtests, higher proportions of correct responses were obtained by nondisabled students than by students with disabilities. Third, nondisabled students tended to omit fewer responses or failed to reach fewer items than did the students with disabilities. These data

15

provide indications that different groups of students on different subtests are affected differentially by the speededness imposed by the four timing conditions and by item difficulty.

Discussion and Implications

Summary of Findings

It is clear that the impact of extended time on test performance presents a complex set of issues and decisions for testing organizations. This study provides some answers, but it also raises many questions that can pave the way for necessary future research.

No gender interactions were found in this study on either the math or verbal sections of the test. Ability level and disability seem to affect performance. The sample used in the current study was found to be higher in ability than those from operational administrations of the SAT. Also, despite efforts to involve more students with disabilities, the sample size for this group was relatively small. Students identified here as having a disability did not perform as well on average as students without disabilities, regardless of the test-timing condition. Although this study involved students with learning disabilities and/or ADHD, it is important to note that the analyses here did not distinguish them as separate categories. The comorbidity and the small number of students with ADHD did not permit such analyses.

Extra time seemed to affect the math sections of the SAT more than the verbal sections. For students without disabilities, the best performance was achieved under the 1.5-time condition with section breaks, and the lowest with standard time. These findings held for high- and medium-ability examinees. For low-ability examinees, extra time provided no advantage. Apparently, more time does not help if the students do not have effective solution strategies to answer the question successfully. Incorrect answers are a result of a lack of knowledge, not a lack of time. The 1.5-time condition with section breaks also proved most beneficial for the verbal sections of the test for all ability groups, but the effects were not as great as for the math sections.

Although the results for the students with disabilities were not significant due to the small sample sizes and to the size of the effect, the differences were actually larger than those found in the nondisabled group. The trends suggest that the 1.5-time condition with section breaks was the most beneficial for the high-ability students on both the math and verbal sections. This condition was only marginally better for the low-ability students on the verbal section.

In 2001, the procedure for providing extended time as an accommodation to test-takers with disabilities was to give the examinees a total block of time, unimpeded by section breaks. Students could work through the test, allocating time across sections as needed. A possible conclusion from this study is that the provision of section breaks is beneficial to performance not only for students with disabilities but also for nondisabled students. The pattern of means suggests that tests without section breaks may be detrimental to the performance of the students who are to be assisted by the accommodation. These differences were not, however, statistically significant, and thus point to the need for more research on this question. Section breaks allow students to pace themselves through the test, regardless of the students' ability or disability. One test center supervisor, in fact, reported after the testing that, contrary to her hypothesis, even the most able students struggled without section breaks. They seemed unable to pace themselves effectively in the extended-time conditions according to the timing records. Although there is in theory a benefit for students to work through the test at their own pace, allocating extra time to sections that require more attention, such open timing could actually disadvantage students. Our results suggest this may be true, but further research on the provision of section breaks with extended-time accommodations is needed.

The timing records from students who participated in the two conditions without section breaks provide some insights as to how different groups of students allocated their time across and within sections. High-ability students spent less time on V1, V2, and M1 than on M2, whereas the low-ability students spent less time on M2. The complex interaction on M2 may be explained by the inclusion of the student-generated response and quantitative comparison items in this section. In both the 1.5-time and the double-time conditions, the low-ability students take the least amount of time, possibly because they lack viable solution strategies that can be used to generate answers to these items. Extra time, therefore, is not an asset to them. Conversely, for medium-ability students, particularly in the double-time condition, the provision of extra time may give them the opportunity to work through the items in a more thorough and effective manner, thereby increasing their chances of generating a correct answer. On the multiple-choice items, however, high-ability students are able to work through these sections at a rapid pace, indicating that extra time makes a difference up to a point, after which there is relatively little incremental benefit. The medium-ability students tend to spend the largest amount of time on these sections.

The degree to which time spent relates to performance varies across sections of the test, ability, and disability group. No significant relationship was found between

time spent on the verbal section and performance. This result may partly be explained by the fact that the verbal sections contained more items (nearly 60 percent) for which extra time will not be helpful (verbal analogy and sentence completion versus reading comprehension). Either the student knows the answer or does not, and extra time will not help. Extended time on the math items, in contrast, may allow the student the opportunity to work through the problems and obtain correct answers. For students with disabilities in the 1.5-time condition, the more time reported spent on M1 and M2 items, the better the performance. For nondisabled students, more time spent on M2 items in both the 1.5-time and the double-time conditions was associated with better performance. Students with and without disabilities who tended to perform better on M2 also performed better on the other sections of the test. Proportions of students who answered items correctly or omitted them also provide indications of the effects of speededness and extended time.

Limitations and Future Research

A significant amount of effort was devoted to making the design scientifically sound and practically feasible. Although the design went through many revisions and reviews, the study still suffers from certain unavoidable limitations. Despite targeting specific schools with high percentages of students with diagnosed learning disabilities, we were unable to attract a large sample of students with LD and/or ADHD, particularly those who would be classified into the high-ability group. Because of the small numbers, we were forced to combine the medium- and high-ability groups, thereby making the ability groups between the nondisabled and disabled students nonparallel. We also were unable to separate the students with LD from those with ADHD. To be consistent with the differentiations among diagnoses, we should have three groups: LD, ADHD, and LD/ADHD. Additional constraints affected the composition and size of the sample.

We sought participation of high schools that first made the decision whether or not to participate. The school counselors then solicited the participation of their junior classes. Some schools did a better job than others at seeking a broad level of participation. Participation was strictly voluntary, but it was clear that students and schools had different rationales for participating and different levels of motivation. The final composition of the sample was not what we expected. Schools sent us a final list of students from which rosters were prepared where students were randomly assigned to testing conditions. There was substantial attrition between the time that students agreed to participate and the testing day. We lost an additional 100 students because we were unable to locate their PSAT/NMSQT scores, a requirement for participation.

If all students who had registered with us had participated, the number of subjects would have been 3,194. Thus, our final sample was only 60 percent of what was expected. Future research will need to improve upon the composition of the sample.

There were two instances where the schools failed to follow the instructions for the distribution of test books, and therefore testing condition. We were very specific about the assignment of students to testing condition, to maintain random assignment and the integrity of the design. Although the schools assured us that the test booklets were randomly assigned, the distribution, in addition to the attrition, caused an imbalance in the design.

There were three design issues that need to be addressed in future studies. First, in randomizing the order of the sections within each timing condition, M1 always preceded M2 and V1 always preceded V2. The verbal sections are not an issue because they contain the same item types. M1 and M2, however, do contain different item types and therefore the order of presentation should be balanced so that M1 followed M2 half of the time. Second, future research should include a double-time condition with section breaks to balance the 1.5-time condition. In the current study, we had a specific rationale for the two 1.5-time groups. The group that did not receive separately timed sections replicated how extended-time accommodations were administered in 2001. The intention of including a group that received time and a half with section breaks was to compare their performance with the standard-time condition to determine the effects of the extra time, and then compare them with the 1.5-time condition without breaks to examine the impact of open timing versus section breaks. For a complete set of comparisons, we introduced a confounding element by not including a double-time group with section breaks. Third, future studies need to have better and more reliable estimates of how students used their time across sections. We only requested that the two groups who were not given section breaks log their time. Admittedly, this information was self-reported, but it was the best data we could obtain. We also needed to obtain parallel information from the other two groups. We assumed that the omitted/not reached item analyses would provide some informative data.

The current study begins to address some of the issues related to the impact of extended time on the performance of various groups of students. Should testing time be lengthened, the number of items decreased, or different instructions given? Are there differential effects of time limits on the verbal and mathematics sections of the test? Does extra time or too much time negatively impact performance? Do separately timed sections help or hinder performance? Is there a differential impact of extended time or section breaks for students with and without

disabilities, for those of different abilities, or for males and females? What is the interaction between item type and extended time in terms of effects on test performance? This study provides evidence of three major findings:

- Lower-ability test-takers gain little or no benefit from extra time. If students do not have the knowledge or skills, no amount of extra time will improve performance.

- Section breaks appear to help test-takers at different ability levels, regardless of their disability status.

- Extra time helps medium- and high-ability test-takers with and without disabilities. Extra time, however, does not help and actually may hinder low-ability students with disabilities.

The results point to a number of potential follow-up research questions to clarify and extend the current findings. In particular, we need to examine further the effects of section breaks as well as how extended time affects performance on particular item types. Another topic is to distinguish students with learning disabilities from those with ADHD and examine the impact of extended time on their performance. The examination of extended time also needs to be conducted on other tests such as the SAT Subject Tests™ and Advanced Placement Program® Examinations.

 The current study, however, leaves unanswered questions, particularly in terms of operational issues. What are the practical and operational implications from the results? What operational recommendations should be made, and what requires additional research? How can operational decisions that may differentially impact one group of students over others be reconciled? Decisions that will ensue from this line of research could have substantial implications for the delivery of future standard administrations as well as those with accommodations. Indeed, partly as a result of this research, section timing is now part of the standard extra-time accommodation. Follow-up studies are needed to determine if this accommodation is functioning as intended. One objective is not to differentially advantage or disadvantage any particular group of test-takers with potential changes to test timing, procedures, or administration. The commitment to provide appropriate and reasonable accommodations for test-takers with disabilities must be maintained. Because any potential changes must recognize the technical, political, ethical, and legal issues surrounding extended time, it is important for the College Board and ETS to continue zealously to pursue this line of research to establish more definitive performance trends across groups and testing conditions.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bridgeman, B., Trapani, C., & Curley, E. (2004). Impact of fewer questions per section on SAT I scores. *Journal of Educational Measurement*, *41*, 291–310.

Camara, W. J., Copeland, T., & Rothschild, B. (1998). *Effects of extended time on the SAT I: Reasoning test score growth for students with disabilities* (College Board Report No. 98–7). New York: The College Board.

Caperton, G. (2000, December 15). SAT: Give extra time to qualified students. *Los Angeles Times*. Retrieved December 16, 200l, from http://www.latimes.com/cgi-bin/latreg2?16203783.

College Board. (1997). *Handbook for the SAT Program 1997-98: Admissions staff*. New York: The College Board.

College Board. (2001). *College Board services for students with disabilities (SSD) brochure*. New York: The College Board.

College Board. (2002). *Test characteristics of the SAT I: Reliability, difficulty levels, completion rates*. Retrieved February 26, 2002, from http://www.collegeboard.com/sat/cbsenior/yr2001/pdf/five.pdf.

Donlon, T. F. (Ed.) (1984). *The College Board technical handbook for the Scholastic Aptitude Test and Achievement Tests*. New York: The College Board.

