CollegeBoard
connect to college success™

Research Report
No. 2006-5

A Historical View of Subgroup Performance Differences on the SAT Reasoning Test™

Jennifer L. Kobrin, Viji Sathy, and Emily J. Shaw

www.collegeboard.com

College Board Research Report No. 2006-5

A Historical View of Subgroup Performance Differences on the SAT Reasoning Test™

Jennifer L. Kobrin, Viji Sathy, and Emily J. Shaw

The College Board, New York, 2007

Jennifer L. Kobrin is a research scientist at the College Board.

Viji Sathy is an associate research scientist at the College Board.

Emily J. Shaw is an assistant research scientist at the College Board.

The research reported in this paper is collaborative in every respect, and the order of the authorship is alphabetical.

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

The College Board: Connecting Students to College Success

The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 5,000 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 #060481915) may be obtained from College Board Publications, Box 886, New York, NY 10101-0886, 800 323-7155. The price is $15. Please include $4 for postage and handling.

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

Printed in the United States of America.

Acknowledgments

We would like to thank Robert Linn, Roberto Noya, Amy Schmidt, and Brian O'Reilly for their helpful comments on this research report.

Contents

Figures

Abstract

This paper presents and reviews gender, racial/ethnic, language, and socioeconomic subgroup performance differences on the SAT® over nearly the last two decades. Theories on the existence of subgroup differences are examined. Substantial revisions to the SAT were made in 1994, and again in 2005. The short-term and long-term impact of these revisions on subgroup differences is evaluated. Furthermore, the trends in subgroup differences on the SAT are compared to those documented for other large-scale standardized tests (i.e., the ACT Assessment, National Assessment of Educational Progress), as well as those found in high school grades. This documentation of trends in subgroup differences on the SAT can be a useful resource for individuals interested in standardized tests as well as those studying how students' experiences might account for differences on the test.

Introduction

The SAT Reasoning Test™ (SAT) is one of the most widely used standardized tests for college admissions. The SAT has a long-standing history of being a rigorous measure of academic potential with well-documented reliability and validity for predicting college performance. However, the SAT also receives its share of criticisms along with other standardized tests. One of the most persistent criticisms of the SAT centers on the large differences in average performance for different gender, racial/ethnic, language, and socioeconomic groups. Although the average score gaps between many demographic groups have been narrowing over the years, large differences continue to occur (Young and Kobrin, 2001). To ensure that the SAT is equally reliable and valid for its intended uses for all subgroups, it is necessary to continually monitor subgroup performance.

Claims of test bias abound in the media and among persons both familiar and unfamiliar with the testing industry and/or psychometrics (e.g., FairTest, 2001; Freedle, 2003; Kohn, 2000; Rosner, 2003; Sacks, 1997). To many critics of standardized tests, mean group differences are sufficient evidence to conclude that a test is biased (Linn, 1990). Although persistent group mean differences are a cause for concern, it does not necessarily mean that the tests producing these differences are biased. "By themselves, test scores cannot reveal causality. Nor can the differences, by themselves, prove or disprove the existence of bias" (Linn, 1986).

A plethora of research has shown that the SAT is not biased against minority groups. Since the SAT is used to predict early college performance, if the SAT was biased, one would expect to find that minority students perform better in college, on average, than their scores predict. However, studies of the predictive validity of the SAT continually show that African American and Hispanic male students achieve somewhat lower freshman grades than their SAT scores predict, and African American and Hispanic females achieve on average about the grades that their test scores predict (Koretz, 2002). Furthermore, research has shown that subgroup performance differences are not limited to the SAT; these differences are found in virtually all measures of educational outcomes, including other large-scale standardized tests, high school grades and rank, high school graduation, and college attendance (Camara and Schmidt, 1999; National Center for Education Statistics, 2003; Zwick, 2002a, 2002b).

Purpose and Organization of This Report

The purpose of this report is to present and review gender, racial/ethnic, language, and socioeconomic subgroup performance differences on the SAT over the last 20 years (from 1987 to 2006), and to identify and discuss trends in these performance differences. Investigating subgroup differences on the SAT is in keeping with the College Board's commitment to excellence and equity in education, which are central aspects of the organization's mission. Although many other subgroups have been studied, including learning disabled students (e.g., Cahalan, Mandinach, and Camara, 2002) and younger test-takers (Wendler, Ninneman, and Feigenbaum, 2001), this report focuses on gender, race/ethnicity, first language, best language, and family income.

Each year, the College Board publishes the mean SAT scores for college-bound seniors by subgroup, so that subgroup performance differences can be assessed. However, there has not been one comprehensive source that has published subgroup means over time so that long-term trends in subgroup differences can be examined. This report fills that gap. Another important impetus for this report was the recent revision of the SAT in 2005. The revisions included the elimination of analogy items from the verbal section (SAT-V) and renaming of the section to critical reading; the elimination of quantitative comparison items from and the addition of items covering content from third-year college-preparatory math courses to the mathematics section (SAT-M); and the addition of a writing section. One of the most important assumptions underlying the revisions to the SAT was that subgroup performance differences would not be exacerbated. If any changes in subgroup differences were to occur, they would be in the direction of reducing the differences. As subgroup performance data on the new SAT are gathered, they

will be compared to historical subgroup data to ensure that the gaps that exist on the current test do not widen. This report provides the documentation necessary to make this comparison.

