CollegeBoard
connect to college success™

Research Report
No. 2007-3

Time Requirements for the Different Item Types Proposed for Use in the Revised SAT®

Brent Bridgeman, Cara Cahalan Laitusis, and Frederick Cline

www.collegeboard.com

College Board Research Report No. 2007-3
ETS RR-07-35

Time Requirements for the Different Item Types Proposed for Use in the Revised SAT®

Brent Bridgeman, Cara Cahalan Laitusis, and Frederick Cline

The College Board, New York, 2007

Brent Bridgeman is principal research scientist at Educational Testing Service (ETS).

Cara Cahalan Laitusis is a research scientist at ETS.

Frederick Cline is a research supervisor at ETS.

The College Board: Connecting Students to College Success

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

For further information, visit www.collegeboard.com.

Additional copies of this report (item #070482286) 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

Figures

Abstract

The current study used three data sources to estimate time requirements for different item types on the now current SAT Reasoning Test™. First, we estimated times from a computer-adaptive version of the SAT® (SAT CAT) that automatically recorded item times. Second, we observed students as they answered SAT questions under strict time limits and recorded the amount of time taken for each question. Finally, we asked high school students to record the amount of time taken for test subsections that were composed of items of a single type. The rules of thumb used by test developers were quite accurate in rank ordering the item types from least to most time-consuming, but the time actually spent was generally higher than assumed in the rules of thumb.

Introduction

Knowing the amount of time needed for particular item types is useful when test forms must be created that combine several item types in a strictly timed test section. Test developers have rules of thumb that have proven useful for estimating such times, but previously there has been no direct evidence supporting these rules.

The amount of time needed to answer the different item types on the SAT Reasoning Test is well established in the folklore of test developers. For the verbal item types:

> We have found that estimating .5 minutes per analogy, .7 minutes per sentence completion, and 1.0–1.2 minutes per reading item (depending on the length of the passage) helps us to configure sections that are not speeded. However, in adding up these figures, one cannot total, e.g., 30 minutes for a 30-minute section—one must end up [with] around 27–28 minutes worth of items for a 30-minute section not to be speeded. Again, this is very rough and useful only for estimating. (E. Curley, personal communication, November 2002)

For the mathematics item types, the working assumption was 1.0 minutes for each quantitative comparison item, 1.2 minutes for each five-choice regular mathematics question, and 1.5 minutes for each student-produced response (SPR) item (F. Schuppan, personal communication, November 2002).

For the multiple-choice writing items that are now part of the SAT, the working assumption is 30 seconds for sentence error questions and 42 seconds for improving sentences questions. Improving paragraphs questions require three minutes to read the paragraph and one minute to answer each question, or an average of 90 seconds per item for the six-item sets (M. Kubota, personal communication, May 2003). Kubota warned that these times are "very rough estimates because the time involved would depend on the difficulty of the individual item."

There is virtually no empirical research to support the rules of thumb used by test developers. The three studies described here are a first step to address this issue. The first study analyzed existing data from a computer-adaptive version of the SAT (SAT CAT). Computer delivery allows very accurate estimates of the time taken on each item, but because of generous time limits on the SAT CAT, it cannot be used to estimate the time taken under strict time pressure. For the second study, examinees were observed as they answered SAT questions under strict time limits, and then the average time it took to answer each type of question was determined. The third study used the same test forms as the observational study, but students were asked to record their starting and stopping times for sections consisting of homogeneous item types. Together, the three studies provide a picture of the amount of time that was taken to answer different types of questions.

Study 1—Item-Type Times from a Computer-Adaptive SAT®

Item timing information is not available from a typical paper-and-pencil test administration, but it is routinely collected with tests delivered by computer. The SAT CAT is one such test. Although currently used for selecting very high-achieving middle school students for talent search programs, it was developed to the same general difficulty specifications as a paper-and-pencil SAT and uses the same item types (albeit slightly modified for screen presentation). Two of the item types are no longer used on the SAT. We include these two types—analogies from the verbal section and quantitative comparisons from the math section—in the current analysis to provide some context for evaluating the time demands of different item types.

The time limits for the SAT CAT were very generous relative to the paper-and-pencil SAT; 33 verbal questions in one hour for the CAT (or 1.8 minutes per question) compared to 78 questions in 75 minutes for the paper version (or .96 minutes per question). The math CAT section had 28 questions to be answered in an hour (or 2.14 minutes per question) compared to 60 questions in 75 minutes on the paper test (or 1.25 minutes per question). Therefore, absolute time spent on the items on the CAT is not a good indicator of the time that would be taken

during a paper-and-pencil administration. Nevertheless, the relative amount of time spent on different item types should still generalize to a paper-based test, though the generalizations are imperfect.

Sample

The SAT CAT timing data was obtained as part of the study that put the SAT CAT and the paper-and-pencil SAT on the same scale. Details of this study are described in Lawrence and Feigenbaum (1997). Briefly, participants were high school juniors who had registered to take the SAT and who lived near a computer-based testing (CBT) center. They took a paper-and-pencil version of the SAT in May and the CAT version in June. The participants were highly motivated to do their best because only the higher score from these two administrations would become part of their official SAT record. From the sample of 1,732 tested with both the paper-and-pencil and CAT versions of the SAT, we removed the few students who were apparently not taking the test seriously because they spent less than five minutes on an entire math or verbal measure, resulting in a final sample of 1,719. The sample was reasonably representative of the college-bound, SAT-taking population, though means were slightly above average (530 and 542 for verbal and math, respectively).

SAT CAT Analysis

For each person, we computed the time it took to answer all questions of a particular type, then divided by the number of those items administered. This average time per question for each individual provided an estimate that could be compared across sections of different lengths. Because this test had very generous time limits, and virtually nobody ran out of time, these averages reflected solution times under minimal time pressure.

Verbal

The distributions of the average times for the verbal item types are shown in Figures 1–4. Comparing Figures 1 and 2 suggests that examinees under minimal time pressure take about 11 seconds longer on average to answer sentence completion questions than analogy questions, or that the time required for a sentence–completion item is 1.3 times the time taken for an analogy. The rule of thumb that the verbal test development experts use is that analogy items take about 30 seconds each, and sentence completion items take about 42 seconds each. By the rule of thumb, sentence completion items require about 12 more seconds than analogy questions, or 1.4 times as much time. Thus, the difference in the actual data is similar to the difference in the rule of thumb.

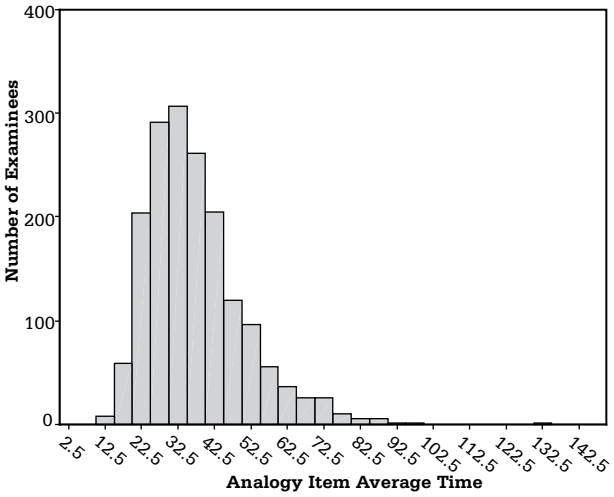

Figure 1. Distribution of mean time per item (in seconds) for analogy items. (The mean of these means is 37 seconds with an SD of 13; the 80th percentile is 46 seconds.)

We also investigated mean times by ability levels on the previously administered PSAT/NMSQT®. This is a paper-and-pencil test that high school juniors take as preparation for the SAT and to qualify for scholarships. It contains the same item types as the SAT. We used the PSAT/NMSQT scores to divide the sample into a bottom quarter, middle half, and top quarter. In a nonadaptive test, higher-ability students would generally be expected to be faster, but in an adaptive test, higher-ability students receive more difficult items. When the difficulty of the items is more closely matched to the ability of the examinees, higher-ability examinees may no longer have a time advantage. Indeed, time differences were generally small and inconsistent across the ability groupings.

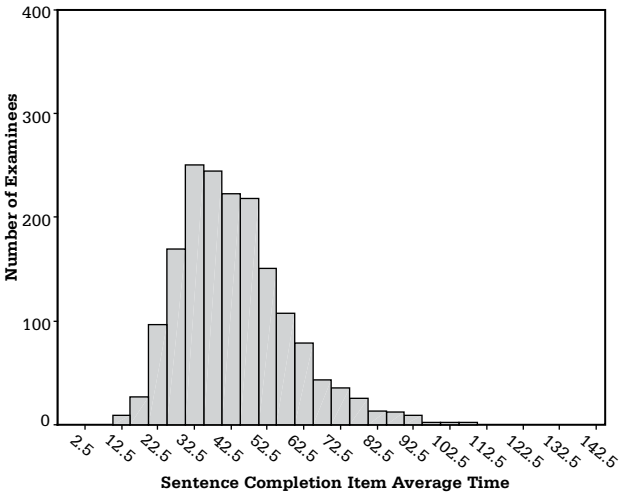

Figure 2. Distribution of mean time per item (in seconds) for sentence completion items. (The mean of these means is 48 seconds with an SD of 15; the 80th percentile is 59 seconds.)

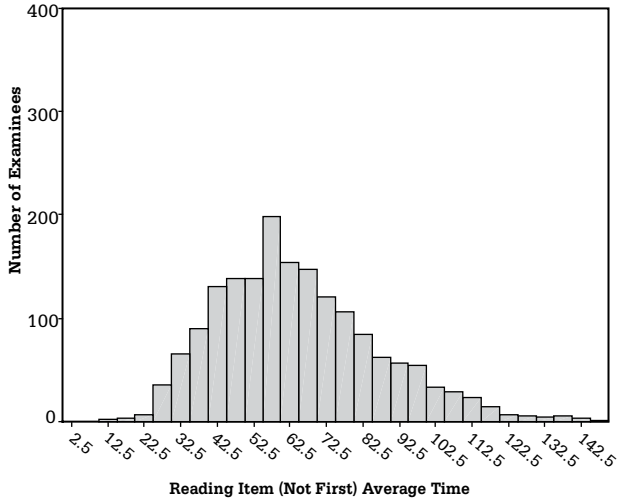

Figure 3. Distribution of mean time per item (in seconds) for reading items following the first item. (The mean of these means is 65 seconds with an SD of 22; the 80th percentile is 82 seconds.)