Elliott, S. N., & McKevitt, B. (2000, April). *Experimental analysis of the effects of testing accommodations on the scores of students with disabilities: Design issues and initial results*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Evans, F. R. (1980). *A study of the relationshi p among speed and power aptitude test scores and ethnic identity* (College Board Research and Development Report, RDR 80-81, No. 2). Princeton, NJ: Educational Testing Service.

Greenberg, B. (2000, February 21). *Learning-disabled flocking to colleges: Trend dispels myths, questions labeling of kids*. Retrieved February 22, 2000, from http://www.desertnews.com/dn/view/0,1248,150015387.00.html.

Halla, J. W. (1988). A psychological study of psychometric differences in graduate record examinations general test scores between learning disabled and non-learning disabled adults (Doctoral dissertation, Texas Tech University). *Dissertation Abstracts International, 49*, 3341A.

Henderson, C. (1999). *1999 College freshmen with disabilities: A biennial statistical profile*. Washington, DC: American Council on Education.

Heyboer, K., & McCarron, B. (1999, August 15). 'Joining the club': Claim a disability, get an edge in SATs. *Star Ledger, 1*.

Hill, G. A. (1984). Learning disabled college students: The assessment of academic aptitude (Doctoral dissertation, Texas Tech University). *Dissertation Abstracts International, 46*, 642B.

Huesman, R. L. Jr., & Frisbie, D. A. (2000, April). *The validity of ITBS reading comprehension test scores for learning disabled and non-learning disabled students under extended-time conditions*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.

Leatherman, C. (2000, December 4). California study finds racial disparities in granting extra time on SAT. *The Chronicle of Higher Education*. Retrieved December 6, 2000, from http://chronicle.com/daily/2000/12/2000120405nhtm.

Lee, L. (2000, April 9). To teach, or merely accommodate. *New York Times Education Life*, pp. 30–31, 40.

Mandinach, E. B., Cahalan, C., & Camara, W. J. (2002). *The impact of flagging on the admission process: Policies, practices, and implications* (ETS RR 02-03, College Board Report 2002-2). New York and Princeton, NJ: The College Board and Educational Testing Service.

Marquart, A. M. (2000, June). *The use of extended time as an accommodation on a standardized mathematics test: An investigation of effects on scores and perceived consequences for students of various skill levels*. Paper presented at the annual meeting of the Council of Chief State School Officers, Snowbird, UT. Retrieved December 12, 2001, from www.wcer.wisc.edu/testacc/publications/marquart%20ccsso%20paper.htm.

Mehrens, W. A., Millman, J., & Sackett, P. R. (1994, November). Accommodations for candidates with disabilities. *The Bar Examiner, 63*, 33–47.

Munger, G. F., & Loyd, B. H. (1991). Effect of speededness on test performance of handicapped and nonhandicapped examinees. *Journal of Educational Research, 85*(1), 53–57.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Packer, J. (1987). *SAT testing time for students with disabilities*. Unpublished report. Princeton, NJ: Educational Testing Service.

Phillips, S. E. (1994). High-stakes testing accommodations: Validity versus disabled rights. *Applied Measurement in Education, 7*(2), 93–120.

Pitoniak, M. J., & Royer, J. M. (2001). Testing accommodations for examinees with disabilities: A review of psychometric, legal, and social policy issues. *Review of Educational Research, 71*(1), 53–104.

Ragosta, M., & Wendler, C. (1992). *Eligibility issues and comparable time limits for disabled and nondisabled SAT examinees* (College Board Research Report No. 92-5). New York: The College Board.

Runyan, M. K. (1991). The effect of extra time on reading comprehension scores for university students with and without learning disabilities. *Journal of Learning Disabilities, 24*(2), 104–108.

Swineford, F. (1974). *The test analysis manual* (ETS SR 74-06). Princeton, NJ: Educational Testing Service.

Weaver, S, M. (1993). *The validity of the use of extended time and untimed testing for postsecondary students with learning disabilities*. Unpublished doctoral dissertation, University of Toronto, Toronto, Ontario, Canada.

Weiss, K. R. (2000a, January 9). New test-taking skill: Working the system. *Los Angeles Times*. Retrieved January 10, 2000, from www.Latimes.com/cgibin/slwebeli?DBLIST=lt000&D...:%20Working%20the%System.

Weiss, K. R. (2000b, December 1). Audit confirms disparities in SAT testing. *Los Angeles Times*. Retrieved December 2, 2000, from www.latimes.com/news/state/20001201/t000115187.html.

Wild, C., & Durso, R. (1979). *Effect of increased test-taking time on test scores by ethnic group, age, and sex* (GREB No. 76-6R). Princeton, NJ: Educational Testing Service.

Wright, N., & Wendler, C. (1994, April). *Establishing timing limits for the new SAT for students with disabilities*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA.

Zuriff, G. E. (2000). Extra examination time for students with learning disabilities: An examination of the maximum potential thesis. *Applied Measurement in Education, 13*(1), 99–117.

Appendix: Proportions Correct and Omit/Not Reached by Item for V1, V2, M1, and M2 by Disability Status

Figure 1. Proportion Omit/Not Reached: M1–Non-LD.

	Regular Time—With Sections N = 425			1.5 Time—With Sections N = 423			1.5 Time—No Sections N = 435			Double Time—No Sections N = 382		
	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR
Item 1	3.3		3.3	2.1		2.1	2.1		2.1	1.6		1.6
Item 2	0.9		0.9	0.2		0.2	1.4		1.4	1.8		1.8
Item 3	0.7		0.7	1.2		1.2	0.5		0.5	0.3		0.3
Item 4	3.1		3.1	2.1		2.1	2.3		2.3	3.4		3.4
Item 5	0.5		0.5	0.2		0.2	0.0		0.0	0.0		0.0
Item 6	1.9		1.9	1.2		1.2	1.4		1.4	0.8		0.8
Item 7	1.6		1.6	0.7		0.7	0.9		0.9	0.8		0.8
Item 8	4.0		4.0	5.7		5.7	6.0		6.0	6.3		6.3
Item 9	6.8		6.8	6.9		6.9	4.8		4.8	4.7		4.7
Item 10	15.5		15.5	11.6		11.6	12.6		12.6	10.5		10.5
Item 11	6.1		6.1	4.5		4.5	6.4		6.4	4.5		4.5
Item 12	34.8		34.8	33.1		33.1	38.9		38.9	31.2		31.2
Item 13	10.4		10.4	5.7		5.7	5.5		5.5	5.5		5.5
Item 14	20.7		20.7	13.5		13.5	17.5		17.5	13.1		13.1
Item 15	12.2	0.5	12.7	5.4		5.4	7.8		7.8	9.2		9.2
Item 16	12.5	0.7	13.2	8.7		8.7	12.4		12.4	11.0		11.0
Item 17	39.5	0.9	40.5	27.7		27.7	39.3		39.3	32.7	0.3	33.0
Item 18	6.1	1.4	7.5	4.5		4.5	3.0		3.0	2.6	0.3	2.9
Item 19	12.7	2.4	15.1	9.9		9.9	9.7	0.2	9.9	8.9	0.3	9.2
Item 20	3.5	2.6	6.1	2.1		2.1	0.9	0.2	1.1	2.9	0.3	3.1
Item 21	11.1	4.7	15.8	9.0	0.2	9.2	9.7	0.7	10.3	9.9	0.5	10.5
Item 22	2.4	8.0	10.4	1.2	1.2	2.4	1.6	1.1	2.8	1.0	1.0	2.1
Item 23	16.5	16.2	32.7	12.8	3.1	15.8	16.3	1.6	17.9	14.4	2.6	17.0
Item 24	2.4	22.8	25.2	2.1	4.0	6.1	2.8	5.1	7.8	2.4	5.5	7.9
Item 25		44.9	44.9		17.3	17.3		27.1	27.1		21.7	21.7

Figure 2. Proportion Correct: M1–Non-LD.

Figure 3. Proportion Omit/Not Reached: M1–LD.

	Regular Time—With Sections N = 66			1.5 Time—With Sections N = 61			1.5 Time—No Sections N = 68			Double Time—No Sections N = 69		
	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR
Item 1	7.6		7.6	1.6		1.6				1.4		1.4
Item 2	4.5		4.5	1.6		1.6	2.9		2.9	2.9		2.9
Item 3	1.5		1.5	1.6		1.6	1.5		1.5	4.3		4.3
Item 4	7.6		7.6	4.9		4.9	1.5		1.5	2.9		2.9
Item 5	0.0		0.0	0.0		0.0	1.5		1.5	0.0		0.0
Item 6	3.0		3.0	0.0		0.0	1.5		1.5	1.4		1.4
Item 7	1.5		1.5	3.3		3.3	2.9		2.9	1.4		1.4
Item 8	15.2	1.5	16.7	11.5		11.5	20.6		20.6	10.1		10.1
Item 9	7.6	1.5	9.1	9.8		9.8	14.7		14.7	8.7		8.7
Item 10	21.2	1.5	22.7	11.5		11.5	17.6		17.6	14.5		14.5
Item 11	15.2	1.5	16.7	3.3		3.3	7.4		7.4	10.1		10.1
Item 12	31.8	1.5	33.3	29.5		29.5	25.0		25.0	27.5		27.5
Item 13	24.2	1.5	25.8	9.8		9.8	26.5		26.5	7.2		7.2
Item 14	24.2	1.5	25.8	24.6		24.6	27.9		27.9	24.6		24.6
Item 15	15.2	1.5	16.7	14.8		14.8	14.7		14.7	15.9		15.9
Item 16	13.6	1.5	15.2	11.5		11.5	7.4		7.4	5.8		5.8
Item 17	54.5	3.0	57.6	45.9		45.9	41.2		41.2	27.5		27.5
Item 18	10.6	3.0	13.6	9.8		9.8	16.2		16.2	8.7		8.7
Item 19	6.1	7.6	13.6	13.1		13.1	17.6		17.6	8.7		8.7
Item 20	1.5	9.1	10.6	8.2		8.2	4.4		4.4	2.9		2.9
Item 21	10.6	13.6	24.2	18.0		18.0	20.6	1.5	22.1	8.7		8.7
Item 22	6.1	21.2	27.3	6.6		6.6	2.9	1.5	4.4	1.4		1.4
Item 23	12.1	27.3	39.4	13.1	4.9	18.0	20.6	8.8	29.4	17.4	4.3	21.7
Item 24	4.5	37.9	42.4	4.9	14.8	19.7	1.5	13.2	14.7	1.4	5.8	7.2
Item 25		47.0	47.0		23.0	23.0		35.3	35.3		21.7	21.7