The first section of this report summarizes the literature documenting performance differences on the SAT and other educational assessments and provides possible explanations for these differences offered by other researchers. Next, trends in mean scores on the SAT and SAT Subject Test in Writing are presented for each of the major subgroups, as well as the trends in the demographic composition of test-takers over the last 20 years. This report includes trends for the SAT Subject Test in Writing because this test was the precursor for the new writing section on the SAT. Trend data on the other 19 SAT Subject Tests™ are not included in this report. Substantial revisions to the SAT were made in 1994, and as mentioned earlier, again in 2005. The short-term and long-term impact of the 1994 revisions on subgroup differences is assessed, providing implications for the impact of the 2005 changes to the SAT on subgroup differences. Then subgroup differences on the SAT and SAT Subject Test in Writing are compared with subgroup differences in high school grades (as self-reported by students taking the SAT), the ACT Assessment, and National Assessment of Educational Progress (NAEP). This report concludes with a summary of the findings and a discussion of the trends in subgroup differences with a view toward what may be expected in the future.

Literature Documenting Subgroup Performance Differences on the SAT Reasoning Test™ and Other Tests

Theories for Subgroup Differences

Scheuneman and Slaughter (1991) presented typical explanations for subgroup performance differences on tests in five broad, overlapping categories: historical, cultural, biological, educational, and psychometric. Historical explanations generally include descriptions of past practices, such as segregation and discrimination, which produced unequal access to facilities and resources (e.g., libraries, museums, higher educational institutions) to enhance an individual's knowledge base. Cultural explanations usually describe behaviors, language issues, and styles of learning that appear to have a relationship with performance on standardized tests. Explanations that center on socioeconomic status as a determinant of test score differences are also considered cultural in nature. Scheuneman and Slaughter also cite that biological explanations have been offered by some researchers to account for subgroup differences in test scores. These explanations attribute such performance differences to some innate feature of individuals or groups. Educational explanations include discussions of the number and quality of courses taken by specific groups, the amount of time on task, the quality of teachers and teaching given to specific groups, and student motivation. Finally, psychometric explanations for subgroup test differences point to the tests themselves as the source of the differences. Psychometric explanations include issues of differential test speededness or the timed conditions under which the test was administered, test center conditions or environment, and test-taking skills for different subgroups.

Wiesen (2005) provided a comprehensive list of more than 95 possible reasons for the mean score differences between African Americans and whites found on many cognitive ability tests. These reasons were categorized somewhat differently from Scheuneman and Slaughter (1991), in that Wiesen's primary categories included: physiological, economic and socioeconomic, psychological (not related to test construction), societal, cultural, and test construction/validation explanations. Physiological explanations included issues related to prenatal and postnatal conditions; economic and socioeconomic explanations included issues related to health care, criminal justice, education, finances, employment, and housing; psychological explanations included issues related to stereotype threat, motivation, and the effects of discrimination; societal and cultural explanations were similar to Scheuneman and Slaughter's cultural explanations; and test construction/validation explanations were similar to Scheuneman and Slaughter's psychometric explanations of subgroup performance differences.

It should be noted that several of the theories mentioned above have little supporting evidence, and most are plausible rather than proven. It is quite difficult to disentangle the effects of these different explanations. Wiesen (2005) noted that no one reason is likely to account for more than a small fraction of the typical one standard deviation difference in mean test scores between African American and white students, but that it is possible that together, these explanations may account for much or all of that difference.

Gender Differences

For many years since the SAT was first introduced, women tended to score higher on the verbal section and men scored higher on the mathematics section. However, starting in the early 1970s, women began to lose the advantage they once held on the SAT verbal section and made little progress in mathematics (Murphy, 1992). Performance differences between men and women have also been observed on many other tests used in higher education admissions. For example, Zwick (2002a) reported that, on average, men score higher than women on the ACT Mathematics Test, Science Test, and Composite; the Medical College Admissions Test (MCAT) Verbal Reasoning, Physical Sciences, and Biological Sciences tests; the Graduate Record Examinations (GRE) Verbal, Quantitative, and Analytical tests; the Graduate Management Admissions Test (GMAT) Verbal and Quantitative tests; and the Law School Admission Test (LSAT).

A variety of hypotheses have been offered as an explanation for standardized test score differences by gender. Some of these have included test bias, biological differences, differing interests and aspirations, societal influences, and stereotype threat (Zwick, 2002a). Wilder and Powell (1989) believe that gender differences in test performance yield no single explanation, but rather a variety of possible contributing factors including differences in spatial ability, sex-role development, social and educational phenomena such as treatment from teachers or parents, interests and attitudes, achievement motivation, and patterns of course-taking. Another potential contributor can be the test itself, or more specifically, items on the test that function differently for males and females. Most testing programs, including the SAT, perform routine analyses called differential item functioning (DIF) to identify test items that function differently for males and females and remove any such items. Therefore, this last explanation is not a likely cause of the gender differences found on the SAT.