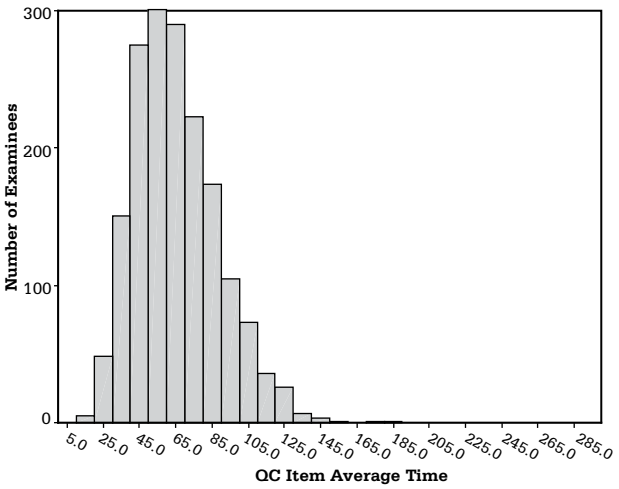

Figure 5. Distribution of mean time per item (in seconds) for quantitative comparison items. (The mean of these means is 65 seconds with an SD of 23; the 80th percentile is 84 seconds.)

The reading portion of the SAT CAT consisted of three passages with five questions each. Passage length was similar to passage length in the paper-and-pencil test except that there was no long (800-word) passage. For the reading questions, the time for the first item starts when the passage is first displayed and runs until that question is answered. Because this initial passage reading time is included in the time for the first item, Figure 3 includes the average time only for items *after* the first item in a set. Figure 4 shows the time for the *first* item; it includes the reading time. Note that Figures 1–3 can be directly compared because they are on the same scale. Figure 4 uses a different horizontal scale to accommodate the longer times.

The average time for the entire set, including initial reading time and time spent answering the five questions, was 484 seconds. Dividing this time by five results in an estimate of 97 seconds per item. This compares to the rule-of-thumb estimate of 60 to 72 seconds.

Math

The distributions of the average times for math items are presented in Figures 5–7. All of these figures are on the same scale, though the scale is different from the scale used for the verbal items because average times are generally longer for the math items.

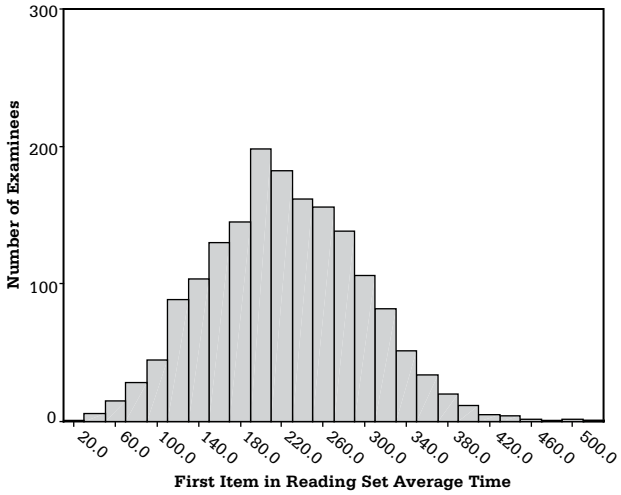

Figure 4. Distribution of mean time per item (in seconds) for initial reading item in a set. (The mean of these means is 224 seconds with an SD of 75; the 80th percentile is 286 seconds.)

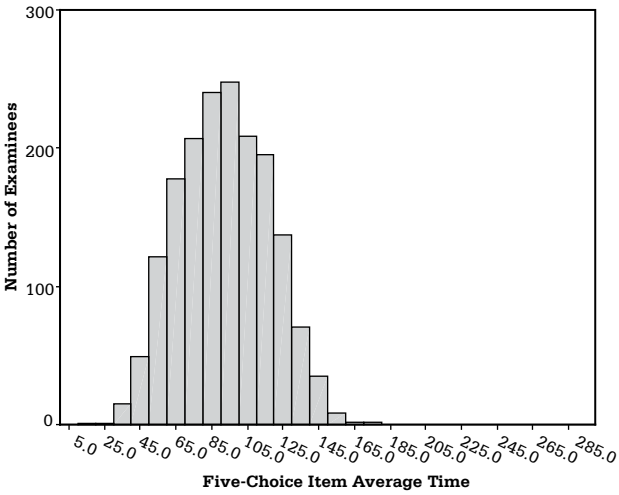

Figure 6. Distribution of mean time per item (in seconds) for five-choice regular math items. (The mean of these means is 92 seconds with an SD of 25; the 80th percentile is 116 seconds.)

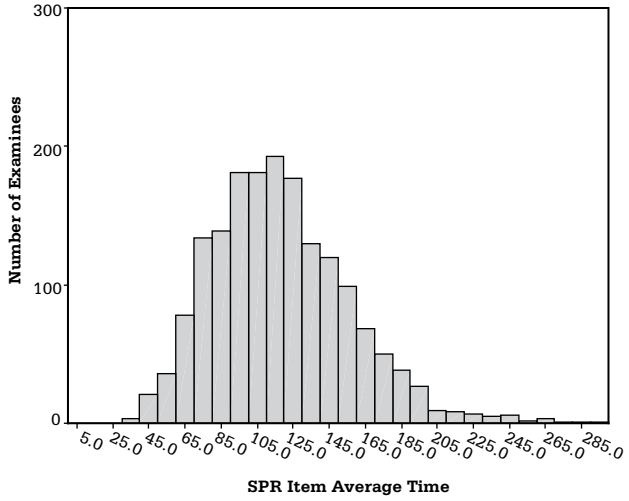

Figure 7. Distribution of mean time per item (in seconds) for SPR math items. (The mean of these means is 118 seconds with an SD of 38; the 80th percentile is 148 seconds.)

The mean time taken on the SAT CAT for the five-choice items was 27 seconds longer than the quantitative comparison (QC) items (or 1.4 times as long), confirming the belief that QC items can generally be answered more quickly. The test developer rule of thumb for the paper-and-pencil SAT is 60 seconds for QC items and 72 seconds for five-choice items (or 1.2 times as long), which appears to underestimate the actual time advantage of the QC format.

Student produced response (SPR) items took an average of about 26 seconds longer than five-choice items (1.3 times as long). The test developer rule of thumb is 72 seconds for five-choice items and 90 seconds for SPR items, for a difference of 18 seconds (or 1.3 times as long).

Conclusion

For all of these analyses, examinees were not under typical SAT time pressure; the relative time advantage of a particular item type could shift with stricter time limits. Nevertheless, the relative amount of time that was taken to answer each item type agreed remarkably closely with the rules of thumb used by test developers.

Study 2—Item-Type Times from an Observational Study

In Study 2, examinees were observed (via one-way mirror and/or video recording) as they answered SAT questions under timed conditions. This study extended the findings from Study 1 in several important ways. First, the time limits in Study 1 were unrealistically long for a typical paper-and-pencil SAT. The time limits in Study 2 were set to be more realistic. Second, Study 2 used a paper-and-pencil format with answers gridded on a standard SAT answer sheet. Third, the SAT CAT contained only item types from the old version of the SAT; Study 2 included item types from the current SAT.

Development of Test Forms

Expert SAT test developers assembled sections in critical reading, writing, and mathematics with the same item types as on the SAT. Each section was designed for a 35-minute test block. Although the time given for each section of the SAT is 25 minutes, we used 35-minute blocks to be able to assess all item types in a single block; thus, the 35-minute blocks contained more items than are included in a 25-minute SAT section. Because there is no way to estimate solution times for unreached questions, two forms were created for each section, with blocks of items presented in reverse order so that items near the end that might be unreached in one form would occur early in the other form. The test forms contained the standard directions for each item type. In order for the observers looking through the one-way mirror to tell which question an examinee was working on, each question was written on a separate page with a large (about 2 inches) question number at the top of each page. In order to avoid the need to flip back over up to nine pages between reading passages and their associated questions, items on a single reading passage were printed on the same page, and time was recorded for the block of items. Samples of the different item types evaluated are in the Appendix.

Critical Reading

The SAT has replaced the analogy items on the old version of the test with more passage-based reading questions. To reflect this change, the verbal scale is now named the critical reading scale. The critical reading section in this study contained 9 sentence completion items, 10 paragraph reading items based on 100-word passages (2 questions for each of 5 passages), and 18 items based on two 650-word reading passages (9 per passage). One of the 650-word passages was a personal narrative, and the other was a more textbook-like science text. We did not have any 800-word or paired passages, or any 500-word passages, but the 650- and 100-word passages were deemed representative of the full range of passages included on the SAT. The test developers believed that the configuration for this study would be slightly more speeded than a standard SAT (E. Curley, personal communication, September 2002). One critical reading test form (R1) began with the sentence completion items, while the other (R2) began with the 100-word passages and concluded with the sentence completion items.

Writing

The writing section was taken from an existing PSAT/NMSQT form that was administered with a 30-minute time limit. The test was thus slightly less speeded in this study, but we had to keep the timing for all tests the same (35 minutes) because students in the same room would be working on different tests. The test contained 19 items on identifying sentence errors, 14 on improving sentences, and 6 items on improving paragraphs. Form W1 began with the sentence error items, and Form W2 began with the improving sentences items and concluded with the sentence error items.

Mathematics

Form M1 started with 13 five-choice items followed by 10 SPR items in which numerical answers are gridded in a special block on the answer sheet. Form M2 started with the SPR items.

Participants

Invitation letters were sent to a sample of high school juniors who had previously taken the PSAT/NMSQT and who lived within a 25-mile radius of the ETS Princeton office. The letters invited the juniors to apply to participate in a research project that would give them practice with real SAT questions in return for a payment of $25, plus an additional $25 if they performed about as well on the experimental tests as they did on the PSAT/NMSQT. This additional payment was intended to make sure that the participants would take the experimental test seriously. (All students did appear to take the test seriously, and all were paid $50.) Participants were selected from among the applicants to ensure that a broad range of PSAT/NMSQT scores was represented; specifically, the first 20 volunteers in each of three score levels based on the combined verbal and math scores (40–80; 81–120; 121–160) were scheduled to be observed.

Procedures

Up to three participants were seated around a round table that was placed next to the one-way mirror connected to the observation room. When the participants arrived, the light was on in the observation room, allowing them to see the video cameras and observers. The participants read general directions that explained the procedures, were informed that they would be watched and videotaped from behind the one-way mirror, and were told that they would be taking two tests of 35-minutes each. They were further informed, "Directions for each question type are printed at the beginning of the questions of that type. These directions are the same as you saw previously when you took the PSAT/NMSQT." They were given the standard directions for marking answers (e.g., "Make sure each mark is dark and completely fills the oval," "If you erase, do so completely," "Use the test book for scratchwork") and standard scoring directions (e.g., "For each correct answer, you receive one point; for questions you omit, you receive no points; for wrong answers to multiple-choice questions, you lose a fraction of a point…"). The administrator asked if there were any questions and told them to begin when they saw the light go out in the observation room. Typically, each person in the room was working on a different section (critical reading, writing, or mathematics). After 35 minutes, participants moved on to a second section. Thus, each participant took sections in two of the three areas. The observer could record timing information for one or two examinees in real time; timing information that could not be recorded live was obtained from viewing the videotape.

