

Research Summary

Office of Research and Development

RS-11, January 2005

The Research Behind the New SAT®

Jennifer L. Kobrin and Amy Elizabeth Schmidt

Introduction

The College Board announced in June 2002 that a new SAT® would be introduced in March 2005, intended for the class of 2006. The content of the new test was inspired by the 1990 blue-ribbon panel report, *Beyond Prediction*, and the development of the actual test specifications has been informed by expert advice from test development committees in reading, writing, and math and by extensive research studies. The College Board is committed to ensuring that the changes to the SAT will have a positive impact on students and meet the changing needs of the education community. Every step along the way in the development of the new SAT has been informed by research, and research will continue well after the new test becomes operational. Here is a brief summary of the projects that have been conducted to support the development of the new SAT.

Research on the new SAT can be classified into three major areas: *pre-field-trial studies*, which were conducted to inform development of the types of items to appear on the test and the overall design of the test; a *large-scale field trial*, which was conducted in March 2003 to evaluate a prototype of the new SAT; and *post-field-trial studies*, which were conducted to follow up on questions that arose during and/or after the field trial, and to examine the validity of a prototype version of the new SAT for predicting college performance. All of the studies described in this summary are or will soon be available on www.collegeboard.com/research.

Pre-Field-Trial Studies

Curriculum Surveys

Early on in the development of the new SAT, more than 2,000 high school and college teachers completed curriculum surveys to indicate what reading and writing skills were important for students entering their first year of college. The results showed that the skills rated as important by the survey respondents were in line with the content covered by the critical reading and writing sections of the new SAT. For more information, see *A survey to evaluate the alignment of the new SAT writing and critical reading sections to curricula and instructional practice* (Milewski, Johnsen, Glazer, and Kubota, 2005).

Simulation of Item Performance

One of the most salient changes to the critical reading section of the new SAT is the elimination of analogy items. Before considering the removal of these items, it was necessary to ensure that the reliability or measurement precision of the test could be maintained without these items. Using actual SAT data, researchers simulated verbal test forms without the analogy items. The results indicated that it was indeed possible to maintain the high reliability of the verbal test without the analogy items, but that it would be necessary to modify the distribution of item difficulty in order to obtain adequate precision at the top and bottom of the score scale. In other words, a larger number of very easy and very difficult items in the other verbal item types (sentence completion and reading comprehension) would be needed. For more information see *A simulation study to explore configuring the new SAT critical reading section without analogy items* (Liu, Feigenbaum, and Cook, 2004).

Investigation of Item Timing

Since the new SAT would have different types of questions from the current test, it was also necessary to determine the amount of time, on average, that students would need to answer each question so that the number of questions and time limits of the test would be reasonable and fair. The College Board addressed these issues by commissioning a series of studies in which item timing data from a computerized version of the SAT, and observation of students as they took new SAT questions under timed conditions, were used to estimate the amount of time needed to answer each type of question. For more information, see *Time requirements for different item types for use in the revised SAT* (Bridgeman, Cahalan, and Cline, in press).

Effects of Enhanced Math Content

The new SAT math section will measure more advanced content than is currently assessed. Researchers explored whether the addition of the items with more advanced content would impact test-taker performance. The findings supported the notion that test-taker performance is not affected by the mere presence of these items, but that performance is based on the difficulty level of these items. For more information, see *Evaluating SAT II: Mathematics IC items in the SAT I population* (Liu, Schuppan, and Walker, in press).

Effects of Fatigue

With the addition of a writing section to the new SAT, the entire test will take 45 minutes longer than the current test. Before increasing the length of the test, the College Board wanted to ensure that the additional time needed to complete the test would not fatigue students and diminish their performance. A review of the literature on fatigue in testing suggested that student performance on reading and math tests is resistant to fatigue for periods of up to five to six hours. Students may feel fatigued, but such feelings are not likely to have a negative effect on test performance. The literature also suggested that more fatigue is associated with simple tasks than complex tasks, and low-stakes tasks compared to high-stakes tasks. These notions were supported by a small study with about 100 students. For more information, see *A study of fatigue effects from the new SAT* (Liu, Allspach, Feigenbaum, Oh, and Burton, 2004).

Type of Essay Prompt

Research was also needed to determine the type of essay prompt to include on the writing section. A study was con-

ducted to investigate the impact of the new type of essay prompt proposed for the new SAT on ethnic, language, and gender groups. The results of this study indicated that the essay prompt type that will be used on the new SAT did not disadvantage any particular group of students. For more information, see *New SAT writing prompt study: Analyses of group impact and reliability* (Breland, Kubota, Nickerson, Trapani, and Walker, 2004).