Several studies were conducted in the late 1980s and early 1990s to uncover the reasons for performance differences between men and women on the SAT. Most of these studies showed that score differences are not related to test content, but are more a result of the changing population of students taking the SAT. Starting in the mid-1970s, women began to comprise the majority of SAT takers. While in 1972 there were 24,696 fewer women than men taking the SAT (48.8 percent of test-takers were female), by 2005, 103,027 *more* women than men took the SAT (53.5 percent of test-takers were female) (College Board, 1972, 2005). These "additional" women have been found to be less likely to have taken rigorous academic courses than other students, more likely to come from lower economic backgrounds, more likely to be the first in their families to apply to college, and more likely to be of African American, Native American, or Hispanic descent (Murphy, 1992; Young and Fisler, 2000).

Burton, Lewis, and Robertson (1988) similarly speculated that gender test score differences are the result of background differences between males and females who take the SAT based on their racial/ethnic group, income as a proxy for socioeducational status, presence or absence of certain basic high school courses, and proposed college major. They found that differences in these characteristics are significantly related to SAT score differences. For example, when women's verbal/critical reading scores were adjusted on the four aforementioned variables together, they were two points higher than men's scores on the SAT verbal/critical reading section. This suggests that if the women and men taking the SAT were more similar on these four characteristics, women would score equal to or higher than men on the SAT verbal/critical reading section. When this adjustment was done for SAT mathematics scores, women still scored lower than men, but on average they scored 17 points lower as opposed to 51 points lower.

Young and Fisler (2000) used analysis of covariance to adjust SAT scores for men and women taking the November 1990 administration of the SAT, based on father's and mother's education, parental income, race/ethnicity, number of years of high school English and math courses, honors courses in English and math, and Advanced Placement Program® (AP®) courses in English and math. The results indicated that all of the covariates were significantly related to SAT scores. Adjusting for these background variables actually led to an increase of about 5 points in the difference between men's and women's average SAT verbal scores, but produced a substantial decrease of about 11.5 points in the difference between men's and women's SAT mathematics scores. These findings corroborate the Burton et al. (1988) results in mathematics, but do not support the finding that matching females and males by background characteristics would decrease the score gap on the SAT verbal section. The conflicting findings may be due to the fact that the Young and Fisler study examined data from only one SAT administration (November 1990), while the Burton et al. study analyzed data for three full cohorts of college-bound seniors in 1975, 1980, and 1985.

Willingham, Cole, Lewis, and Leung (1997) designed a study to advance our understanding of gender differences in assessment to better ensure the fairness of current tests and learn what is most salient in developing the fairest assessments possible. They studied gender differences by taking into account the three main sources of variation in tests that are most useful in describing gender differences: construct, cohort, and sample differences. Variation by construct refers to differences in the knowledge, skills, or content that different tests purport to measure. Variation by cohort refers to differences in defined populations of

students taking the test (e.g., by age, grade, or from year to year). Variation by sample refers to the effects of having an "available" sample as opposed to a representative sample, in other words, a self-selected sample in which students elect to take the test (which generally results in a statistical phenomenon referred to as *range restriction*). In order to disentangle the effects of these differences, Willingham et al. studied multiple and diverse tests, and multiple age and year cohorts; and used tests with large and stable samples.

Their findings indicated that there were no differences in the overall average of test performance for females and males. On most of the tests studied, the gender difference in mean scores was negligible based on the average of the standardized mean differences between males and females across tests. In grade 12, the difference between the average score for males and females accounted for no more than 1 percent of all score variation. On average, women tended to score well on verbal sections, especially writing, and men tended to score well in the technical subjects. They found that these differences became more pronounced as students moved from grade 4 to grade 12. Also, there was somewhat greater variability in the scores of males than in those of females.

Despite the fact that males score higher than females on many measures of academic performance, there is some indication that females perform better in college than their test scores predict. The SAT actually predicts how well women will do in college better than it predicts for men. That is, the correlation of SAT scores with both first-year grade point average (FGPA) and individual course grades are higher for women than for men (College Board, 1998). High school grade point average (HSGPA), when used alone, tend to slightly underpredict FGPA for women. However, adding SAT scores to the equation reduces the underprediction. Differences in college course selection and high school course preparation account for much of the gender differences found when FGPA is used. A smaller proportion of women than men complete rigorous science and math courses in college. Science and math courses have been consistently shown to have more stringent grading standards than courses in the humanities, arts, social sciences, and English, which likely impacts the predictive validity of the SAT (College Board, 1998).

In their review of the literature on underprediction, Burton and Ramist (2001) demonstrated that analyzing FGPA within a single college course rather than across all college courses reduced the underprediction of FGPA for women (Burton and Ramist, 2001). Also, Stricker, Rock, and Burton (1991) found that controlling for gender differences in several explanatory variables regarding academic preparation, studiousness, and attitudes about mathematics was successful in reducing the over- and underprediction of FGPA. Women appeared to differ from men on these variables, which Stricker et al. (1991) believed to be linked to college grade attainment and were not necessarily measured by the SAT or HSGPA.