Results and Discussion

Computation of the mean solution time for an item type was typically straightforward and was similar to the computation in Study 1 (compute total time on an item type for each examinee and divide by the number of items of that type). However, items that were not attempted because time ran out were problematic. The time spent on such unreached items was zero seconds; including these items would clearly distort computations of the time necessary to answer a question of that type. Although "unreached" items are typically at the end of a test, we observed a number of cases in which an examinee glanced over an earlier item for a few seconds without truly considering it, apparently hoping to return to it at the end of the test. Such skipped items would also distort computations of mean time to answer. Therefore, we decided that only items considered for at least 10 seconds would be included in the computation of mean time to answer. When item times and correlations of times with PSAT/NMSQT scores were comparable across the test forms that presented the subsections in different orders, they were combined. When results appeared to differ by form, we have reported them separately.

Critical Reading

Because of taping problems, timing information on two examinees was lost. Usable data were obtained from 24 R1 examinees and 22 R2 examinees.

Sentence Completion

Among the R2 examinees (who took the passage-based items before the sentence completion items), four ran out of time before getting to any sentence completion items. Figure 8 shows the amount of time, in 10-second intervals, spent on reading only the instructions for the sentence completion items for the remaining 42 examinees. The directions for the sentence completion items consist of three sentences and a sample item.

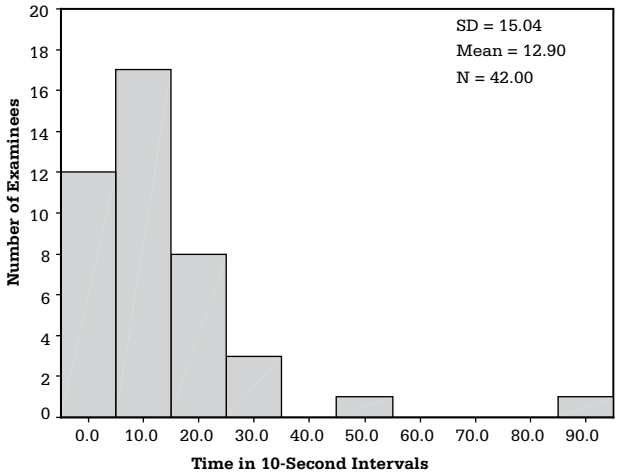

Figure 8. Histogram of time taken to read the instructions for sentence completion items.

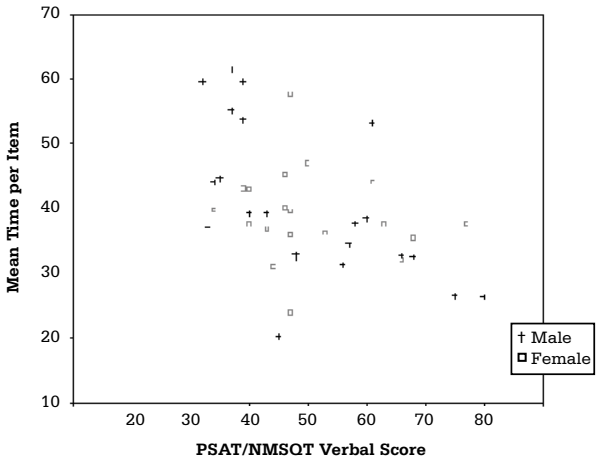

Figure 10. Scatterplot of time taken for sentence completion items and PSAT/NMSQT verbal scores by gender.

Over half of the examinees took 15 seconds or less to read the instructions, including 12 students who skipped the directions entirely.

All of the examinees appeared to make a serious effort to answer the sentence completion items on their first attempt. Sixteen examinees (38 percent) went back to review at least one of these items. The mean total time, adding the review time to the first attempt time, was about five seconds longer than the mean time for first attempt alone. Note that this average review time resulted from reviewing just a few items and not reviewing most items. Therefore, the time spent reviewing an item that the examinee chose to review would typically be much longer than five seconds. The histogram for total time is presented in Figure 9.

The mean time per item was 40 seconds, with 24 percent of the examinees taking longer than 45 seconds. This is eight seconds less than the mean time found in the Study 1 (SAT CAT) sample; shorter times are to be expected because time limits were stricter in Study 2 than in Study 1. Lower-ability examinees tended to take somewhat longer than higher-ability examinees. The correlation of mean time with PSAT/NMSQT verbal score was -.48. This relationship can be seen in the scatterplot in Figure 10, which also shows that lower-ability males took the longest time. The plot also suggests that the correlation of time with ability is stronger in men than in women, which is indeed the case (-.60 for men and -.16 for women).

At the level of the individual item, results were clearest when restricted to the form in which the sentence completion items were administered first. In this form, six of the nine items were administered in an average time in the 30- to 40-second range. The average time for the fastest item was 24 seconds, and the slowest item was 63 seconds. The correlation of item difficulty

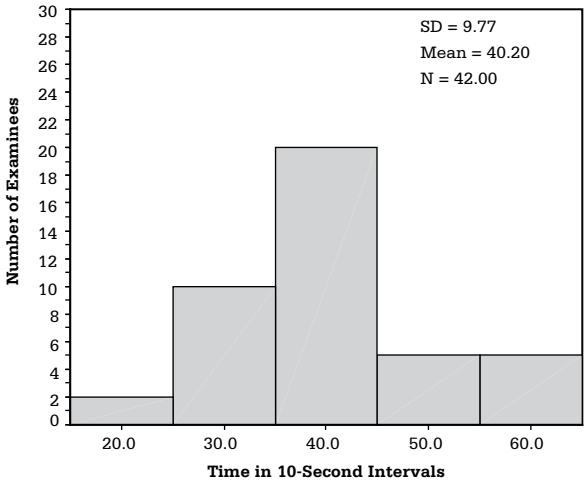

Figure 9. Histogram of time taken for sentence completion items.

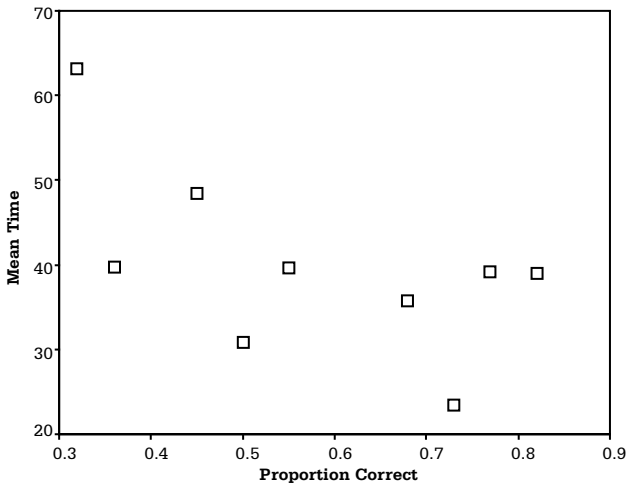

Figure 11. Scatterplot of the mean time taken for each sentence completion item and the proportion correct for that item.

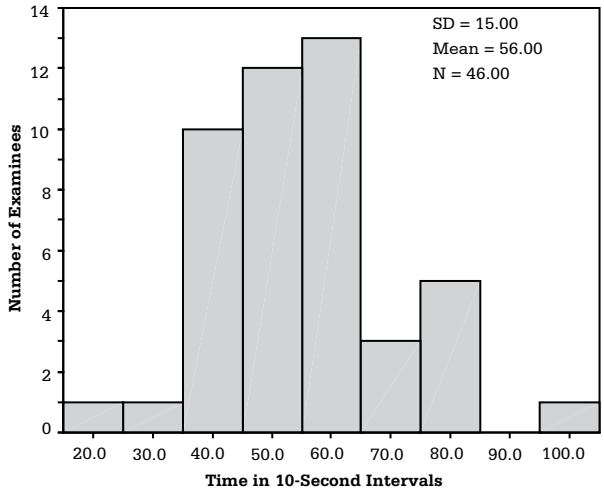

Figure 12. Histogram of the time taken for items from paragraph reading passages.

(percent correct) with item time was -.59, indicating that hard items generally take longer than easier items. This relationship can be seen in the scatterplot in Figure 11.

Paragraph Reading

Although nominally "100-word passages," the five passages in this section contained a total of 383 words, or about 77 words per passage. A single set of instructions was used for the 100-word passages and the longer passages. Examinees spent relatively little time on these instructions. The longest time spent was 40 seconds, and only five students took more than 20 seconds.

We timed each passage and its pair of questions as a block and divided by two to get per-question time. There was relatively little review of these items with only seven students spending any time beyond the first attempt; the mean total time was only one second longer than the mean time for first attempt. Figure 12 is the histogram of the mean total time.

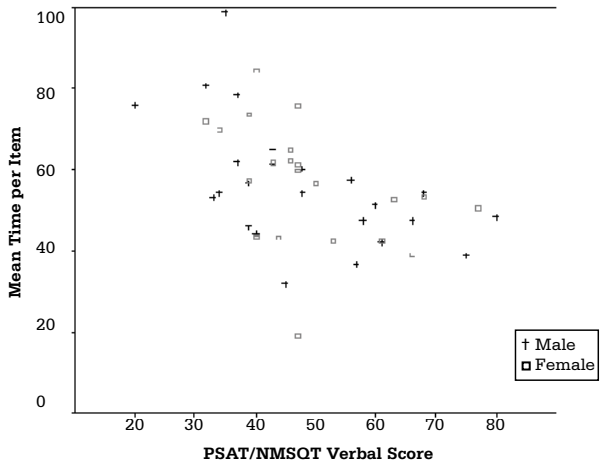

Figure 13. Scatterplot of the mean time taken for items from paragraph reading passages and PSAT/NMSQT verbal scores by gender.

The mean time per item was 56 seconds and the median was 54 seconds. About 80 percent of the examinees took less than 65 seconds and only 13 percent took more than 75 seconds. The correlation pattern with PSAT/NMSQT verbal score was similar to that for the sentence completion items (r = -.49), though the gender difference was smaller (r = -.56 and -.40 for males and females, respectively). This pattern can be seen in the scatterplot in Figure 13. For examinees in this sample with PSAT/NMSQT verbal scores over 50, the mean item time was under a minute for every examinee; for PSAT/NMSQT scores under 50, there were about as many means over a minute as under.

650-Word Passages

Among the 24 examinees who took Form R1 (with reading passages last), 12 clearly ran out of time before completing the second 650-word passage, and a few others had such short times that it seems likely that they also ran out of time. Therefore, we decided to use only R2 data in which the passages were administered before the sentence completion items. We found it impossible to measure separately the amount of time spent reading the passage and the amount of time answering questions because about a third of the examinees started with the questions and then flipped back through the passage to find the answers. We recorded the total time from the time they turned the page to begin the first passage until they answered the last question on that passage and divided that time by the number of items on a passage (nine) to get the per-item time. This per-item time therefore includes both the time spent reading the passage and the time spent reading and answering the questions. The first passage was a personal narrative in which "the narrator considers his family's history and migration from Mexico to Texas, which was once part of Mexico." The histogram for the time taken for this passage and its associated questions is presented in Figure 14.