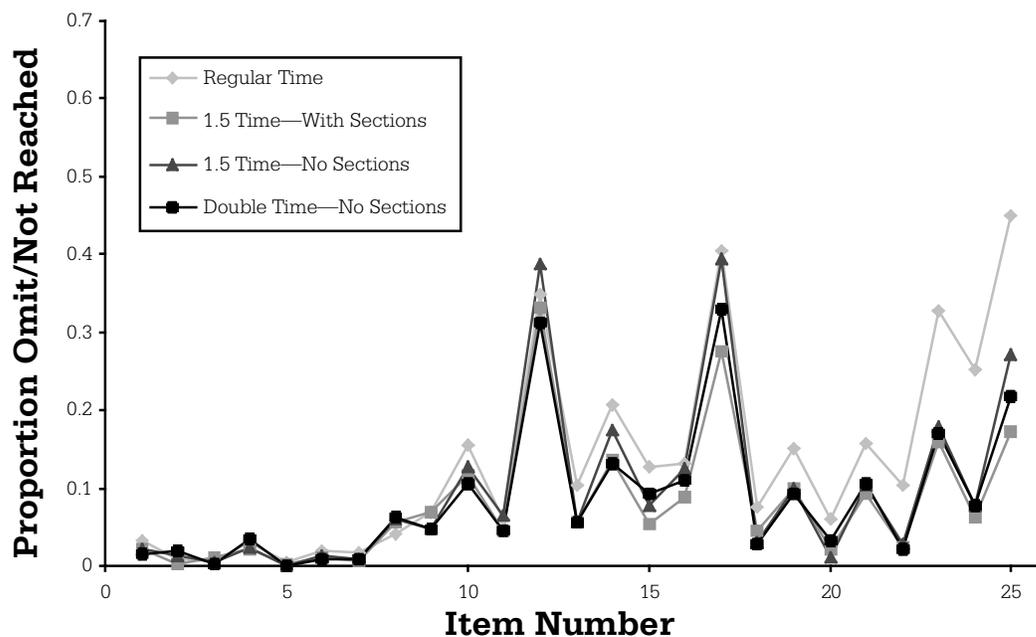**Figure 4.** Proportion Correct: M1–LD.
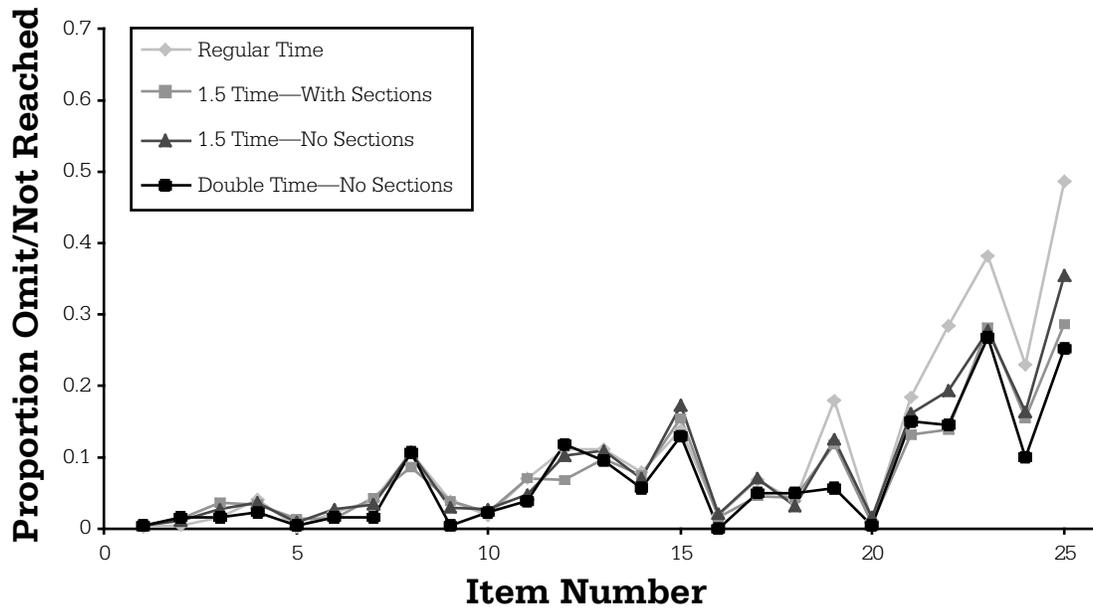
Figure 5. Proportion Omit/Not Reached: M2–Non-LD.

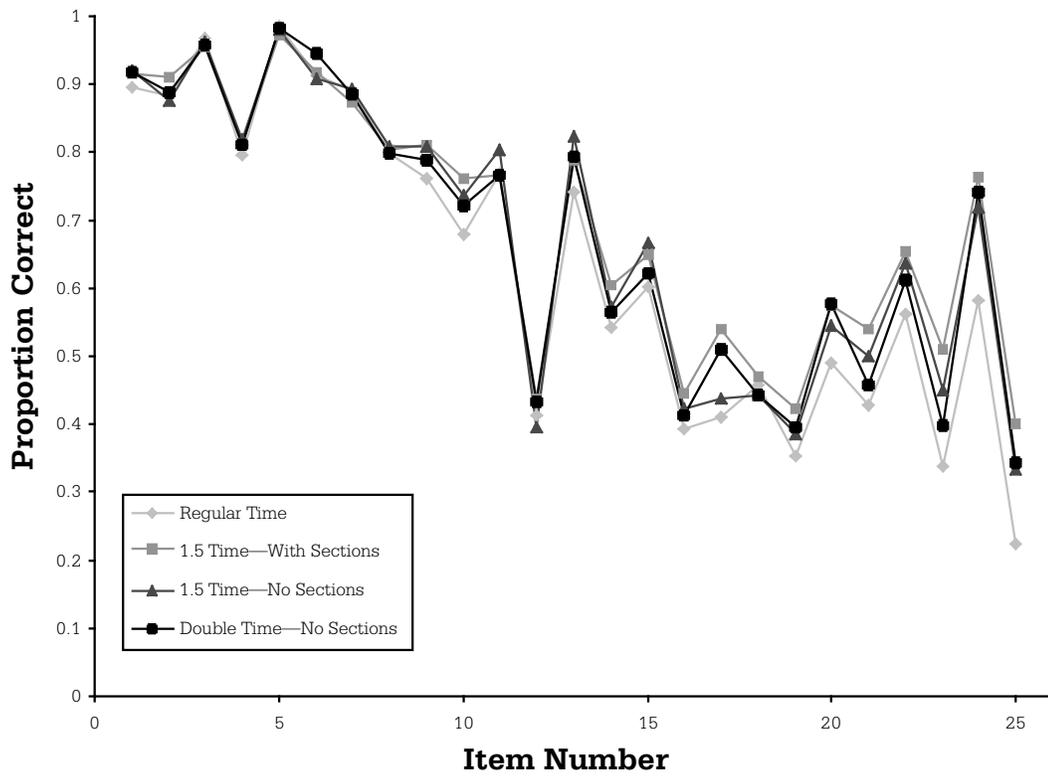|	Regular Time—With Sections N = 425			1.5 Time—With Sections N = 423			1.5 Time—No Sections N = 435			Double Time—No Sections N = 179		
	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR
Item 1	0.2		0.2	0.5		0.5	0.2	0.2	0.5	0.6		0.6
Item 2	0.5		0.5	1.4		1.4	0.9	0.2	1.1	1.7		1.7
Item 3	1.6		1.6	3.5		3.5	2.5	0.2	2.8	1.7		1.7
Item 4	4.0		4.0	3.3		3.3	3.4	0.2	3.7	2.2		2.2
Item 5	0.2		0.2	1.4		1.4	0.7	0.2	0.9	0.6		0.6
Item 6	2.4		2.4	1.4		1.4	2.5	0.2	2.8	1.7		1.7
Item 7	3.8		3.8	4.3		4.3	3.2	0.2	3.4	1.7		1.7
Item 8	10.6		10.6	8.7		8.7	10.3	0.2	10.6	10.6		10.6
Item 9	3.8		3.8	3.8		3.8	2.8	0.2	3.0	0.6		0.6
Item 10	1.9		1.9	2.4		2.4	2.5	0.2	2.8	2.2		2.2
Item 11	7.1		7.1	7.1		7.1	4.6	0.2	4.8	3.9		3.9
Item 12	11.1		11.1	6.9		6.9	9.9	0.2	10.1	11.7		11.7
Item 13	11.1		11.1	9.7		9.7	10.6	0.2	10.8	9.5		9.5
Item 14	8.0		8.0	7.6		7.6	6.9	0.2	7.1	5.6		5.6
Item 15	13.9		13.9	15.4		15.4	16.8	0.5	17.2	12.8		12.8
Item 16	1.6	0.2	1.9	1.4	0.2	1.7	1.6	0.5	2.1	0.0		0.0
Item 17	6.6	0.2	6.8	4.3	0.2	4.5	6.7	0.5	7.1	5.0		5.0
Item 18	3.3	0.5	3.8	4.0	0.2	4.3	2.8	0.5	3.2	5.0		5.0
Item 19	17.2	0.7	17.9	11.1	0.7	11.8	12.0	0.5	12.4	5.6		5.6
Item 20	0.5	0.7	1.2	0.0	0.7	0.7	1.1	0.5	1.6	0.6		0.6
Item 21	15.1	3.3	18.4	11.3	1.9	13.2	14.0	2.1	16.1	14.5	0.6	15.1
Item 22	21.2	7.3	28.5	10.4	3.5	13.9	15.9	3.4	19.3	12.3	2.2	14.5
Item 23	25.6	12.5	38.1	21.5	6.6	28.1	21.8	6.0	27.8	23.5	3.4	26.8
Item 24	4.5	18.6	23.1	5.7	9.7	15.4	4.6	11.7	16.3	3.9	6.1	10.1
Item 25		48.7	48.7		28.6	28.6		35.4	35.4		25.1	25.1