Racial/Ethnic Group Differences

Beginning in the 1960s, when the composition of students taking the SAT became more racially and ethnically diverse, mean score differences between racial/ethnic minority groups and white students became salient. The racial/ethnic groups that are typically the focus of concern are American Indian or Alaskan natives, African Americans, Asian Americans, and Hispanics. Camara and Schmidt (1999) reported that group differences appear to be fairly consistent across admissions tests, including the SAT, ACT, GRE, GMAT, MCAT, LSAT, and other measures of educational attainment such as the National Assessment of Educational Progress (NAEP), National Educational Longitudinal Survey (NELS), and AP Program Examinations. The largest gaps are between white and African American students, followed by white and Hispanic students. Hedges and Nowell (1998) reviewed several large-scale studies of test score differences since 1965 among secondary school students and concluded that African American students scored between .82 and 1.18 standard deviations below white students in composite test scores. Camara and Schmidt (1999) also found that Asian American students' test performance was nearly identical to that of white students, with two exceptions: (1) Asian American students scored about one-quarter standard deviation unit lower than white students on the SAT verbal section, and (2) Asian American students scored nearly one-half a standard deviation unit higher than white students on the GRE Quantitative test.

A number of different theories have been offered to account for racial/ethnic subgroup differences on standardized tests such as the SAT (Phillips, Brooks-Gunn, Duncan, Klebanov, and Crane, 1998; Sackett, Hardison, and Cullen, 2004; Scheuneman and Slaughter, 1991; Steele, 1997, 1998; Steele and Aronson, 1995; Stewart, 1999; Wiesen, 2005). Stewart (1999) highlighted the inequities minority groups have endured with regard to poor-quality academic preparation, including rundown school facilities and underprepared teachers, poverty, less family support, and discrimination. Phillips et al. (1998) noted differences between African Americans and whites in a set of family environment indicators including grandparents' educational attainment, mother's household size, quality of mother's high school, mother's perceived self-efficacy, children's birth weight, children's household size, and mother's parenting practices. These factors were found to explain more than half of the test score gap between 5- and 6-year-old African American and white children on the Peabody Picture Vocabulary Test—Revised. The addition of mother's cognitive skills

enabled the researchers to account for even more of the test score gap, or approximately two-thirds of the gap. It is thought that some of these disadvantages are expressed and become evident in standardized test scores.

The theory of stereotype threat is often cited as a possible contributor to mean differences in test scores, especially among different racial and ethnic groups. According to this theory, advanced by Claude Steele and Joshua Aronson (Steele, 1997, 1998; Steele and Aronson, 1995), when members of any group about whom negative stereotypes exist encounter threatening situations that highlight their awareness of the stereotype(s), the awareness is likely to cause disruptive effects of its own (Steele and Aronson, 1995). For many minority groups, it can be the common finding that members of certain minority groups tend to score lower than average on tests, leading to the concern or threat that they may do poorly on the test, which ultimately results in the confirmation of the stereotype (Sackett, Hardison, and Cullen, 2004).

It is important to note that the theory of stereotype threat has not been consistently demonstrated. Sackett et al. (2004) remarked that Steele and Aronson's (1995) findings of stereotype threat are quite misleading because Steele and Aronson statistically equated the prior SAT scores of the African American and white students in the study. When the two groups were exposed to stereotype threat prior to taking a test in a laboratory setting, African American students performed significantly more poorly on the test than white students. In the absence of threat, there were no performance differences. However, instead of showing that the elimination of stereotype threat eliminates African American–white test score differences, Steele and Aronson (1995) showed that score differences in the nonthreatening condition were consistent with what one would expect based on the differences in SAT scores for the two groups, while the test score difference was larger than would be expected (based on prior SAT scores) in the threatening condition.

Stricker (1998) also studied the effects of inquiring about examinees' gender and ethnicity on AP Calculus AB performance, and was unable to confirm Steele and Aronson's findings. Inquiring about gender and ethnicity prior to taking the exam produced no statistically significant effects on the test performance of African American, female, and other subgroup examinees. However, the participants in the Stricker study were of higher ability than the average SAT taker, and this may have impacted the results.

Disparities in the predictive validity of the SAT among different races/ethnicities have also been found. The SAT actually predicts how well white and Asian American students will do in college better than it predicts African American and Hispanic student performance. The correlation of SAT scores and HSGPA with FGPA is higher for white and Asian American students than for African American and Hispanic students (Young, 2004). Young and Kobrin (2001) reported that on average, African American and Hispanic students tend to earn slightly lower FGPAs than are predicted from their SAT scores and HSGPAs.

Young (2004) offered three possible explanations for the differential prediction of FGPA among racial/ethnic backgrounds. The first explanation was that traditional admissions measures tend to overestimate the academic skills of African American and Hispanic students with regard to their first-year course work. The second explanation was that college grades do not as accurately reflect the academic performance of African American or Hispanic students in their first year of college as they do for Asian American or white students. The third explanation was that an unknown factor exists that impacts the college GPA for African American and Hispanic students in a negative way. Fleming (1985) demonstrated that in predominantly black colleges, standardized test scores were a strong predictor of college grades, producing validity coefficients similar to those found in studies of white students. She uses these findings as evidence that the black college environment is more facilitative than the average "white" college environment, allowing black students to show more evidence of improvement in grades, cognitive skill, competitive abilities, and academic motivation.