The second 650-word passage was about rapid climate change toward the end of the latest Ice Age. One student apparently ran out of time in the middle of this passage and was dropped from the analysis. As indicated in Figure 15, the time taken for this passage was reasonably close to the time taken for the first 650-word passage; the means were virtually identical and no one averaged less than 45 seconds per item for either passage. However, only two students took more than 75 seconds. It is not clear whether the absence of longer times for the second passage is content related, or merely reflects increased hurrying as examinees realize time is running short and they still must complete all of the sentence completion items.

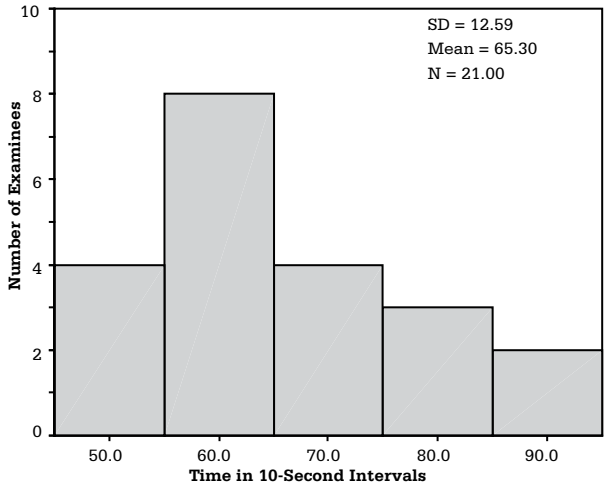

Figure 14. Histogram of time taken for items from the first 650-word passage.

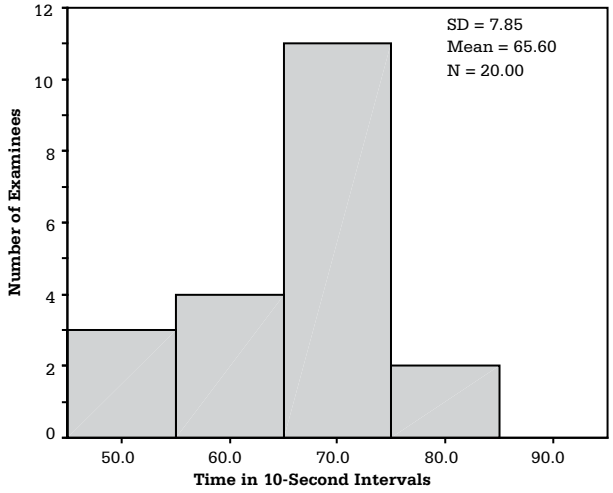

Figure 15. Histogram of time taken for items from the second 650-word passage.

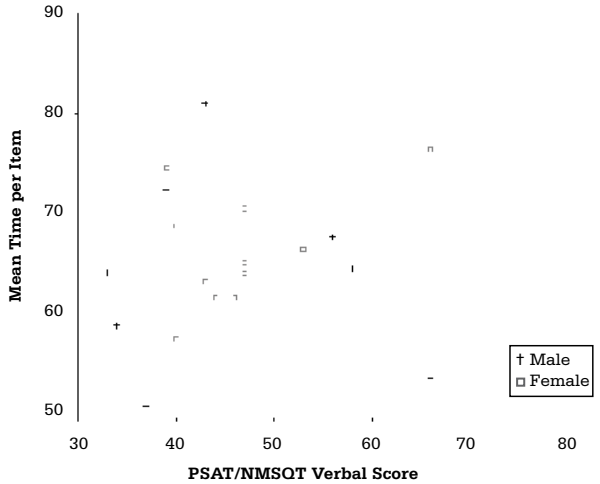

Figure 16. Scatterplot of the mean time taken for items from combined 650-word passages and PSAT/NMSQT verbal scores by gender.

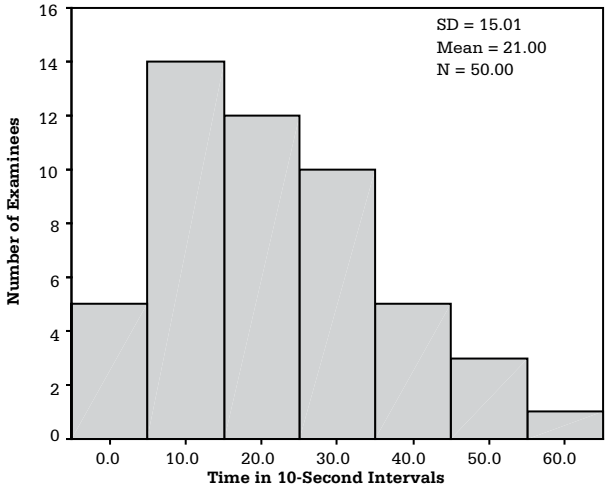

Figure 17. Histogram of time taken to read instructions for identifying sentence error items.

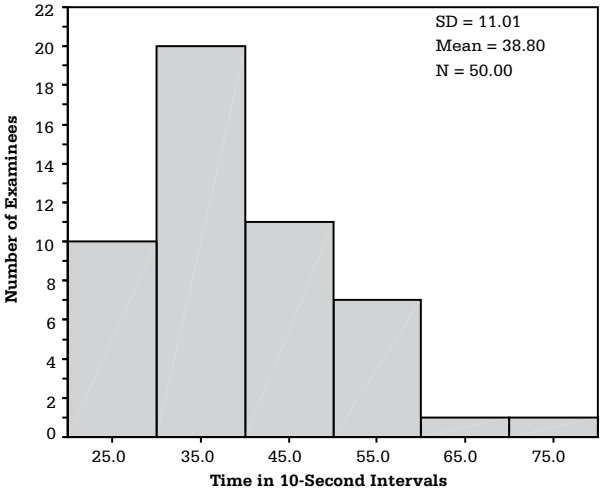

Figure 18. Histogram of time taken for identifying sentence error items.

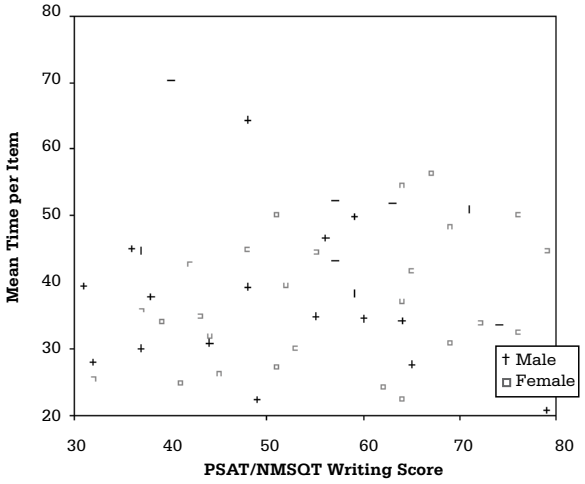

Figure 19. Scatterplot of the mean time taken for sentence error items and PSAT/NMSQT writing scores by gender.

For the scatterplot in Figure 16, average times for both 650-word passages were combined. The correlation between the time taken and PSAT/NMSQT verbal score was somewhat different for the two passages and was also dependent on form. For Form R1, in which the second 650-word passage was the last subsection, there was a substantial and statistically significant positive correlation (r = .68), apparently because high-ability students who worked quickly on other parts of the section had time remaining to spend on this last subsection. For Form R2, the correlation was still positive for this passage, but the correlation was only .20. For the first 650-word passage, the correlation was -.29.

Purely from a time perspective, sentence completion items are more efficient than items based on short or long passages; about three sentence completion items can be asked in the same time needed for two (or even slightly less than two) passage-based reading comprehension questions. Of course, time efficiency is not the only consideration, and it is difficult to imagine a reading test in which examinees never had to read more than one or two sentences at a time. Observed times were remarkably close to the rules of thumb used by test developers (42 seconds for sentence completion items, and 60 to 72 seconds for reading comprehension items from longer passages; there was no rule of thumb yet established for the new 100-word passage items).

Writing

Although examinees experienced some time pressure while taking the writing section (39 items in 35 minutes), the pressure was not as extreme as during the reading section. Only one student ran out of time before getting to the last item, and the mean times were within three seconds whether the section was the first administered or the last.

Identifying Sentence Errors

The histogram of the time spent on the instructions is shown in Figure 17. Although most students spent little time on these instructions, 38 percent spent more than 25 seconds.

The histogram of the time per item is in Figure 18. With a mean time per item of 39 seconds and 98 percent of examinees with a mean of less than a minute, the sentence error items are comparable in time-efficiency to the sentence completion items in the critical reading section. Nevertheless, they appear to be more time-consuming than the test developer estimate of 30 seconds per item. Unlike the sentence completion items, the solution time was unrelated to ability, with a correlation of only .07 (or .27 for the 24 examinees who took the form in which this subsection was not in the last position; this correlation is not statistically significant in this small sample). This lack of correlation is evident in the scatterplot presented in Figure 19.

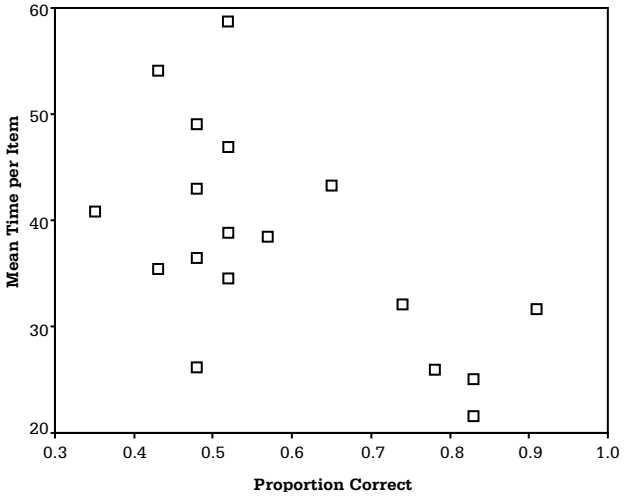

Figure 20. Scatterplot of the mean time taken for each sentence error item and the proportion correct for that item.

For individual items on the form in which sentence error items were first, the mean times ranged from 22 seconds to 59 seconds. Four of the 19 sentence error items had mean times under 30 seconds and four had mean times over 45 seconds. The correlation of the item difficulty and the mean time for an item was -.59. This relationship can be seen in Figure 20.