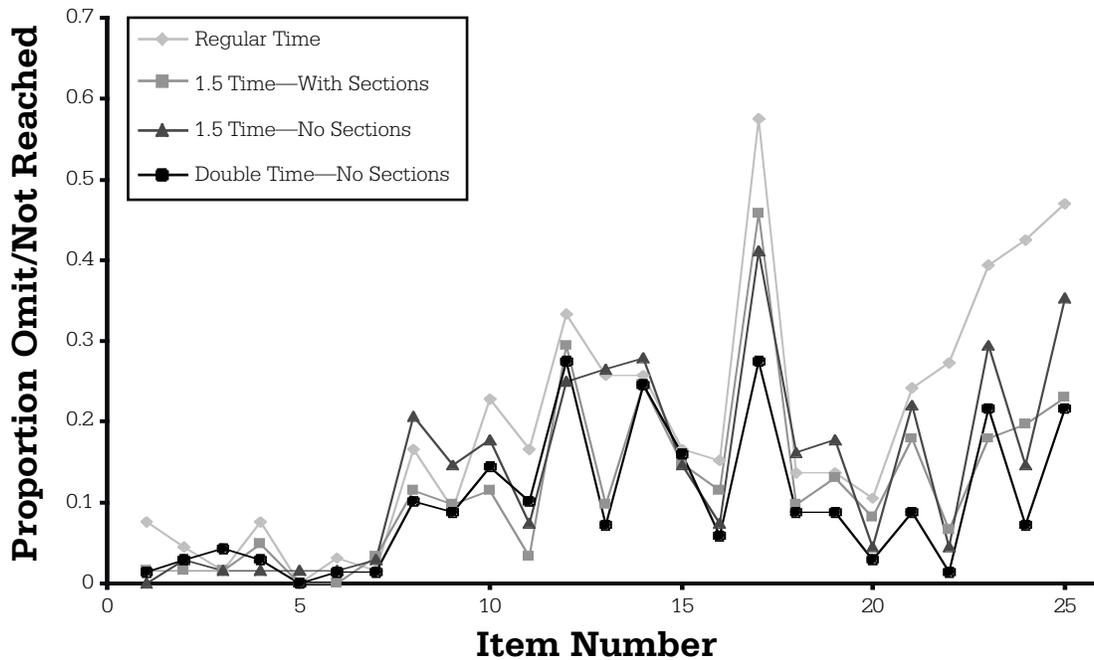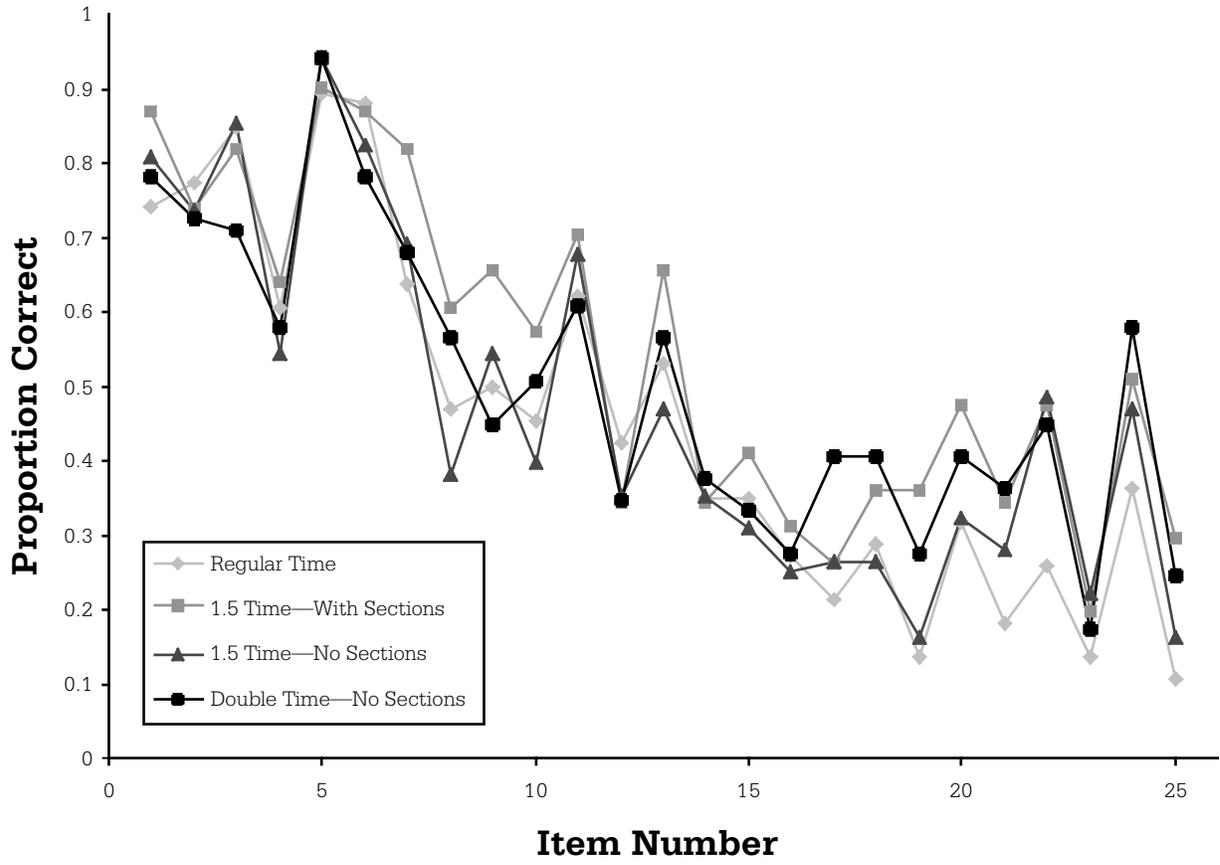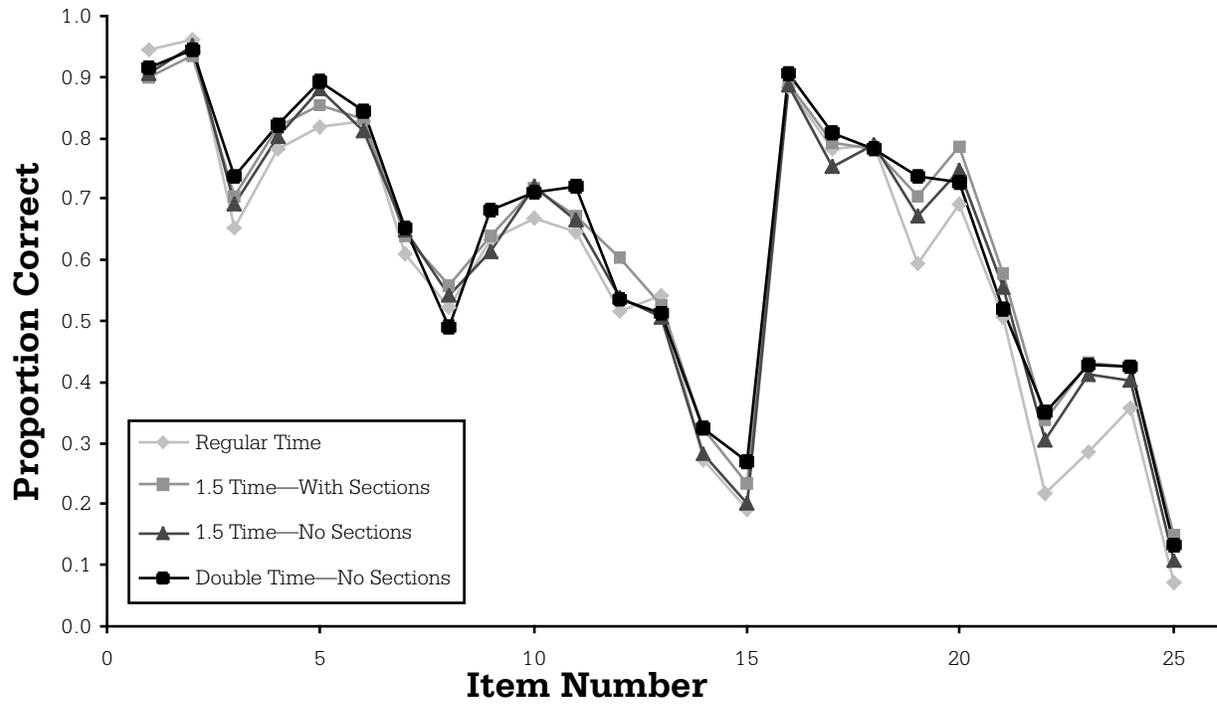**Figure 6.** Proportion Correct: M2–Non-LD.

Figure 7. Proportion Omit/Not Reached: M2–LD.

	Regular Time—With Sections N = 66			1.5 Time—With Sections N = 61			1.5 Time—No Sections N = 68			Double Time—No Sections N = 32		
	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR
Item 1	1.5		1.5	0.0	1.6	1.6	2.9		2.9	0.0	3.1	3.1
Item 2	6.1		6.1	3.3	1.6	4.9	11.8		11.8	0.0	3.1	3.1
Item 3	10.6		10.6	8.2	1.6	9.8	10.3		10.3	6.3	3.1	9.4
Item 4	9.1		9.1	3.3	1.6	4.9	19.1		19.1	6.3	3.1	9.4
Item 5	1.5		1.5	3.3	1.6	4.9	2.9		2.9	0.0	3.1	3.1
Item 6	3.0		3.0	0.0	1.6	1.6	7.4		7.4	3.1	3.1	6.3
Item 7	7.6		7.6	9.8	1.6	11.5	8.8		8.8	3.1	3.1	6.3
Item 8	7.6		7.6	8.2	1.6	9.8	11.8		11.8	12.5	3.1	15.6
Item 9	3.0		3.0	6.6	1.6	8.2	8.8		8.8	3.1	3.1	6.3
Item 10	6.1		6.1	3.3	1.6	4.9	7.4		7.4	3.1	3.1	6.3
Item 11	6.1		6.1	3.3	1.6	4.9	19.1		19.1	9.4	3.1	12.5
Item 12	18.2		18.2	16.4	1.6	18.0	25.0	1.5	26.5	6.3	3.1	9.4
Item 13	22.7	1.5	24.2	13.1	1.6	14.8	14.7	1.5	16.2	9.4	3.1	12.5
Item 14	9.1	1.5	10.6	11.5	1.6	13.1	17.6	1.5	19.1	12.5	3.1	15.6
Item 15	16.7	1.5	18.2	11.5	1.6	13.1	26.5	2.9	29.4	15.6	3.1	18.8
Item 16	3.0	1.5	4.5	3.3	1.6	4.9	5.9	2.9	8.8	3.1	3.1	6.3
Item 17	16.7	1.5	18.2	11.5	1.6	13.1	13.2	4.4	17.6	6.3	3.1	9.4
Item 18	13.6	1.5	15.2	9.8	1.6	11.5	8.8	7.4	16.2	12.5	3.1	15.6
Item 19	30.3	3.0	33.3	18.0	1.6	19.7	23.5	7.4	30.9	18.8	3.1	21.9
Item 20	3.0	3.0	6.1	1.6	1.6	3.3	2.9	7.4	10.3	3.1	3.1	6.3
Item 21	21.2	7.6	28.8	18.0	3.3	21.3	19.1	17.6	36.8	21.9	3.1	25.0
Item 22	37.9	16.7	54.5	18.0	13.1	31.1	22.1	25.0	47.1	28.1	12.5	40.6
Item 23	31.8	21.2	53.0	24.6	14.8	39.3	13.2	27.9	41.2	34.4	15.6	50.0
Item 24	1.5	31.8	33.3	6.6	23.0	29.5	5.9	30.9	36.8		18.8	18.8
Item 25		66.7	66.7		44.3	44.3		48.5	48.5		46.9	46.9