Gender differences within ethnic groups have also been studied, although not as comprehensively as gender or ethnicity alone (Coley, 2001; Willingham et al., 1997). Willingham et al. (1997) examined ethnic differences by gender on the SAT and ACT. In general, they found that for each ethnic group, there was no difference between male and female test performance. The only exception was found among African American test-takers. African American female students were more likely to sit for the SAT or ACT than males, and were more likely to perform better than African American males on both tests. When GRE verbal, quantitative, and analytical data were examined, Willingham et al. found that African American and Asian American female test-takers performed slightly better in relation to males than white females did.

Language Group Differences

Examining subgroup differences within language groups is important because of the increasing linguistic diversity in the U.S. student population resulting from the growth of immigrant populations (Pennock-Román, 2002). For students who are limited in English proficiency, it is common to find mean verbal scores on standardized admissions tests considerably below norms (Pennock-Román, 1999). Pennock-Román (1990) noted that for a person whose first language is not English, a test administered in the English language becomes primarily

a test of language proficiency rather than a test measuring the skills and abilities it is intended to measure. In a study of how language characteristics of Hispanic students might affect their entrance to college, Duran et al. (1985) found that overall SAT scores were lower for people whose first language was not English, even when they indicated that English was their best language.

A number of researchers have found that due to the particular linguistic features of certain item types such as analogies or antonyms, items can be differentially easier or harder for bilingual test-takers (Alderman and Holland, 1981; Breland et al., 1974; Chen and Henning, 1985; Scheuneman, 1982; Schmitt, 1986; as cited in Pennock-Román, 1990). Also, bilingual students have been found to perform more slowly when completing problem-solving tasks in their less familiar language, presenting a potential problem on speeded or timed tests (Duran, Enright, and Rock, 1985).

In a study of the relationship between language background and the predictive validity of the SAT, Pennock-Román (1990) found that SAT verbal scores, in particular, were lower for students with less proficiency in English. These lower scores, however, did not systematically under- or overpredict Hispanic students' college achievement. As part of a larger study, Ramist, Lewis, and McCamley-Jenkins (1993) looked at student group differences in predicting college grades by language group based on student data supplied from 45 colleges and universities in the United States. They found that students whose best language was not English were more likely to be male (55 percent) and nonwhite (70 percent), and were twice as likely to be in the lower academic group based on their SAT scores and HSGPA. Compared to students whose best language was English, students with limited English proficiency had much lower mean SAT verbal scores (117 points lower or slightly more than one standard deviation). However, the same students had a higher mean SAT mathematics score (9 points higher).

Pennock-Román (2002) studied the effects of English proficiency on the GRE General Test and GRE Subject Tests in Psychology and Biology in comparison to the Prueba de Admisión para Estudios Graduados (PAEG), a test in Spanish used for admission to graduate schools in Puerto Rico, for a sample of students whose native language was Spanish. The findings from this study suggested that quantitative and subject tests were more highly related to underlying academic skills than were verbal sections. Pennock-Román advised that quantitative and subject tests be weighted more heavily than verbal sections in admissions decisions for students with limited English proficiency.

Socioeconomic Group Differences

Zwick (2002b) acknowledged that some critics of standardized tests have called the SAT a better measure of parental income than of verbal or math ability. Such critics have often cited positive correlations between test scores and family income as evidence of the test's biased nature (Zwick, 2004). However, studies have also shown that family income is related to many other measures of educational outcomes including HSGPA, the completion of certain courses, enrollment in college immediately after high school, greater expectations of attending a four-year college, completing an admission test and applying to a four-year college, and acceptance at a four-year college (Camara and Schmidt, 1999; Owings, McMillen, and Burkett, 1995; Zwick, 2002a, 2004).

Many hypotheses have been offered to explain performance differences by socioeconomic status (SES). Camara, Kobrin, and Sathy (2005) showed that academic rigor, or the amount of advanced or AP courses taken by a student in a number of different subject areas in high school, is related to parental income and parental education level. Therefore, the greater academic rigor of the higher SES students' course work may explain their better performance on the SAT.

Two additional explanations are commonly offered for performance differences by income level on standardized admissions tests (Zwick, 2004). The first explanation for this relationship is that ingrained in the content of these tests is a white, middle-class way of thinking that disadvantages lower-income students who think differently. Those subscribing to this hypothesis believe that tests of material taught in the classroom result in smaller score differences by income level. The second explanation offered for score differences by income level is that test-takers from higher-income families are able to afford expensive coaching and test-preparation materials which, in turn, boost the wealthier students' scores. Zwick (2004) investigated these two explanations by examining data from different tests and assessments such as the ACT, SAT Subject Tests, California High School Exit Exam, NAEP, as well as high school grades and other outcome measures. She found that even when the content of a test was more focused on material learned in the classroom there were still comparable differences in average score by income level.

Also, after reviewing the work of Camara and Schmidt (1999) and Owings, McMillen, and Burkett (1995), Zwick demonstrated that students' HSGPA, and their ability to meet selective admissions criteria (based on SAT score, HSGPA, participation in at least two extracurricular activities, and positive teacher perceptions) were also correlated with parental income level. Therefore, it is not only standardized test scores expressing the educational disadvantage of students from lower-income families, but many other factors related to academic achievement.