Improving Sentences

The time taken for instructions is presented in Figure 21. The distribution of times is very similar to that for the sentence error items. Students who take a long time on one set of instructions also tend to take a long time on the other (r = .59; p < .01) and tend to be students with lower writing skills as indexed by the correlation with the PSAT/NMSQT writing score of -.29 (p < .05). Wasting valuable testing time reading instructions is

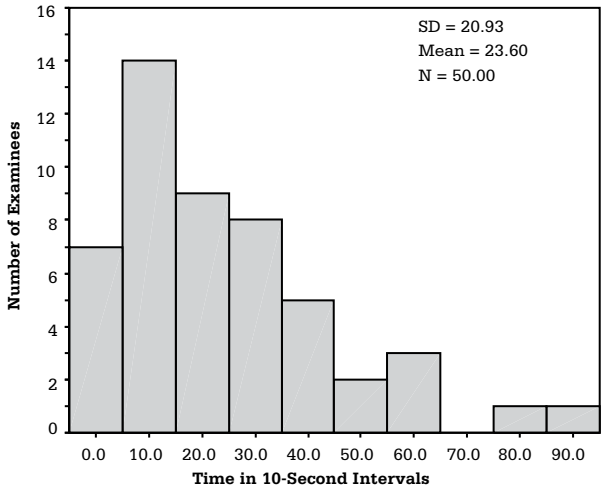

Figure 21. Histogram of time taken to read instructions for improving sentences items.

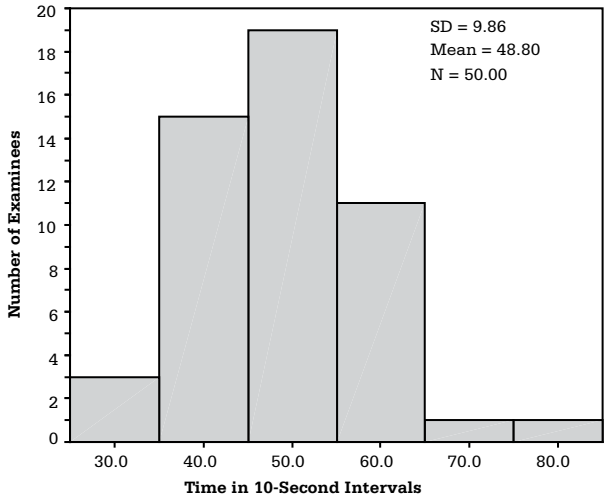

Figure 22. Histogram of time taken for improving sentences items.

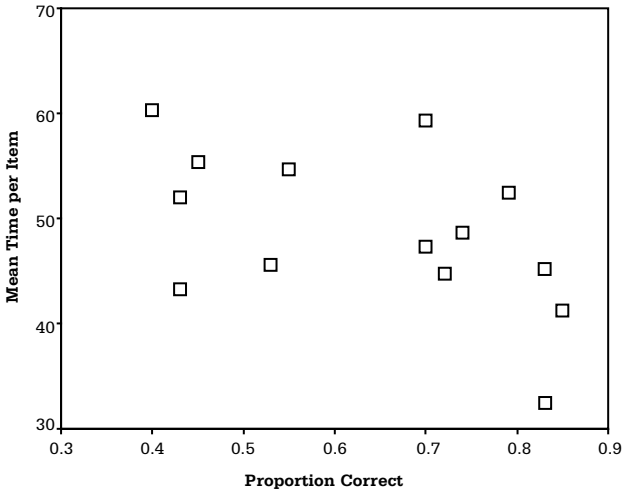

Figure 24. Scatterplot of the mean time taken for each sentence completion item and the proportion correct for that item.

inefficient. Test-preparation materials should continue to emphasize that the more time that is spent reading instructions, the less time that is available to answer the questions.

Figure 22 shows the mean time spent for the improving sentences questions.

These items took about 10 seconds longer per item to answer than did the sentence error questions, and were somewhat more time-consuming than the test developer estimate of 42 seconds. As with the sentence error items, time and writing ability (PSAT/NMSQT writing score) were uncorrelated (r = -.06). This is also evident in Figure 23.

For individual improving sentences items, the mean times ranged from 32 to 60 seconds. The scatterplot of the proportion correct and the mean time for an item is in Figure 24. The correlation represented in this figure was -.50.

Improving Paragraphs

The distribution of time spent reading the instructions was virtually identical to the distribution for the improving sentences subsection.

As indicated in Figure 25, the time spent on initially reading the passage was quite variable, with nine examinees spending less than 15 seconds and five examinees spending more than 95 seconds.

As we did with computing the times for the critical reading section, we took the total time for this subsection, including the time spent reading the passage and the time spent answering questions, and divided by the number of questions in the subsection (six) to get the mean time per item. The histogram of these mean times is presented in Figure 26. In terms of both mean and distribution, the time for these items closely resembles the time taken for the critical reading items. Most examinees were within

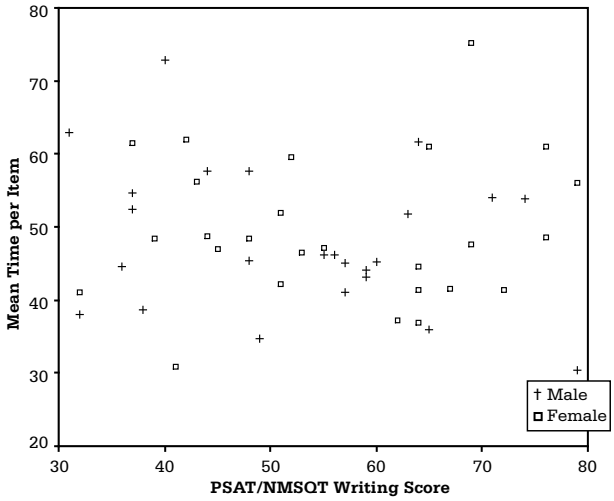

Figure 23. Scatterplot of the mean time taken for improving sentences items and PSAT/NMSQT writing scores by gender.

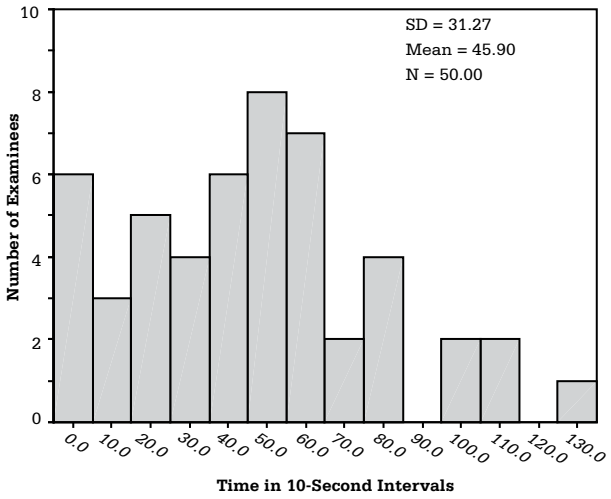

Figure 25. Histogram of time taken for reading a passage associated with improving paragraphs items.

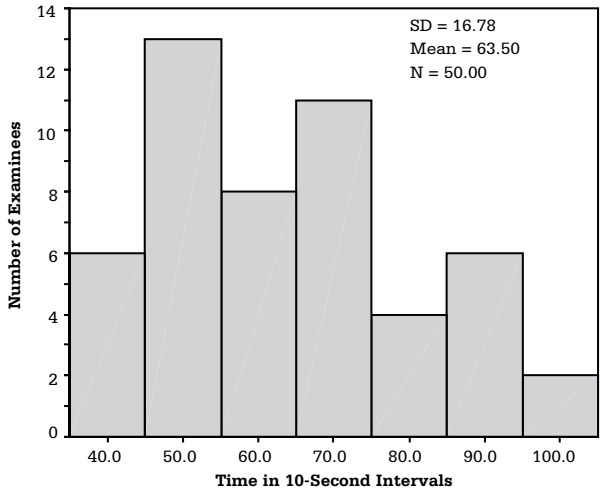

Figure 26. Histogram of the mean time taken for improving paragraphs items.

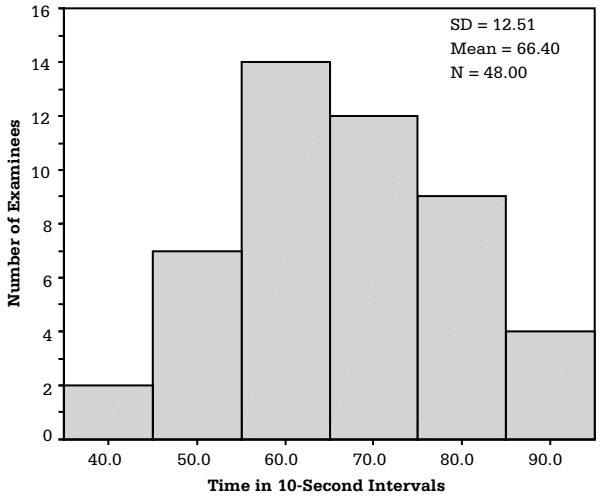

Figure 28. Histogram of the mean time taken for multiple-choice items.

the test developer estimate of 90 seconds per item for this item type.

As indicated in Figure 27, the mean time for improving paragraphs items was apparently unrelated to PSAT/NMSQT writing score. The correlation was a nonsignificant -.12. However, when the form in which this was the last subsection is excluded, the correlation jumps to a statistically significant -.48, which is almost identical to the correlation of PSAT/NMSQT verbal score and time taken on paragraph reading passages.

Mathematics

Multiple Choice

The mean time to complete the multiple-choice items was about the same whether they appeared at the end of the test (67 seconds) or at the beginning (69 seconds). The histogram of the time taken for the multiple-choice questions is presented in Figure 28. The mean time was considerably less than the SAT CAT mean time for this item type of 92 seconds, and was slightly less than the rule-of-thumb estimate of 72 seconds. However, over a quarter of the sample took longer than the rule-of-thumb estimate.

PSAT/NMSQT math scores correlated -.31 with the time scores (or -.44 for the form in which these items were first), suggesting that higher-ability students answer these questions more quickly than lower-ability students. The time-by-ability scatterplot is presented in Figure 29.

For the form in which the multiple-choice items were administered first, the mean times ranged from 42 to 103 seconds. Difficult items tended to take longer than easy items, but as seen in Figure 30, there were exceptions. The

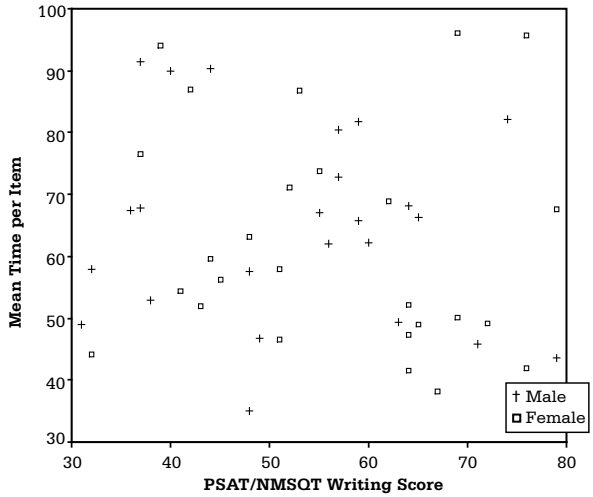

Figure 27. Scatterplot of the mean time taken for improving paragraphs items and PSAT/NMSQT writing scores by gender.