Figure 8. Proportion Correct: M2–LD.
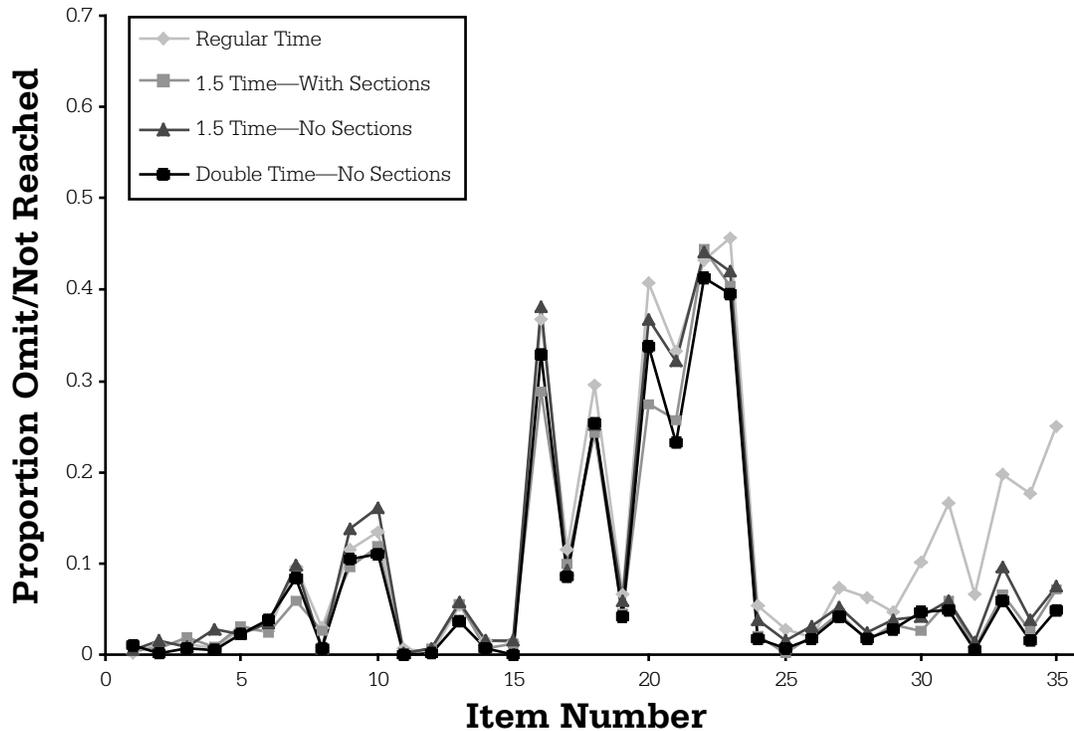
Figure 9. Proportion Omit/Not Reached: V1–Non-LD.

	Regular Time—With Sections N = 425			1.5 Time—With Sections N = 423			1.5 Time—No Sections N = 435			Double Time—No Sections N = 382		
	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR
Item 1	0.2		0.2	0.5		0.5	0.5		0.5	1.0		1.0
Item 2	1.2		1.2	0.7		0.7	1.6		1.6	0.3		0.3
Item 3	1.2		1.2	1.9		1.9	0.9		0.9	0.8		0.8
Item 4	0.7		0.7	0.9		0.9	2.8		2.8	0.5		0.5
Item 5	2.8		2.8	3.1		3.1	2.3		2.3	2.4		2.4
Item 6	2.6		2.6	2.4		2.4	3.4		3.4	3.9		3.9
Item 7	9.6		9.6	5.9		5.9	9.9		9.9	8.4		8.4
Item 8	3.1		3.1	2.6		2.6	0.7		0.7	0.8		0.8
Item 9	11.5		11.5	9.7		9.7	13.8		13.8	10.5		10.5
Item 10	13.4		13.4	11.8		11.8	16.1		16.1	11.0		11.0
Item 11	0.7		0.7	0.2		0.2	0.2		0.2	0.0		0.0
Item 12	0.2		0.2	0.7		0.7	0.7		0.7	0.3		0.3
Item 13	5.4		5.4	5.7		5.7	5.7		5.7	3.7		3.7
Item 14	1.6		1.6	0.7		0.7	1.6		1.6	0.8		0.8
Item 15	1.4		1.4	1.2		1.2	1.6		1.6	0.0		0.0
Item 16	36.7		36.7	28.8		28.8	38.2		38.2	33.0		33.0
Item 17	11.5		11.5	9.9		9.9	9.2		9.2	8.6		8.6
Item 18	29.6		29.6	24.3		24.3	25.3		25.3	25.4		25.4
Item 19	6.6		6.6	4.5		4.5	6.0		6.0	4.2		4.2
Item 20	40.7		40.7	27.4		27.4	36.8		36.8	33.8		33.8
Item 21	32.9	0.2	33.2	25.8		25.8	32.2		32.2	23.3		23.3
Item 22	43.1	0.2	43.3	44.4		44.4	44.1		44.1	41.4		41.4
Item 23	45.4	0.2	45.6	40.4		40.4	42.1		42.1	39.5		39.5
Item 24	4.9	0.5	5.4	2.1		2.1	3.9		3.9	1.8		1.8
Item 25	2.4	0.5	2.8	0.2		0.2	1.6		1.6	0.8		0.8
Item 26	1.4	0.5	1.9	2.1		2.1	3.2		3.2	1.8		1.8
Item 27	5.9	1.4	7.3	4.5		4.5	5.3		5.3	4.2		4.2
Item 28	4.5	1.9	6.4	1.7		1.7	2.5		2.5	1.8		1.8
Item 29	2.4	2.4	4.7	3.3		3.3	3.9		3.9	2.9		2.9
Item 30	6.8	3.3	10.1	2.6		2.6	3.9	0.2	4.1	4.7		4.7
Item 31	11.5	5.2	16.7	5.4	0.5	5.9	4.8	1.1	6.0	4.7	0.3	5.0
Item 32	0.7	5.9	6.6	0.2	0.5	0.7	0.2	1.1	1.4	0.3	0.3	0.5
Item 33	7.5	12.2	19.8	5.4	1.2	6.6	8.3	1.4	9.7	5.2	0.8	6.0
Item 34	1.6	16.0	17.6	0.9	1.7	2.6	1.6	2.3	3.9	0.5	1.0	1.6
Item 35		24.9	24.9		7.1	7.1		7.6	7.6		5.0	5.0