In their investigation of subgroup differences on standardized tests, Camara and Schmidt (1999) demonstrated another trend related to SES that is not

unique to the SAT. In their study, SES was investigated as a function of parental education and family income. Camara and Schmidt found that African American and Hispanic students from comparable SES scored lower on the SAT I and the NELS than Asian American and white college-bound students. The middle-third SES white students were more likely to score higher on such tests than the upper-third SES Hispanic and African American students.

Each of the studies described in the preceding section of this report provides important context to assess subgroup performance differences on the SAT as well as other assessments and academic performance measures. Most of these studies focused on cohorts of students in one particular year or a few consecutive years, thus offering only a snapshot of subgroup differences. In the following sections of this report, the scope is broadened to discuss the trends in SAT performance over nearly two decades.

Academic Performance Measures

The SAT Reasoning Test (formerly known as the SAT I: Reasoning Test) assesses student reasoning based on knowledge and skills developed by the students in their course work. It is typically taken by juniors or seniors in high school interested in attending colleges and universities in the United States. SAT scores in conjunction with HSGPA and other criteria are frequently used by colleges for admissions selection decisions.

In most of the years covered by this report (1987 to 2005), the SAT was a three-hour multiple-choice test in verbal and mathematical reasoning. Beginning in March 2005, a writing section was added to the SAT and the total time for the test was increased to 3 hours and 45 minutes. The SAT Subject Test in Writing (formerly known as SAT II: Writing Subject Test) was a one-hour multiple-choice test with one constructed response item (an essay) that measures how much students know about writing and how well they can apply that knowledge. This test is no longer administered and has been replaced by the SAT writing section. Students who took the SAT Subject Test in Writing generally scored higher on the SAT. Subject Tests are required by the University of California system schools as well as many other prestigious schools in the United States.

Most students taking either the SAT Reasoning Test or any of the SAT Subject Tests also complete the optional SAT Questionnaire (formerly known as the Student Descriptive Questionnaire) when they register to take SAT Program tests, providing valuable contextual information to aid in interpreting and understanding individual and group scores.

Demographics of SAT® College-Bound Seniors and Subgroup Performance on the SAT and SAT Subject Test in Writing from 1987 to 2006

The term *college-bound seniors* refers to a cohort of graduating students who take the SAT or SAT Subject Tests anytime during their high school career, up until March of their senior year. Each year in August, the College Board releases an annual report presenting data for high school graduates for that most recent year. The average test scores of the graduating class for each year can be found in Table B1. For instance, in the year 2005 the cohort includes students who became seniors in the 2004-05 school year and took the SAT anytime up until March 2005. Students are counted only once, no matter how often they tested, and only their latest scores and most recent SAT Questionnaire responses are summarized in the report. Because the accuracy of self-reported information has been documented (Freeberg, Rock, and Pollack, 1989), and the college-bound population is relatively stable from year to year, SAT Questionnaire responses from these students can be considered highly accurate. Trends in this report are drawn directly from these SAT Questionnaire responses and test scores.[1] A list of the relevant SAT Questionnaire questions and the response options appear in Appendix A.

This section of the report documents trends in subgroup differences on the SAT Reasoning Test and SAT Subject Test in Writing from 1987 to 2006. The primary groups of interest include gender, race/ethnicity, English language ability, and parental income. Because shifts that occur in the demographics of the test-takers are important

[1] From 1987 to 1995, individual student scores were converted to the recentered scale and then the mean was recomputed. From 1996 to 1999, nearly all students received scores on the recentered scale. Any score on the original scale was converted to the recentered scale prior to computing the mean. From 2000 to 2006, all scores were reported on the recentered scale. Interested readers are directed to Dorans (2002).

in understanding performance trends, there will be a brief discussion of the composition of each group over time (i.e., percentage of students in each subgroup over time) as well as the trends in test scores. Furthermore, as mentioned previously, substantial revisions were made to the SAT in 1994, and again in 2005. The short-term and long-term impact of the 1994 revisions on subgroup differences will be assessed, providing a source of comparison for the impact of the 2005 changes to the SAT on subgroup differences.

SAT Score Trends by Gender

Unlike the other subgroups of interest, students' gender is obtained from them both at the time of registration (on the SAT Questionnaire) and during the testing administration. Prior to starting the test, students are prompted by their test supervisor to indicate their name, date of birth, test site information, and gender on their answer sheet. As indicated in Table B2 (SAT-V) and Table B3 (SAT-M), women have made up slightly more than half of the SAT-taking population during this 20-year period.

As indicated in Table B2, over the last two decades, males have continued to score slightly higher than females on the SAT verbal/critical reading section. Over this time period, the difference in the scores has generally not exceeded 10 points, which is less than one-tenth of one standard deviation. Table B3 shows that males have also continued to outperform females on the SAT mathematics section, with score differences between 35 and 40 points. There has been, however, a very small decline in the gap in more recent years. Finally, the introduction of a new test in 1994 seems to have had virtually no impact on the trends for males and females.