The Field Trial of the New SAT

In March 2003, an extensive field trial was conducted to better understand the proposed changes to the SAT. The purpose of the field trial was to evaluate the content, statistical, and timing specifications for the new SAT and PSAT/NMSQT®, as well as whether scores on the new test are comparable to (have the same meaning as) scores on the current test. More than 45,000 students from 679 high schools participated in the field trial. These students were from both public and private schools across rural, suburban, and urban areas, and represented every geographic region in the United States. Each student completed a full-length, or equivalent, new or current version of the SAT or PSAT/NMSQT. To ensure that the research to address racial/ethnic group differences was based on sufficient numbers of students, a higher proportion of African American and Hispanic/Latino students were included in the field trial sample.

The results showed that:

- the new critical reading and math sections did not impact the difficulty of the test, that is, the difficulties of the items were within the normal range;
- participants' scores on the field trial were in line with their previous PSAT/NMSQT scores;
- the new test was very similar in reliability to the current SAT;
- the correlations between the current and new tests were very high (between .95–.97) for all three sections, indicating that the critical reading and math scores on the new SAT can be equated to the verbal and math scores on the current SAT;
- overall, the content and format changes do not appear to exacerbate the score differences between gender and ethnic groups.

For more information, see *Prototype analysis of spring 2003 new SAT field trial* (Liu and Feigenbaum, 2003), and

Equitability analysis of the new SAT to the current SAT I (Liu, Feigenbaum, and Dorans, in press).

Post-Field-Trial Studies

Several studies took place after the field trial to evaluate different ways of refining the SAT before deciding on its final form. A separate study of about 3,000 students from the field trial suggested that the proposed time limits for the critical reading and math sections are appropriate and that extra time has virtually no effect on performance. For more information, see *Effects of extra time on performance on new SAT questions* (Bridgeman and Cline, in press).

The field trial results indicated the need for further research on the number of writing questions to be included and the placement of the essay for optimal performance by students. An additional field test of 6,000 students was conducted to evaluate the effects of varying time limits and number of questions on student performance (Allspach and Walker, 2004). The results led to the College Board's decision to add an additional 10-minute section of writing questions to improve the reliability of the writing score. In a separate study, the College Board learned that the majority of students preferred the essay first on the test, and they performed slightly better on the essay when it was placed first on the exam. Therefore, the essay will always appear first on the new SAT. For more information, see *The effects of essay placement and prompt type on performance on the new SAT* (Oh and Walker, 2003).

Predictive Validity of the New SAT

One of the goals of adding a writing section to the new SAT is to improve the validity of the test for predicting college success. The American Institutes for Research, in collaboration with the College Board, recently completed a study based on approximately 1,200 first-year students from 13 colleges to examine the predictive and placement validity of the new SAT writing section. The prototype that was evaluated was 10 minutes shorter and had 12 fewer questions than the operational new writing section will have. The results indicated that total scores on the writing section correlated about .46 with first-year college grades, and correlated about .32 with English composition grades.

While scores on the new SAT writing section were slightly better than high school grades in predicting first-year college grades (.46 versus .43), the SAT writing scores were

slightly worse than high school grades in predicting English Composition course grades (.32 versus .35). The addition of the writing section added .01–.02 to the validity of the current SAT and high school grades. These results suggest that the writing section will increase the predictive validity of the SAT, and will be valuable in course placement. For more information, see *The College Board SAT writing validation study* (Norris, Oppler, Kuang, Day, and Adams, 2004).

Future Research

When the new SAT becomes operational in March 2005, the College Board has an intensive research agenda planned to investigate all aspects of the new test and its impact on all student groups. Among the important studies planned are determining how to set the new writing section on scale, conducting a new baseline validity study, and monitoring the technical characteristics of the new test as it goes operational.

Office of Research and Development
The College Board
45 Columbus Avenue
New York, NY 10023-6992
212 713-8000

Copyright © 2005 by College Board. All rights reserved. College Board, SAT, and the acorn logo are registered trademarks of the College Board. Connect to college success is a trademark owned by the College Board. PSAT/NMSQT is a registered trademark of the College Board and National Merit Scholarship Corporation. Visit College Board on the Web: www.collegeboard.com.