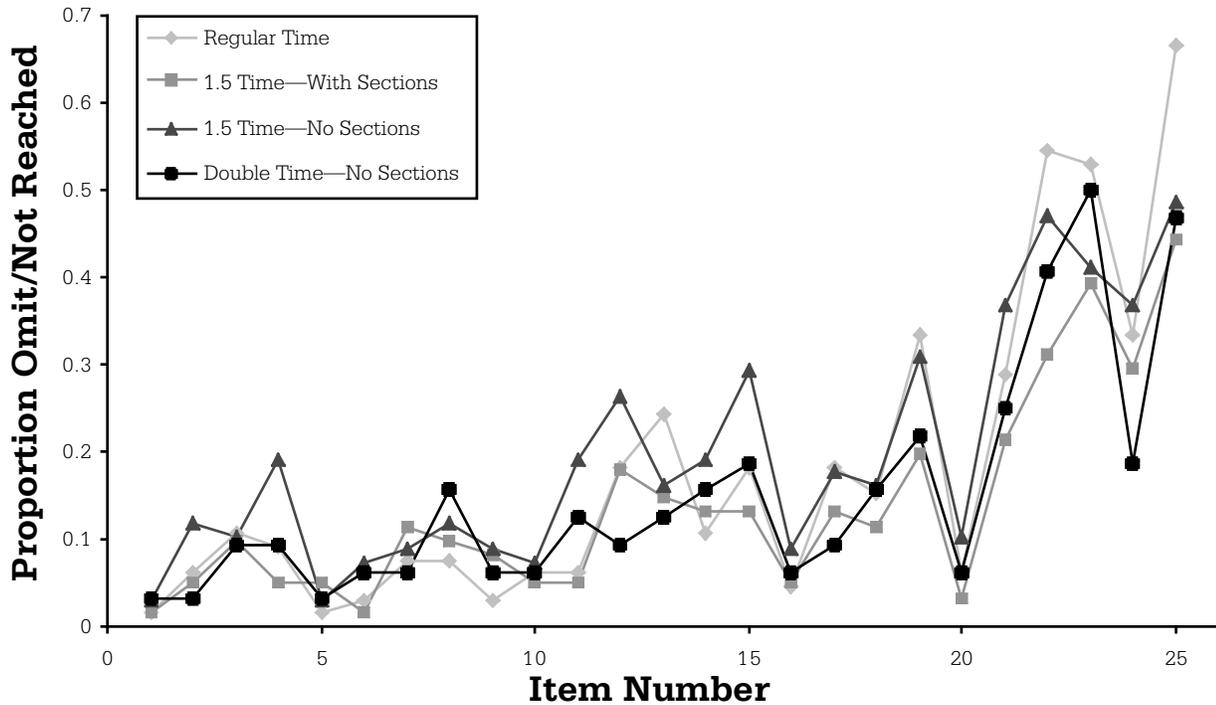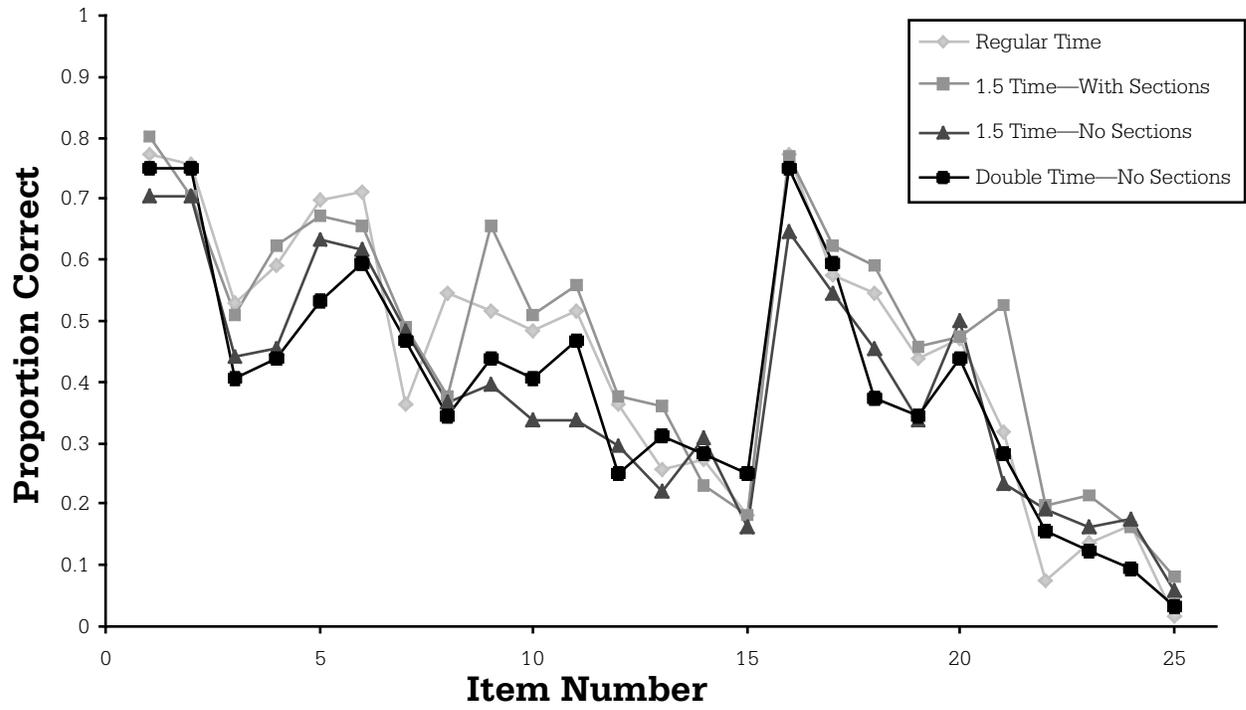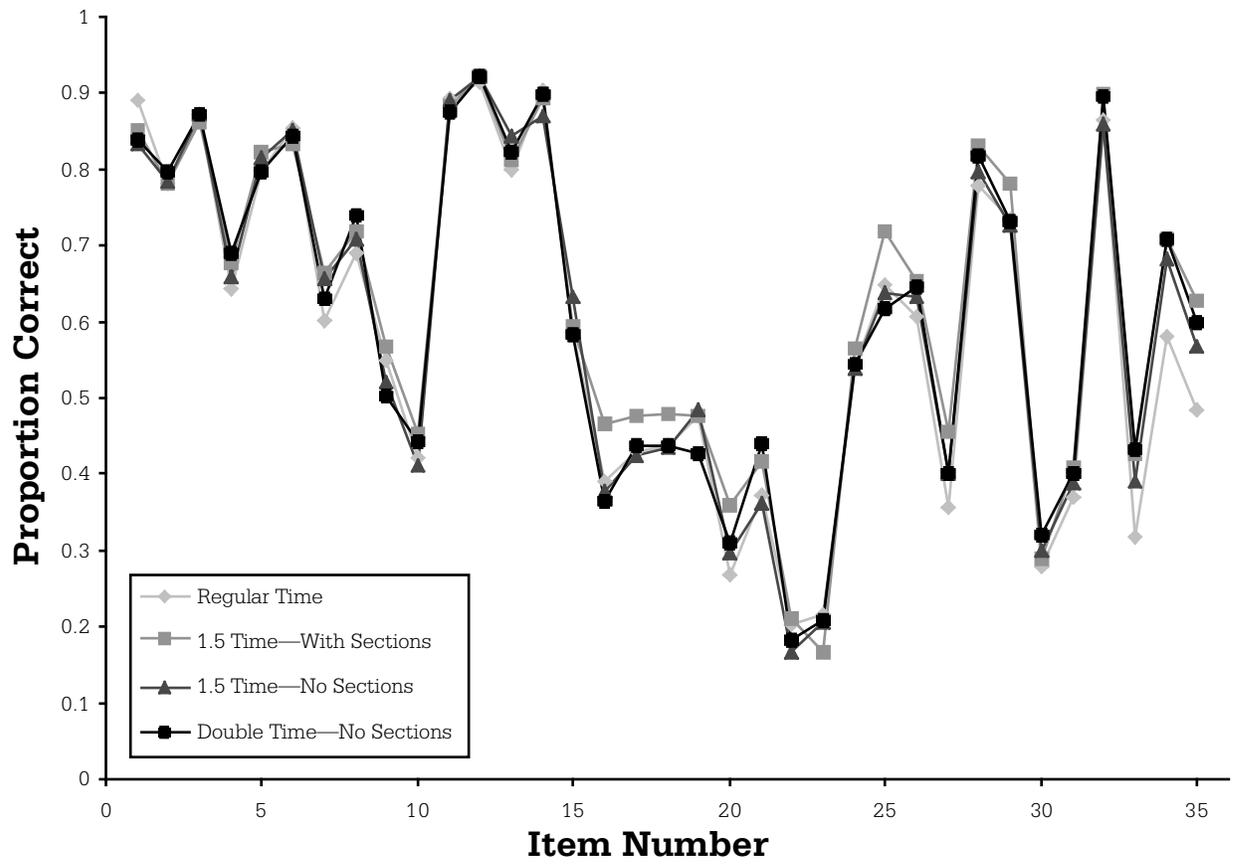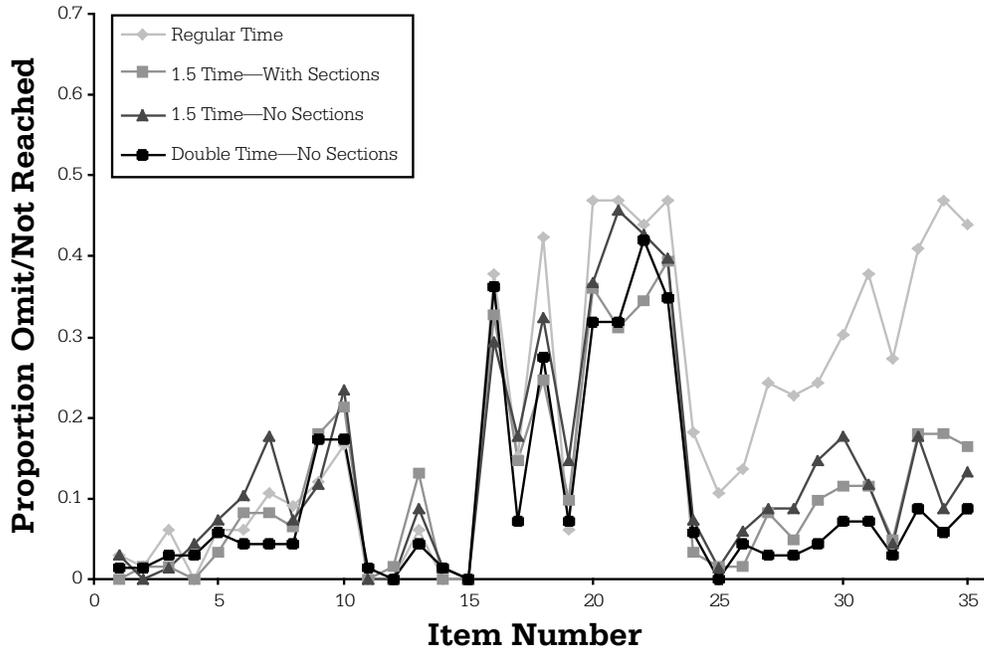**Figure 10.** Proportion Correct: V1–Non-LD.

Figure 11. Proportion Omit/Not Reached: V1–LD.

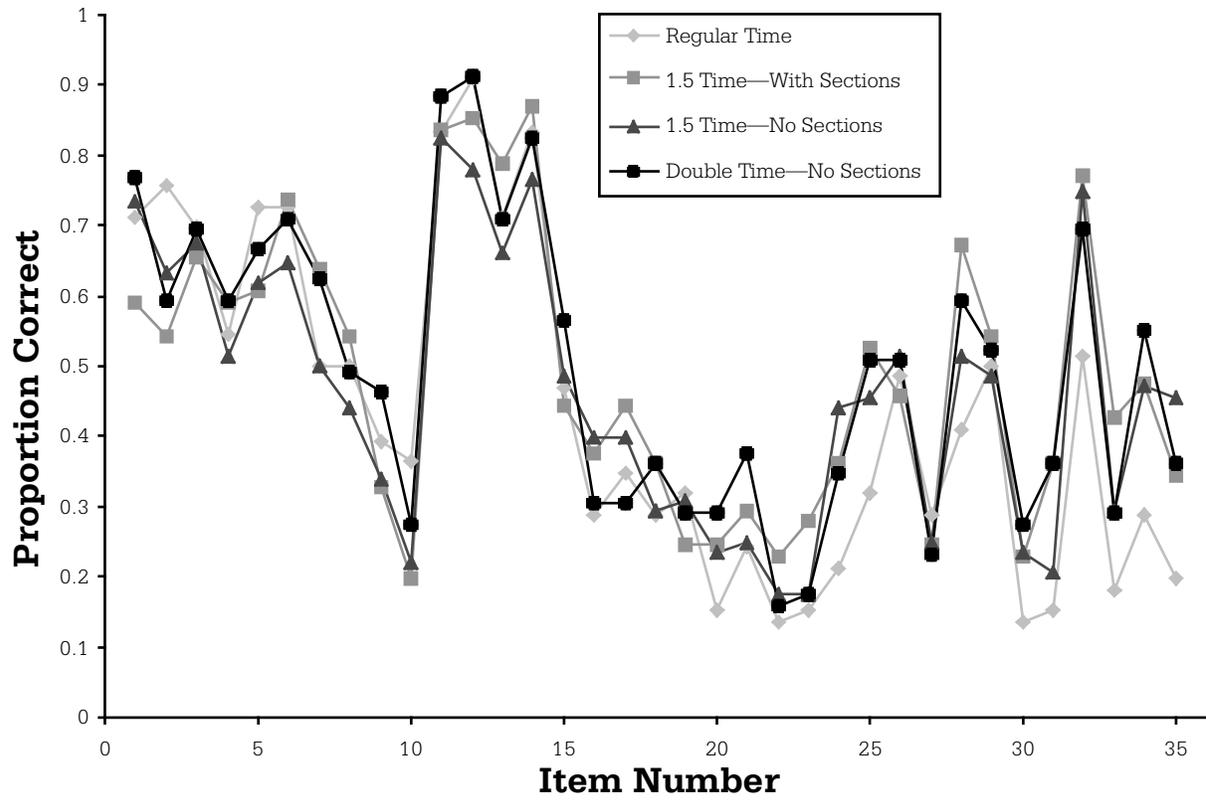|	Regular Time—With Sections N = 66			1.5 Time—With Sections N = 61			1.5 Time—No Sections N = 68			Double Time—No Sections N = 69		
	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR
Item 1	3.0		3.0	0.0		0.0	2.9		2.9	1.4		1.4
Item 2	1.5		1.5	1.6		1.6	0.0		0.0	1.4		1.4
Item 3	6.1		6.1	1.6		1.6	1.5		1.5	2.9		2.9
Item 4	0.0		0.0	0.0		0.0	4.4		4.4	2.9		2.9
Item 5	6.1		6.1	3.3		3.3	7.4		7.4	5.8		5.8
Item 6	6.1		6.1	8.2		8.2	10.3		10.3	4.3		4.3
Item 7	10.6		10.6	8.2		8.2	17.6		17.6	4.3		4.3
Item 8	9.1		9.1	6.6		6.6	7.4		7.4	4.3		4.3
Item 9	12.1		12.1	18.0		18.0	11.8		11.8	17.4		17.4
Item 10	16.7		16.7	21.3		21.3	23.5		23.5	17.4		17.4
Item 11	0.0		0.0	0.0		0.0	0.0		0.0	1.4		1.4
Item 12	0.0		0.0	1.6		1.6	0.0		0.0	0.0		0.0
Item 13	6.1		6.1	13.1		13.1	8.8		8.8	4.3		4.3
Item 14	0.0		0.0	0.0		0.0	1.5		1.5	1.4		1.4
Item 15	0.0		0.0	0.0		0.0	0.0		0.0	0.0		0.0
Item 16	37.9		37.9	32.8		32.8	29.4		29.4	36.2		36.2
Item 17	13.6	1.5	15.2	14.8		14.8	17.6		17.6	7.2		7.2
Item 18	40.9	1.5	42.4	24.6		24.6	32.4		32.4	27.5		27.5
Item 19	4.5	1.5	6.1	9.8		9.8	14.7		14.7	7.2		7.2
Item 20	43.9	3.0	47.0	36.1		36.1	36.8		36.8	31.9		31.9
Item 21	43.9	3.0	47.0	31.1		31.1	45.6		45.6	31.9		31.9
Item 22	40.9	3.0	43.9	34.4		34.4	42.6		42.6	42.0		42.0
Item 23	43.9	3.0	47.0	39.3		39.3	39.7		39.7	34.8		34.8
Item 24	13.6	4.5	18.2	3.3		3.3	7.4		7.4	5.8		5.8
Item 25	1.5	9.1	10.6	1.6		1.6	1.5		1.5	0.0		0.0
Item 26	4.5	9.1	13.6	1.6		1.6	5.9		5.9	4.3		4.3
Item 27	13.6	10.6	24.2	8.2		8.2	8.8		8.8	2.9		2.9
Item 28	7.6	15.2	22.7	3.3	1.6	4.9	7.4	1.5	8.8	2.9		2.9
Item 29	6.1	18.2	24.2	8.2	1.6	9.8	13.2	1.5	14.7	4.3		4.3
Item 30	7.6	22.7	30.3	6.6	4.9	11.5	16.2	1.5	17.6	7.2		7.2
Item 31	12.1	25.8	37.9	6.6	4.9	11.5	8.8	2.9	11.8	5.8	1.4	7.2
Item 32	1.5	25.8	27.3	0.0	4.9	4.9	1.5	2.9	4.4	1.4	1.4	2.9
Item 33	7.6	33.3	40.9	6.6	11.5	18.0	13.2	4.4	17.6	4.3	4.3	8.7
Item 34	7.6	39.4	47.0	1.6	16.4	18.0	2.9	5.9	8.8	1.4	4.3	5.8
Item 35		43.9	43.9		16.4	16.4		13.2	13.2		8.7	8.7