As mentioned earlier, although those taking the SAT Subject Tests are generally higher performing, the population taking the SAT Subject Test in Writing is very similar in gender composition to the general SAT-taking population. As shown in Table B4, females have made up slightly more than half of the group (55 percent) and it has remained so until the last administration of the test in 2005. It should be noted that although trend information was not reported until 1996, the test was introduced in 1994. This resulted in a cohort of scores reported in 1995, but the scores were subsequently rescaled, so 1996 is the first year reported. There was another adjustment to the scale in 1998 that produced an increase of about 20 points between 1998 and 1999. Females have typically outperformed males on the SAT Subject Test in Writing, but in the last few years that gap has also narrowed from nearly 10 points when the test was introduced in 1995 to 4 points in 2004 and 2005.

Ethnicity

Unlike gender, students' racial/ethnic information is obtained from the SAT Questionnaire, which they complete when they register to take the test. (See Appendix A for the wording of the ethnicity question and response options.) As a result, when students fail to complete the questionnaire, there is a loss of information. As Table B5 indicates, there was an increasing number of students for whom ethnicity was not known, which peaked at around 25 percent in 2003. Students in this category can be divided into two groups: students who did not answer this question among other questions and students who didn't complete the entire questionnaire. Until 2002, there is a steady rate of between 5 and 6 percent of students not completing the SAT Questionnaire. There was also an increase in nonresponse to particular questions at this time. In 2003, we find that the nonresponse to the questionnaire jumped to 17 percent. In the mid-1990s the College Board began administering the questionnaire on the Internet, and by 2002, 44 percent of the graduating seniors were registering online. The online implementation of the questionnaire allowed students to more easily skip whole sections of the questionnaire. In 2003, subsequent revisions, which did not allow students to skip through as easily, were made to the Web-based questionnaire. As a result there were reductions in nonresponses (back down to 9 percent in 2006).

As indicated in Table B6, the percentage of white students generally declined over the last 20 years. In 1987, white students comprised just under three-quarters of the test-taking population (73 percent) and in 2006 they made up just over half of the test-taking population (56 percent). There have been small increases in the percentage of Asian American students, Hispanic students, and students identifying their ethnic group as "other."[2]

Students identifying themselves as "white" outperform all other subgroups on the SAT verbal/critical reading section (see Tables B7 and B8). Trends indicate that white students have performed best, followed by Asian American students, American Indian or Alaskan native students, Hispanic students, and African American students. Over the 20-year period, nearly all ethnic subgroups have experienced a small increase in scores, with the exceptions of American Indian or Alaskan native test-takers seeing a nearly 20-point gain and Asian American test-takers seeing

[2] The Hispanic category is composed of the following groups: Mexican/Mexican American; Puerto Rican; and Latin, South, or Central American/Other Hispanic/Latino. Detailed information on the subgroups that compose the Hispanic group can be found in Tables B8 and B10.

a large gain of 30 points. During this same time period, African American and Hispanic test-takers experienced only a modest increase in scores. In the first year after the introduction of a new test in 1994, scores were slightly higher (2–3 points) for all ethnic groups on the verbal section, which is consistent with an upward trend for all of the ethnic subgroups over the 20-year period. Trends on the SAT mathematics section indicate that Asian American students consistently outperform all other subgroups, followed by white students, American Indian or Alaskan native students, Hispanic students, and African American students (see Tables B9 and B10). The changes to the test in 1994 yielded only small increases in mathematics scores (1–2 points) across all subgroups. The most recent changes to the SAT occurred in 2005, and the 2006 data shown in this report represents the first cohort of college-bound seniors taking the new test. The 2006 data show slight score drops for the overall cohort and for most subgroups on the critical reading (formerly verbal) and mathematics sections. Because we have only one cohort of data after introducing the new SAT, the stability of subgroup differences has not yet been established. Since it is too soon to evaluate the impact of the most recent test changes on subgroup differences, this will not be discussed in this report. Over the 20-year period, all groups have seen an increase in mathematics scores ranging from small (10–20 points for African American and Hispanic students) to large (30–40 points for American Indian/Alaskan native, Asian American, and white students).

Trends on the SAT Subject Test in Writing are not in line with trends seen in the SAT verbal/critical reading and mathematics sections due to the nature of this population (Table B11 contains the percentage of students in each ethnic/racial group for the SAT Subject Test in Writing). As mentioned previously, students who take the SAT Subject Test in Writing typically exhibit higher than average SAT scores and generally represent a more selective group of students applying to schools requiring SAT Subject Test scores. As a result, trends from the SAT Subject Test in Writing indicate that while white students outperform all other subgroups, Asian American students and American Indian or Alaskan native students perform equally well but score about 30 points lower than white students, and African American and Hispanic students perform very similarly to each other and about 30 points lower than Asian American and American Indian or Alaskan native students (see Table B12). There does appear to be an indication that the gap between African American students and Hispanic students on Subject Test in Writing scores increased with African American students exhibiting slightly higher scores than Hispanic students in recent years.

English As a First Language

There are two questions on the SAT Questionnaire pertaining to language ability (see Appendix A). The first of which is "What language did you speak first?" and the second is "What language do you know best?" Students can respond: English, English and another language, or another language. Trends in nonresponse for "what language did you learn to speak first" are similar to trends that appeared with nonresponse in ethnicity (see Table B5). At its peak in 2003, the nonresponse rate for this question was just over a quarter of the population. Nonresponse rates in 2006 have returned to rates very similar to those prior to online registration, about 6 percent.