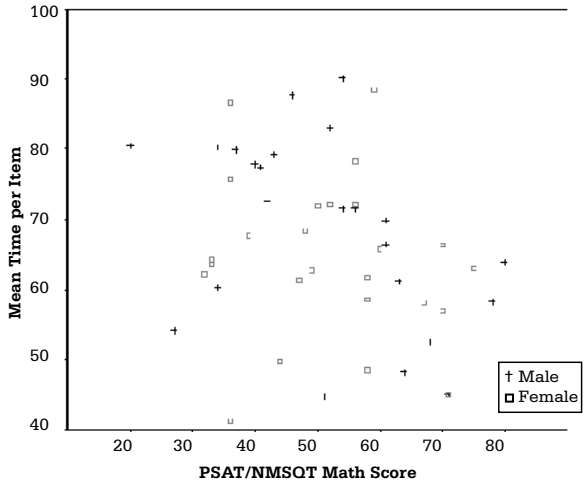

Figure 29. Scatterplot of the mean time taken for multiple-choice items and PSAT/NMSQT math scores by gender.

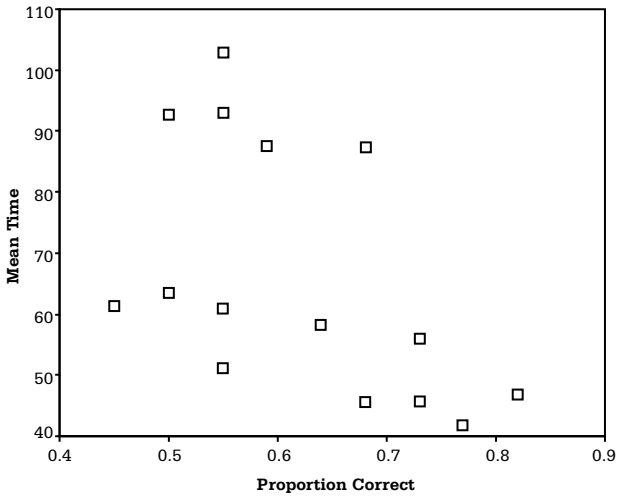

Figure 30. Scatterplot of the mean time taken for each multiple-choice item and the proportion correct for that item.

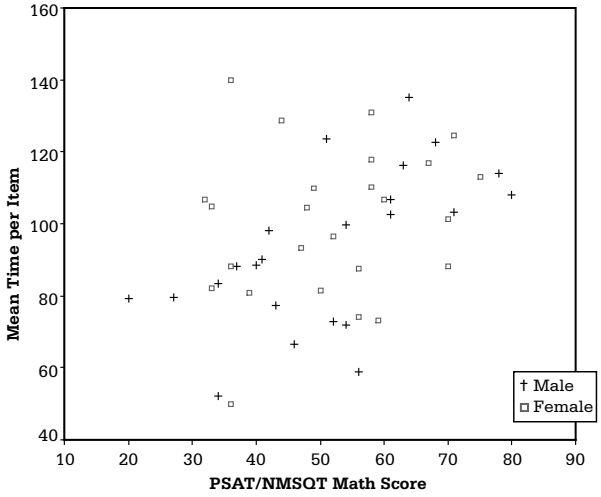

Figure 32. Scatterplot of the mean time taken for SPR items and PSAT/NMSQT math scores by gender.

item that took the longest was of only moderate difficulty (p = .55), and the average time for the most difficult item (p = .45) was only 62 seconds.

Student-Produced Response (SPR)

The mean time per item was only three seconds longer when the SPRs were the first in the section than when they were last; this difference was not statistically significant (F < 1). Figure 31 presents the histogram of mean times for both orders combined. The mean time was slightly longer than the test-developer rule of thumb of 90 seconds but was shorter than the time taken in the SAT CAT (118 seconds). Twenty percent of the sample took longer than 115 seconds, suggesting that more time may need to be allowed for SPRs.

The correlation of the mean time with PSAT/NMSQT math scores was .41, (though this correlation dropped to .23 for the form in which SPR items were first), suggesting that higher-ability students tended to take longer to answer these questions. Note that this is the exact opposite of the pattern for the multiple-choice items. The scatterplot is in Figure 32. A possible explanation of this unusual positive correlation is that lower-ability examinees may give up more quickly, leading to relatively short times. To evaluate this possibility, we correlated the mean time to a *correct* solution with PSAT/NMSQT math scores. This correlation was essentially zero (.01).

The mean times for the SPR items (using only the form in which SPR items were first) were quite variable, ranging from a low of 59 seconds to a high of 132 seconds for the last item. The mean times for 6 of the 10 items exceeded 100 seconds. The relationship of the mean item time to the proportion correct can be seen in Figure 33. There was a strong tendency for the harder items to

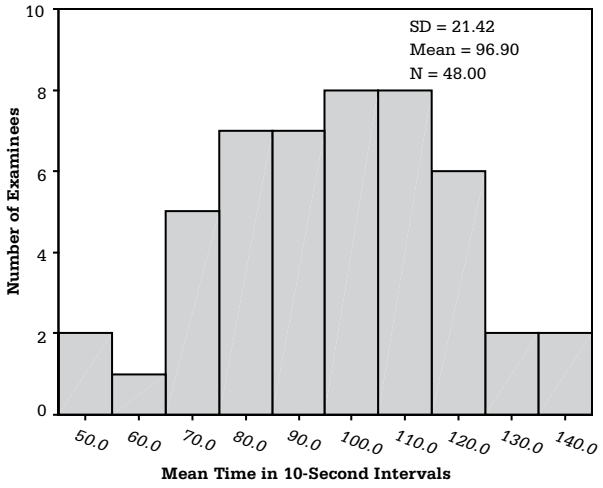

Figure 31. Histogram of the mean time taken for SPR items.

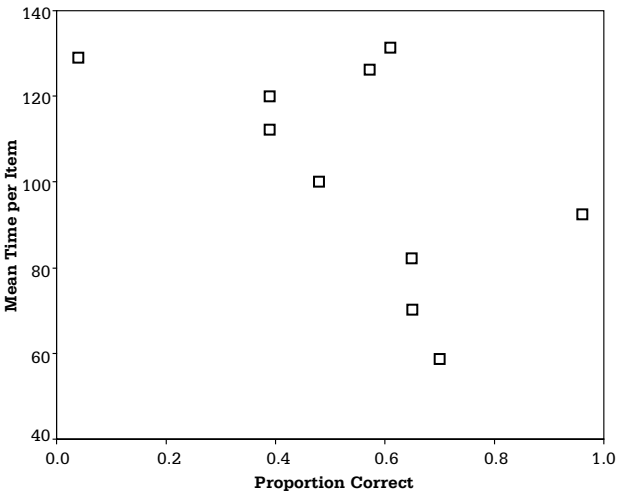

Figure 33. Scatterplot of the mean time taken for each SPR item and the proportion correct for that item.

take longer. The correlation of the mean time with the difficulty (proportion correct) was -.56. But, as noted above, mean times can understate true solution times if substantial numbers of examinees give up before reaching an answer. Across both forms, only five students got the correct answer for the last question; one student got the answer in 172 seconds, and the other four each took more than 300 seconds.

Study 3—Item-Type Times from a Group Administration

For this study, the College Board helped us recruit eight high schools that were willing to administer critical reading, writing, and/or mathematics sections during regular school hours. These sections were treated separately so that a given student might take one, two, or all three sections, depending on the way an individual school was able to arrange the testing. The high schools were selected largely from high-minority neighborhoods so that there would be an adequate representation of African American, Hispanic, and white students in the final sample. About 70 percent of these examinees had previously taken the PSAT/NMSQT. When comparing the PSAT/NMSQT verbal scores of these students to the published means for all college-bound seniors (College Board, 2003), and dropping the third digit to make the scores comparable, the African American sample is somewhat above average (mean = 48 [SD = 7.7]) compared to the national mean of 43 for African American students. (This slightly underestimates the difference between study samples and national samples because the national means are based on seniors, but the students in the study sample were juniors. On the other hand, the students who did not take the PSAT/NMSQT were probably of lower ability than those who did.) The Hispanic sample (mean = 42 [SD = 8.0]) was slightly below the national average of 45 for this group. Similarly, the white sample was slightly below average (mean = 50 [SD = 9.2]) compared to 53. For the PSAT/NMSQT math scores, the African American students in the sample were slightly above average, the Hispanic students were slightly below average, and the white students were comparable to the white students in the national sample. (In the study sample, means [and SDs] for African American students, Hispanic students, and white students, respectively, were: 46 [8.7], 43 [8.7], and 53 [7.8]; national means were 43, 46, and 53.)

The items in the test forms were identical to those in the observational study, and the same scrambling procedures were used so that an item type that was in the last position on one person's test would be in the first position on another person's test. The major difference from the observational study was that individual items could not be timed. Instead, examinees were asked to record the starting and stopping time for each subsection (to the nearest minute). Because each subsection contained a single item type, we could estimate the time per question of a particular type by computing the total time taken on the subsection and dividing by the number of items in the subsection. If a student ran out of time on the last subsection, there was no satisfactory way to estimate item times, so we never used the times from the last subsection. (Recall that the subsection that was last in some forms was first in other forms.) Although this procedure had the unfortunate effect of substantially reducing the available sample sizes, small interpretable samples seemed preferable to larger uninterpretable ones. For sections that contained more than two subsections, the first subsection in one form would be last in another, but the middle subsections would never be last and therefore the usable sample sizes were larger for these middle subsections.

The critical reading and mathematics section time limits were set to 35 minutes, which was the same amount of time allowed in the observational study. The writing section time limit was set to 30 minutes, which was 5 minutes shorter than in the observational study. This was done because most students in the observational study finished easily in 35 minutes, and we wanted to know how quickly students could answer these questions when they were experiencing more severe time pressure.

In anticipation of a possible change in scoring directions from formula scoring to rights-only scoring, we administered half of the forms under each set of directions. We hypothesized that items might take longer under formula-scoring directions because an examinee required time not only to figure out the answer, but also to decide whether to omit a particular question. A random half of the forms included the following standard formula-scoring directions:

"Scoring is the same as on a regular SAT.

- For each correct answer, you receive one point.
- For questions you omit, you receive no points.
- For a wrong answer to a multiple-choice question, you lose a fraction of a point.
- For a wrong answer to a math question that is not multiple choice, you don't lose any points.
    - On a multiple-choice question, if you can eliminate one or more of the answer choices as wrong you increase your chances of choosing the correct answer and earning one point.
    - If you can't eliminate any choice, move on. You can return to the question later if there is time."

The other random half had rights-only scoring directions:

"In a regular SAT, you get no points for a question you omit and you lose a fraction of a point for a wrong answer. THIS TEST DOES NOT HAVE A GUESSING PENALTY. ANSWER EVERY QUESTION EVEN IF YOU HAVE TO GUESS."