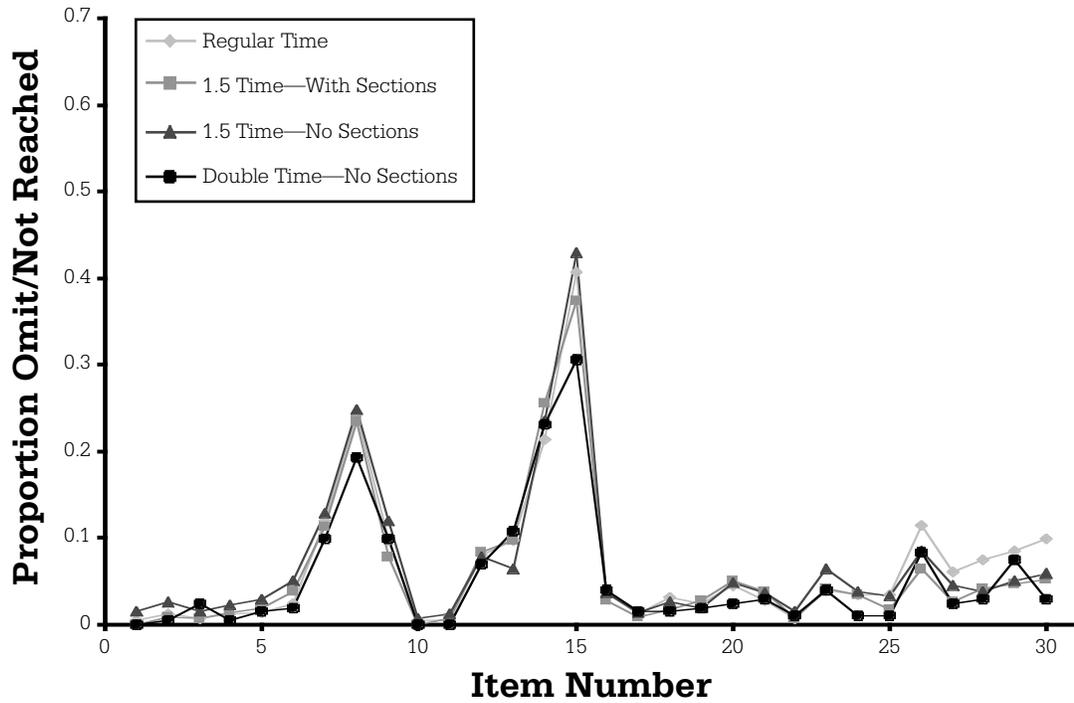**Figure 12.** Proportion Correct: V1–LD.

Figure 13. Proportion Omit/Not Reached: V2–Non-LD.

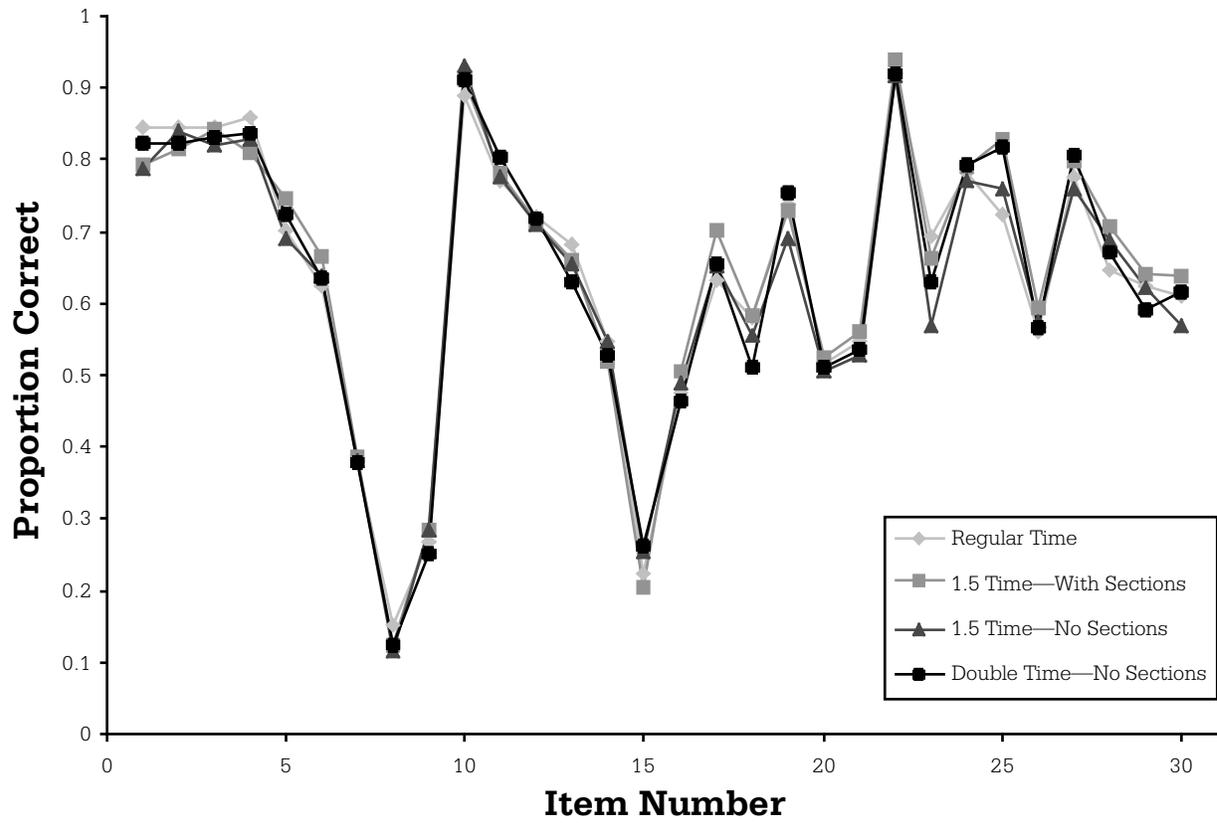|	Regular Time—With Sections N = 425			1.5 Time—With Sections N = 423			1.5 Time—No Sections N = 435			Double Time—No Sections N = 203		
	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR
Item 1	0.5		0.5	0.0		0.0	1.4	0.2	1.6	0.0		0.0
Item 2	1.2		1.2	0.9		0.9	2.3	0.2	2.5	0.5		0.5
Item 3	0.7		0.7	0.7		0.7	1.4	0.2	1.6	2.5		2.5
Item 4	1.2		1.2	1.4		1.4	2.1	0.2	2.3	0.5		0.5
Item 5	1.4		1.4	1.9		1.9	2.8	0.2	3.0	1.5		1.5
Item 6	2.4		2.4	3.8		3.8	4.8	0.2	5.1	2.0		2.0
Item 7	12.0		12.0	11.3		11.3	12.6	0.2	12.9	9.9		9.9
Item 8	24.2		24.2	23.4		23.4	24.6	0.2	24.8	19.2		19.2
Item 9	9.6		9.6	7.8		7.8	11.7	0.2	12.0	9.9		9.9
Item 10	0.5		0.5	0.0		0.0	0.5	0.2	0.7	0.0		0.0
Item 11	0.5		0.5	0.7		0.7	0.9	0.2	1.1	0.0		0.0
Item 12	8.0		8.0	8.3		8.3	7.6	0.2	7.8	6.9		6.9
Item 13	9.6		9.6	9.7		9.7	6.2	0.2	6.4	10.8		10.8
Item 14	21.4		21.4	25.5		25.5	23.2	0.2	23.4	23.2		23.2
Item 15	40.7		40.7	37.4		37.4	42.8	0.2	43.0	30.5		30.5
Item 16	3.8		3.8	2.6	0.2	2.8	3.4	0.2	3.7	3.9		3.9
Item 17	1.2		1.2	0.7	0.2	0.9	1.1	0.2	1.4	1.5		1.5
Item 18	3.1		3.1	1.4	0.2	1.7	2.1	0.5	2.5	1.5		1.5
Item 19	2.4		2.4	2.6	0.2	2.8	1.4	0.5	1.8	2.0		2.0
Item 20	4.5		4.5	4.7	0.2	5.0	4.4	0.5	4.8	2.5		2.5
Item 21	2.8		2.8	3.5	0.2	3.8	3.2	0.5	3.7	3.0		3.0
Item 22	0.5	0.2	0.7	0.5	0.2	0.7	1.1	0.5	1.6	1.0		1.0
Item 23	3.5	0.5	4.0	4.0	0.2	4.3	5.7	0.7	6.4	3.9		3.9
Item 24	2.4	1.2	3.5	3.3	0.2	3.5	3.2	0.7	3.9	1.0		1.0
Item 25	1.4	2.1	3.5	0.9	0.7	1.7	1.8	1.4	3.2	0.5	0.5	1.0
Item 26	8.2	3.3	11.5	5.7	0.7	6.4	6.9	1.6	8.5	7.9	0.5	8.4
Item 27	2.4	3.8	6.1	1.2	1.4	2.6	3.0	1.6	4.6	2.0	0.5	2.5
Item 28	2.4	5.2	7.5	2.8	1.4	4.3	2.3	1.6	3.9	2.5	0.5	3.0
Item 29	2.1	6.4	8.5	3.1	1.7	4.7	3.0	2.1	5.1	6.4	1.0	7.4
Item 30		9.9	9.9		5.2	5.2		6.0	6.0		3.0	3.0