Analyses

We classified examinees into four racial/ethnic groups based on their self-reports: African American, Hispanic, white, and other (all others and nonresponders combined). With the mean time for an item type as the dependent variable type, we ran 4 (racial/ethnic groups) x 2 (genders) x 2 (scoring directions; standard versus rights-only) Analysis of Variances (ANOVAs). For the subsample of examinees for which we could locate PSAT/NMSQT (P/N) scores, we included the relevant score as a covariate (e.g., P/N-V for analysis of reading-item-type times; P/N-W for analysis of writing-item-type times), and we tested whether the covariate had a significant impact on the dependent variable.

Results and Discussion

Critical Reading

Sentence Completion

None of the variables in the ANOVA were significant. In the Analysis of Covariance (ANCOVA), the only significant variable was P/N-V (F [1,42] = 71.6, $p < .001$). For the 58 students with P/N-V scores, the correlation of P/N-V with mean time on the sentence completion items was -0.51. This item type was also negatively correlated with P/N-V scores in Study 2 (r = -0.48), suggesting that there is a relatively strong and consistent tendency for higher-ability students to answer these items more quickly than lower-ability students. Mean times by racial/ethnic group are presented in Table 1.

The mean time over all groups of 44.0 seconds per item is comparable to the 40.2 seconds per item found in the observational study. These items appear to be relatively time-efficient.

Table 1

Means and Standard Deviations of Item-Type Time (in Seconds) for Sentence Completion Items

Racial/Ethnic Group	n	*M*	*SD*
African American	32	44.6	14.6
Hispanic	31	41.5	11.9
White	23	44.4	10.2
Other	14	46.4	15.4

Table 2

Means and Standard Deviations of Item-Type Time (in Seconds) for Paragraph Reading Passage Items

Racial/Ethnic Group	n	*M*	*SD*
African American	45	63.2	19.3
Hispanic	56	63.9	13.8
White	34	60.5	16.2
Other	32	61.7	18.6

Paragraph Reading

Because these items never occurred as the last subsection, the sample size was larger for these items than for the sentence completion items (n = 167). None of the ANOVA or ANCOVA variables were significant, though there was a tendency for examinees with high P/N-V scores to answer more quickly (F [1, 68] = 3.38, p = .07). The overall mean time was 62.6 seconds per item (compared to 56.0 seconds in the observational study). Means by racial/ethnic group are in Table 2.

650-Word Passages

As in the observational study, the mean time for these items includes the time to read the passage (i.e., the total time on subsection divided by nine because there were nine items for each passage). Mean times for the first and second 650-word passages were 66.2 and 57.1 seconds, respectively (compared to 65.3 and 65.6 in Study 2); combined, the mean time for items on both passages was 62.6 seconds. Means by racial/ethnic group for the combined passages are in Table 3. There were no significant effects in the ANOVA or ANCOVA. Time to read the passages and answer the questions was not related to P/N-V scores (r = .07).

Writing

Identifying Sentence Errors

This subsection was taken by 121 examinees. The only significant effect in the ANOVA was race/ethnicity (F [3, 105] = 4.50, p = .005). A follow-up Tukey's HSD test indicated Hispanic students took significantly longer on this question type than did white and African American students. Table 4 shows the means and

Table 3

Means and Standard Deviations of Item-Type Time (in Seconds) for Combined 650-Word Passage Items

Racial/Ethnic Group	n	*M*	*SD*
African American	13	66.9	13.0
Hispanic	25	62.5	12.0
White	11	62.7	11.5
Other	11	57.3	18.7

Table 4

Means and Standard Deviations of Item-Type Time (in Seconds) for Identifying Sentence Error Items

Racial/Ethnic Group	n	*M*	*SD*
African American	33	30.6	10.5
Hispanic	34	40.5	11.3
White	34	31.1	13.6
Other	20	34.4	11.3

standard deviations for each racial/ethnic group. In the observational study, the mean time for these items was 38.8 seconds. In the reduced sample with the P/N-W covariate, the racial/ethnic effect remained statistically significant (F [1,53] = 2.98, p = .04).

Improving Sentences

Because this subsection was never in the last position, sample sizes were somewhat larger for this item type. As shown in Table 5, white examinees answered these questions somewhat faster than examinees in the other groups. This apparent difference was confirmed in the ANOVA (F [3, 218] = 3.96, p = .009) in which racial/ethnic group was the only significant variable. Tukey's HSD indicated that the white group differed from each of the other groups, which did not differ among themselves. The mean time for this item type in the observational study was 48.8 seconds. In the reduced sample of 121 examinees with the P/N-W covariate, the racial/ethnic difference was no longer statistically significant (F [4, 104] = 2.31, p = .08); the P/N-W score itself was unrelated to item time (F < 1; r = -.06).

Improving Paragraphs

Again, the only significant variable in the ANOVA was racial/ethnic group (F [3, 98] = 4.16, p = .008). Tukey's HSD indicated that the white group was faster than the Hispanic and other groups. Mean times for this subsection are in Table 6. There were no significant differences in the sample of 51 students with P/N-W scores. The times for the groups in this study bracketed the time from the observational study in which the mean time for improving paragraphs items was 63.5 seconds.

Table 6

Means and Standard Deviations of Item-Type Time (in Seconds) for Improving Paragraphs Items

Racial/Ethnic Group	n	*M*	*SD*
African American	42	67.4	21.3
Hispanic	35	82.6	20.1
White	14	59.3	20.6
Other	22	80.9	28.4

Mathematics

Multiple Choice

As suggested in Table 7, there were no significant racial/ethnic differences among item times for the five-choice mathematics items (nor were there significant differences on any other variable). Similarly, there were no significant differences on any variable in the sample of 157 students with P/N-M scores. Times were somewhat longer than the 66.4 seconds noted in the observational study.

Student-Produced Response (SPR)

As with the multiple-choice items, no significant differences were noted for any variable in either the full sample (F [3, 156]= 2.03, p = .11) or the sample with P/N-M scores (F [3, 110] = 1.87, p = .14). Mean times are shown in Table 8. Although not significant in these small samples with large within-group standard deviations, the pattern of the means suggests that future research could explore the possibility that these items are especially time-consuming for African American and Hispanic examinees. The mean time of 109 seconds over all groups is somewhat longer than the 96.9 seconds noted in the observational study, but is still shorter than the 118 seconds noted in Study 1 with the SAT CAT.

Scoring Directions

Item times were not significantly related to scoring directions for any of the item types in any of the three content areas (critical reading, writing, and mathematics). Contrary to expectations, times were typically a few seconds longer with rights-only directions.

Table 5

Means and Standard Deviations of Item-Type Time (in Seconds) for Improving Sentences Items

Racial/Ethnic Group	n	*M*	*SD*
African American	75	53.4	17.4
Hispanic	69	53.3	14.1
White	48	45.2	14.3
Other	42	55.0	15.0

Table 7

Means and Standard Deviations of Item-Type Time (in Seconds) for Multiple-Choice Math Items

Racial/Ethnic Group	n	*M*	*SD*
African American	43	81.3	16.8
Hispanic	29	76.6	16.3
White	84	74.2	21.2
Other	50	72.6	22.9

Table 8

Means and Standard Deviations of Item-Type Time (in Seconds) for SPR Math Items

Racial/Ethnic Group	n	*M*	*SD*
African American	39	114.8	32.5
Hispanic	34	118.4	26.6
White	69	100.5	28.9
Other	30	109.0	41.8

Summary and Conclusions

A summary of item times across all three studies is presented in Table 9. Times were uniformly longest in Study 1, which had the most generous time limits. This makes it clear that the question of how long a particular item type takes really has no definitive answer. In general, if students are granted a little more time they will take a little more time. Thus, Studies 2 and 3 provide a better picture of how long students take for each item type when they are feeling substantial time pressure.

There are two major sources of instability in these mean-time estimates. First, although efforts were made to ensure that samples were diverse with respect to racial/ethnic group membership and abilities measured by the PSAT/NMSQT, they are certainly not national probability samples and so can only crudely estimate means in the population. Second, as noted in Study 2, mean times for items of the same item type can vary, so a different mix of items would produce somewhat different estimates of mean time for a particular item type. Given these constraints, there is a reasonable consistency among the three studies and between the studies and the test developers' rule-of-thumb estimates. In particular, there is total agreement on the relative ranking of the times by item type; within the critical reading section, sentence completion items are substantially faster than passage-based items; within the writing section, sentence errors are fastest, followed by improving sentences, with improving paragraphs taking noticeably longer than the other item types; within the mathematics section, SPR items take considerably more time than standard multiple-choice items.

Assuming that 80 percent of the examinees should be able to answer an item within the rule-of-thumb time, it appears that the rules of thumb may be too short for all except the improving paragraphs items, and that they are particularly too short for SPR items. Time demands for the new paragraph reading items (with two items per passage) appear to be roughly comparable to time demands for the 650-word passages (with nine items per passage); in both Studies 2 and 3, 80th percentile times were within four seconds of each other for these two passage types.

Study 2 showed that many students are wasting testing time reading directions that they should have studied before the test began. Test-preparation materials may need to put even more emphasis on the importance of being very familiar with the test directions before the day of the test.

Results from Study 3 suggested that writing items may be especially time-consuming for Hispanic examinees. Although this is not surprising, further

Table 9

Mean and 80th Percentile Item-Type Times Across Three Studies

	Rule-of-Thumb Time	*Study 1 SAT CAT*		*Study 2 Lab Observation*		*Study 3 High School*	
		Mean	*80th Percentile*	*Mean*	*80th Percentile*	*Mean*	*80th Percentile*
Critical Reading							
Sentence completion	42	48	59	40	46	44	53
100-word passage	None	NA	NA	56	68	63	78
650-word passage	60-72*	97	123	65	72	63	80
Writing							
Sentence errors	30	NA	NA	39	50	34	44
Improving sentences	42	NA	NA	49	58	52	64
Improving paragraphs	90**	NA	NA	63	82	74	90
Mathematics							
Multiple choice	72	92	116	66	78	76	92
SPR	90	118	148	97	117	109	132

* Per-item estimate includes time to read passage and answer questions.
** Assumes three minutes to read passage, then one minute for each of six items, or an average of 90 seconds per item, including paragraph reading time.

Table 10

Correlation of Related PSAT/NMSQT Score with Item-Type Time Across Two Studies

	Study 2	*Study 3*
Critical Reading		
Sentence completion	-.45*	-.51*
100-word passage	-.49*	-.24*
650-word passage	-.05	.07
Writing		
Sentence errors	.27	-.24*
Improving sentences	-.06	-.17
Improving paragraphs	-.48*	-.30*
Mathematics		
Multiple choice	-.44*	-.09
SPR	.23	-.08

*$p < .05$.

Note: Sample sizes for Study 2 ranged from 24 to 50 (for subsections that were never in the last position). In Study 3, minimal sample sizes were 27 in critical reading, 51 in writing, and 127 in mathematics.

follow-up is recommended. As a start, completion rates for Hispanics should be compared to completion rates for white students on the P/N-W. Final test specifications should provide ample time for all subgroups to complete the test.

Table 10 presents the correlations of mean item times for individuals with their PSAT/NMSQT scores from the related section (e.g., P/N-V with critical reading items and P/N-W with writing items). We excluded data from the last subsection of each test. The table suggests that higher-ability students tended to answer more quickly, though there were exceptions. For SPRs, the correlation was a low negative in one study and a low positive in the other, suggesting that even higher-ability math students may take time to recheck answers on SPRs.

Further research is needed to confirm the patterns noted here. More importantly, future research should focus on the reasons for the considerable variability noted within item type. Although some of this variability is related to item difficulty, some is not. Items of the same type and difficulty level can still vary substantially in the amount of time needed to answer them. If appropriate time limits are to be set, some attention to the time demands of different item types is necessary, but it is not sufficient without also understanding the factors affecting variations within item type. Such understanding would help to ensure that different forms had comparable time demands.

References

College Board. (2003). *College-Bound Seniors 2002*. Retrieved March 20, 2003 from http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2002/graph10.pdf.

Lawrence, I., & Feigenbaum, M. (1997). *Linking scores for computer-adaptive and paper-and-pencil administrations of the SAT*. (Research Report No. 97-12). Princeton, NJ: Educational Testing Service.

Appendix: Examples of Different Item Types

Critical Reading: Sentence Completion Item Type

Despite the wide-ranging curiosity about her personal life, Eleanor Roosevelt enjoyed a degree of _____ that today's highly scrutinized public figures can only _____.

(A) privacy . . envy
(B) popularity . . celebrate
(C) privilege . . imitate
(D) isolation . . regret
(E) generosity . . refuse

Critical Reading: Paragraph Reading Item Types

Questions 10 and 11 are based on the following passage.

For generations, archaeologists have dug up the tombs and treasures of the pharaohs all over Egypt. Archaeologist Mark Lehrer has focused, instead, on where and how the
Line thousands of laborers who actually built the pyramids lived,
5 attempting to decipher the complex economic system that sustained them over the 80 or so years they labored on this monumental task. Lehrer believes that more than 20,000 workers lived in a "lost" city on the Giza plain in Egypt.

10. The first sentence of the passage ("For generations … Egypt") is primarily intended to

(A) capture the reader's attention with a controversial statement
(B) establish a context by citing a prevailing practice
(C) introduce a theory that is later rejected
(D) express anger about the conventional practices of a profession
(E) point out the distinctive features of archaeological investigations

11. In line 8, "lost" most nearly means

(A) wasted
(B) unsuccessful
(C) vanished
(D) beyond reach
(E) no longer owned

Critical Reading: Reading Comprehension Item Type (Paragraph on this page, and items on next page.)

Questions 20–28 are based on the following passage.

In this passage, the narrator considers his family's history and migration from Mexico to Texas, which was once part of Mexico.

I never understood people's fascination with immortality. The idea of life without end gave me chills. Even as a kid, I wanted to be among my family and my ancestors,
Line walking through our short time together. I wanted to bind
5 Texas and Mexico together like a raft strong enough to float out onto the ocean of time, with our past trailing in the wake behind us like a comet tail of memories.

But the past can be difficult to conjure again when so little has been left behind. Some families in Mexico have troves
10 of their ancestors' belongings, from pottery of the ancients and paintings of Mexico City in the eighteenth century to helmets and shields of the Spaniards. By comparison my family, the Santos, are traveling light through time. Virtually nothing has been handed down, not because
15 there was nothing to give, but after leaving Mexico to come to Texas—so many loved ones left behind, cherished places and things abandoned—they ceased to regard anything as a keepsake. Everything was given away. Or they may have secretly clung so closely to treasured objects that
20 they never passed them on. Then these objects were lost.

My uncle Lico ferreted out the past as a passionate genealogist who used research, fantasy, and spells of breathless madness to craft his ancestral charts of the branches of our family. Some are elaborate discs, in which
25 each outward concentric ring represents a new generation. In others, quickly dashed off as notes to himself, ragged trees and jagged lines are drawn between names like Evaristo, Viviano, Blas, and Hermenegilda. In one, going back to 1763, the capstone slot contains the cryptic entry
30 "King of Spain," from whom, presumably, he believed we were descended. Subtle faculties and proclivities were passed, speechlessly, through the flesh of successive generations. The ghosts of Spanish royalty mingled with Indians, Black people, and others from every part of the
35 world in Uncle Lico's secret genealogy. Yet, despite the ridicule of many, he managed to recover numerous names and stories. Lico knew I had some of the same magnetic attraction to the past that fueled his manic genealogies, as if the molecules of our bodies were polarized in a way
40 that drew us both back in time, back, inexorably, toward the ancestors.

In my dreams, the ancestors who have passed on visit with me in this world. They ask me questions they were once asked: Where did our forbears come from and what have we amounted to in this world? Where have we come to in the span of time, and where are we headed, like an arrow shot long ago into an infinite empty space? What messages and markings of the ancient past do we carry in these handed-down bodies we live in today?

With these questions swirling inside me, I have rediscovered some stories of the family past in the landscapes of Texas and Mexico, in the timeless language of stone, river, wind, and trees. My great-uncle Abrán was a master of making charcoal. He lived in the Texas hill country, where the cedars needed to make charcoal were planted a century ago. Today, long after he worked there, walking in that central Texas landscape crowded with deep cedar, I feel old Abrán's presence, like the whisper of a tale still waiting to be told, wondering whether my intuition and the family's history are implicitly intertwined. Even if everything else had been lost—photographs, stories, rumors, and suspicions—if nothing at all from the past remained for us, the land remains, as the original book of the family. It was always meant to be handed down.

20. The image of the "raft" (line 5) most clearly conveys the narrator's childhood

(A) wish to escape his circumstances

(B) desire to merge his family's Texan and Mexican identities

(C) consideration of leaving Texas and returning to Mexico

(D) belief that Texas and Mexico are more similar than not

(E) awareness that he is neither a Texan nor a Mexican

21. The objects mentioned in lines 10–12 ("from pottery . . . Spaniards") are examples of

(A) artifacts discovered by Uncle Lico

(B) possessions viewed as impediments to a simple life

(C) gifts bestowed on departing loved ones

(D) necessities valued by earlier generations

(E) items bearing both cultural and personal meaning

22. In line 13, "light" most nearly means

(A) unencumbered

(B) illuminated

(C) nimbly

(D) faintly

(E) gently

23. The primary effect of lines 21–35 ("My uncle … genealogy") is to depict the

(A) collaboration between the narrator and his uncle

(B) influence of the uncle on the narrator's generation

(C) unorthodox nature of Uncle Lico's methodology

(D) family's enthusiasm for Uncle Lico's research

(E) rigors of conducting genealogical investigations

24. The scientific language used in lines 37–41 ("Lico … ancestors") emphasizes the

(A) forcefulness of a shared fascination

(B) chaotic methods used by the narrator's uncle

(C) distillation of information about the narrator's past

(D) place of family systems in the natural world

(E) intersection of two separate family lines

25. The narrator indicates that the questions his ancestors pose (lines 43–49) are ones that

(A) he cannot possibly answer truthfully
(B) are meant to forewarn as well as confuse
(C) are not really intended to elicit a response
(D) contain the answers hidden within themselves
(E) have been asked before and will be asked again

26. The characterization of the "bodies" in line 49 underscores the narrator's preoccupation with

(A) genealogical method
(B) personal destiny
(C) family harmony
(D) familial identity
(E) genetic variability

27. The last paragraph suggests that the narrator has discovered

(A) a collection of cedar mementos left by his great-uncle
(B) a way to remain in touch with his family's past without keepsakes
(C) an area in Texas that reminds him of the home he had left
(D) stories that supply direct answers to the questions in his dreams
(E) a method of using the land as a valuable source of income

28. The overall tone of the passage is best described as

(A) analytical
(B) whimsical
(C) dramatic
(D) reflective
(E) speculative

Mathematics: Multiple-Choice Item Types

Which of the following fractions is NOT between $\frac{1}{4}$ and $\frac{3}{4}$?

(A) $\frac{1}{2}$
(B) $\frac{1}{3}$
(C) $\frac{1}{12}$
(D) $\frac{2}{3}$
(E) $\frac{9}{16}$

If a class consists of b boys and g girls and there are twice as many girls as boys, which of the following represents the total number of boys and girls in the class?

I. $b + g$
II. $3b$
III. $3g$

(A) I only
(B) II only
(C) III only
(D) I and II only
(E) I, II, and III

Mathematics: Constructed-Response Item Type

Katie is selling cookies for $3 per box. For each box that Katie sells she earns 10 percent of the selling price. How many boxes must she sell to earn $60 ?

Writing: Identifying Sentence Errors Item Type

The bus driver was <u>not hardly</u> (A) in the mood to wait for passengers; he <u>drove off</u> (B) <u>while</u> (C) Mr. Jeffers was in the store <u>buying</u> (B) souvenirs. <u>No error</u> (C)

Writing: Improving Sentences Item Type

Under a cold, bright blue winter sky, the fantastic city of Florence enchanted the young couple <u>like on the first time they visited it</u>.

(A) like on the first time they visited it

(B) like it did the first time of their visit

(C) as when they were enchanted the first time

(D) like the enchantment when first they visited it

(E) as it had on their first visit

Writing: Improving Paragraphs Item Type

(1) *A good sense of humor is essential.* (2) *Not many people have the ability to transform something grave into something laughable.* (3) *Taking even a mundane occurrence and to make people view it in an offbeat way.* (4) *That is a gift that most people underestimate.* (5) *Also, sometimes a stubborn friend or an inflexible coach is the object of humor.* (6) *Good-humored folk can be late to practice and still beam, "Oh well. How important can practice be?"* (7) *People with a sense of humor are a lot freer than most from societal pressures and expectations.* (8) *They realize that, from a broader perspective, being late or not being dressed perfectly does not really matter.* (9) *They also realize that, in the larger scheme of things, our disappointments are trivial.* (10) *My brother is like this.* (11) *He missed his high school graduation because his car broke down, but he just looked at it all as a big joke.* (12) *I think I would have cried, but he laughed!*
(13) *It is true that you cannot laugh at everything.* (14) *Most people would agree that some things should not be mocked.* (15) *Society cannot function smoothly if they laughed at everything.* (16) *And who likes to be made fun of?* (17) *Yet if we can laugh at ourselves and take things less seriously, our world may become a happier place.*

Sentences 3 and 4 (reproduced below) could best be written in which of the following ways?

Taking even a mundane occurrence and to make people view it in an offbeat way. That is a gift that most people underestimate.

(A) (As they are now)

(B) An underestimated gift is how to make people view even a mundane occurrence in an offbeat way.

(C) Indeed, taking a mundane occurrence and making them view it in an offbeat way is an underestimated gift.

(D) The ability to make people view even a mundane occurrence in an offbeat way is a gift that many underestimate.

(E) Being able to take this and make people view it in an offbeat way is a gift underestimated by most people.


CollegeBoard
connect to college success™