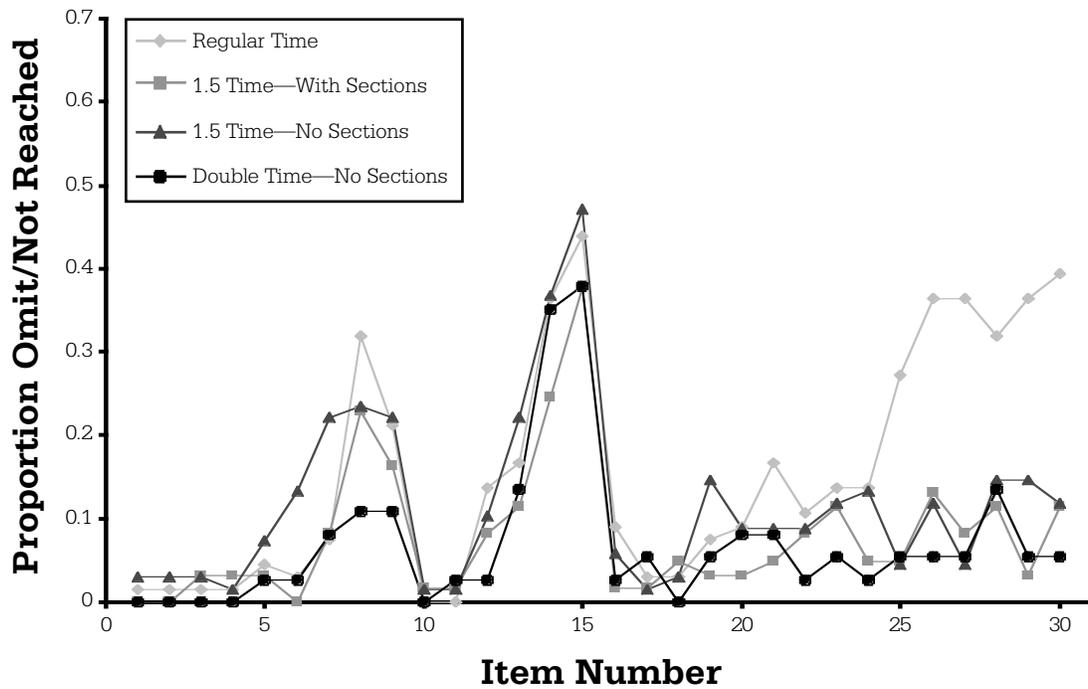**Figure 14.** Proportion Correct: V2–Non-LD.

Figure 15. Proportion Omit/Not Reached: V2–LD.

|	Regular Time—With Sections N = 66			1.5 Time—With Sections N = 61			1.5 Time—No Sections N = 68			Double Time—No Sections N = 37		
	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR	omit	NR	omit/NR
Item 1	1.5		1.5	0.0		0.0	2.9		2.9	0.0		0.0
Item 2	1.5		1.5	0.0		0.0	2.9		2.9	0.0		0.0
Item 3	1.5		1.5	3.3		3.3	2.9		2.9	0.0		0.0
Item 4	1.5		1.5	3.3		3.3	1.5		1.5	0.0		0.0
Item 5	4.5		4.5	3.3		3.3	7.4		7.4	2.7		2.7
Item 6	3.0		3.0	0.0		0.0	13.2		13.2	2.7		2.7
Item 7	7.6		7.6	8.2		8.2	22.1		22.1	8.1		8.1
Item 8	31.8		31.8	23.0		23.0	23.5		23.5	10.8		10.8
Item 9	21.2		21.2	16.4		16.4	22.1		22.1	10.8		10.8
Item 10	0.0		0.0	1.6		1.6	1.5		1.5	0.0		0.0
Item 11	0.0		0.0	1.6		1.6	1.5		1.5	2.7		2.7
Item 12	13.6		13.6	8.2		8.2	10.3		10.3	2.7		2.7
Item 13	15.2	1.5	16.7	11.5		11.5	22.1		22.1	13.5		13.5
Item 14	34.8	1.5	36.4	24.6		24.6	36.8		36.8	35.1		35.1
Item 15	42.4	1.5	43.9	37.7		37.7	45.6	1.5	47.1	37.8		37.8
Item 16	7.6	1.5	9.1	1.6		1.6	4.4	1.5	5.9	2.7		2.7
Item 17	1.5	1.5	3.0	1.6		1.6	0.0	1.5	1.5	5.4		5.4
Item 18	1.5	1.5	3.0	4.9		4.9	1.5	1.5	2.9	0.0		0.0
Item 19	4.5	3.0	7.6	3.3		3.3	13.2	1.5	14.7	5.4		5.4
Item 20	6.1	3.0	9.1	3.3		3.3	7.4	1.5	8.8	8.1		8.1
Item 21	12.1	4.5	16.7	4.9		4.9	5.9	2.9	8.8	8.1		8.1
Item 22	6.1	4.5	10.6	8.2		8.2	5.9	2.9	8.8	2.7		2.7
Item 23	9.1	4.5	13.6	11.5		11.5	8.8	2.9	11.8	5.4		5.4
Item 24	7.6	6.1	13.6	4.9		4.9	10.3	2.9	13.2	2.7		2.7
Item 25	6.1	21.2	27.3	3.3	1.6	4.9	1.5	2.9	4.4	0.0	5.4	5.4
Item 26	15.2	21.2	36.4	11.5	1.6	13.1	8.8	2.9	11.8	0.0	5.4	5.4
Item 27	10.6	25.8	36.4	6.6	1.6	8.2	1.5	2.9	4.4	0.0	5.4	5.4
Item 28	4.5	27.3	31.8	9.8	1.6	11.5	10.3	4.4	14.7	8.1	5.4	13.5
Item 29	6.1	30.3	36.4	1.6	1.6	3.3	8.8	5.9	14.7	0.0	5.4	5.4
Item 30		39.4	39.4		11.5	11.5		11.8	11.8		5.4	5.4
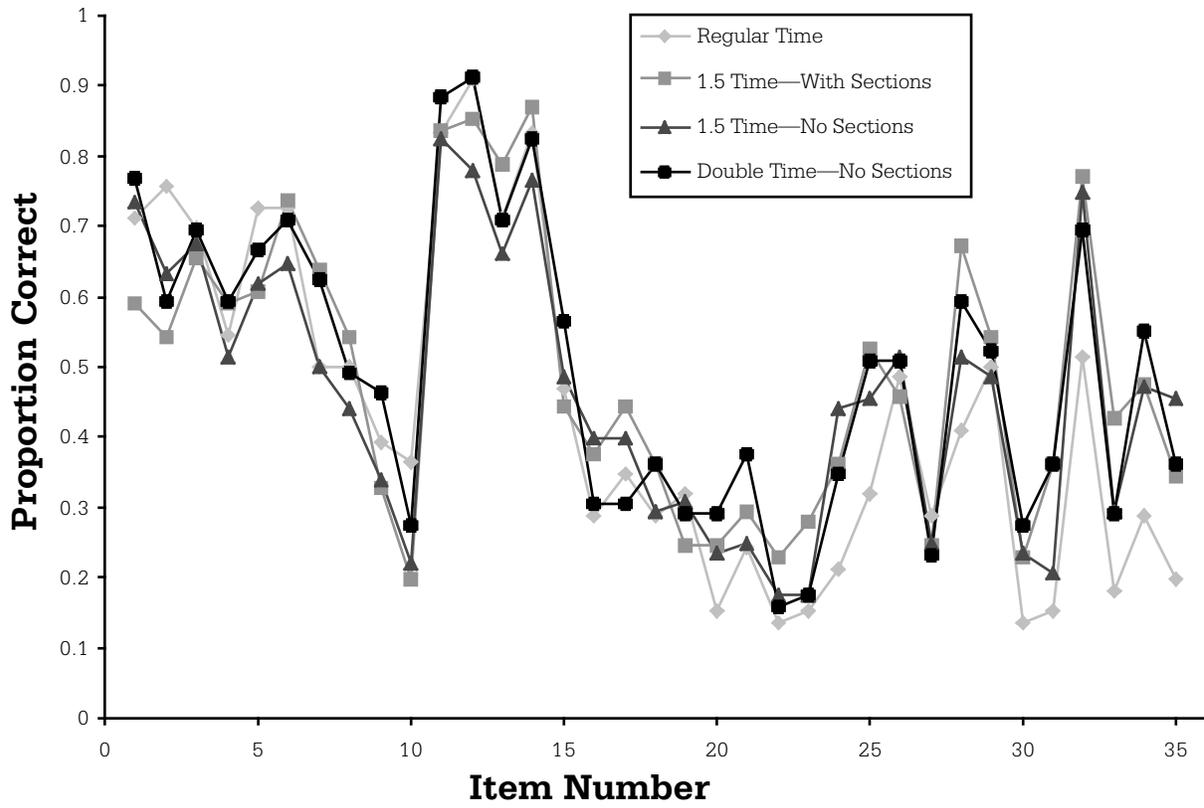
Figure 16. Proportion Correct: V2–LD.