The number of students reporting English as their first language has steadily declined over the last 20 years (see Table B13). The number of students reporting that they speak English and another language has increased from 8 percent in 1987 to 13 percent in 2006. Students who report speaking another language make up a very small part of the test-taking population but have increased from 5 percent in 1987 to 8 percent in 2006.

Table B13 indicates that students who learned English first or English and another language have performed consistently on the SAT verbal/critical reading section over this 20-year period with students reporting English only performing an average across years of 514 and students who report speaking English and another language first scoring around 481. Students who learned another language first have scored lowest on the verbal/critical reading portion of the SAT, but have seen a steady improvement from about 436 in 1987 to about 467 in 2006. Language differences appear to have little impact on SAT mathematics section scores, but there is a small gap in scores for students who report speaking English and another language first (around 10 points lower than students who report speaking only English first or only another language first; see Table B14). Changes to the test in 1994 appear to have had little impact on the trends for language subgroups in both the verbal/critical reading and mathematics sections.

On the SAT Subject Test in Writing, students were very similar in language ability to the SAT Reasoning Test population (see Table B15). Over time the Subject Test population comprised fewer students who spoke only English and more students who spoke English and another language first and those who spoke another language first. As can be seen in Table B15, trends in the SAT Subject Test in Writing for language ability are very similar to trends in the SAT verbal/critical reading section: Students who report speaking English first score about 50 points higher than those who report speaking English and another language first, and they score about 20 points higher than students who report speaking another language first. The gap between students who

report speaking another language first and students who report speaking English and another language first has shrunk slightly over this time period.

English As Best Language

The second question pertaining to language involves what language a student knows best. This question was added to the questionnaire in 1995.[3] For English as best language, the same pattern of nonresponse trends appears as for the other questions from the SAT Questionnaire (see Table B5). Since this question was introduced to the SAT Questionnaire in 1995, the test-taking population has remained fairly consistent. About 80 percent of the students who take the SAT report knowing English best, while a very small percentage (2.5 percent) report knowing another language best. There has been a very small increase in students who report knowing another language and English best (bilingual students) from 6 percent in 1995 to 8 percent in 2006.

Gaps in scores for best language are much larger than they are for first language (see Table B16). Students who speak English best consistently score a little more than 50 points higher than students who know English and another language best on the SAT verbal/critical reading section, and they score a little more than 100 points higher than students who know another language best. However, students who speak another language best have seen a steady improvement in test scores from about 384 in 1995 to about 414 in 2006.

Much like English as a first language, best language appears to have little impact on SAT mathematics section scores (see Table B17). Students who report speaking English best have seen increases of about 11 points from 1995 to 2006, and students who report speaking another language best have seen increases of about 20 points in this time period. Students who are bilingual have also seen increases of about 20 points but still score about 10 to 20 points lower than the other groups.[4]

For the SAT Subject Test in Writing, students were very similar in composition in terms of best language, with students who speak English best making up the majority of the population (see Table B18). Trends in the SAT Subject Test in Writing for language ability are also very similar to the SAT verbal/critical reading section: Students who report speaking English best score about 50 points higher than those who report speaking English and another language best, and they, in turn, score about 50 points more than students who report speaking another language best. The gap between those who report speaking another language best and bilingual speakers has shrunk over the last few years from 55 points in 1996 to 36 points in 2005.

Parental Income

The final set of subgroups under investigation included income groups. Parental income is self-reported by the student on the SAT Questionnaire (Appendix A). As Table B5 indicates, unfortunately, parental income has suffered the largest loss of information due to nonrespondents.[5] The average nonresponse rate for income has been just over a quarter of the sample, with its peak in 2003 at 47 percent.[6] Starting in 1996, the income category "more than $100,000" was added to the list of options. Since that time, between 9 and 15 percent of students indicated a parental income in this range. Due to the large rate of nonresponse it is not possible to determine how this population has shifted over time, but it does appear that the number of students reporting lower incomes (less than $30,000) has reduced. The reader should also note that the relative economic condition of students within each of the income categories on the SAT Questionnaire has changed over the years due to inflation. In other words, students reporting an annual income less than $30,000 in 1987 were much better off than students reporting an annual income less than $30,000 in 2006. The reverse is also true of students in the highest income group (greater than $100,000).

Trends for income are, at best, tenuous due to the large number of students for whom we have no income data and the changes in the meaning of the income categories due to inflation. In light of these limitations, however, there are a few observations that can be noted. Students who report higher parental incomes perform better on the SAT verbal/critical reading section (see Table B19). The largest gap in verbal/critical reading scores occurs for students who report incomes less than $30,000, with scores about 100 points lower than those who report being in the highest income bracket. The gap in scores among students who report parental incomes between $30,000 and $100,000 does, however, appear to be shrinking slightly. Similarly, students who report higher parental incomes perform better on the SAT mathematics section (see Table B20). The same 100-point gap exists for the SAT mathematics section for students in the lowest versus

[3] In the early 1990s different versions of the question regarding best language were used. The first year that a consistent format was used and data were available was 1995.

[4] The student groups reporting "English and another language about the same" and "Another language" as their best language both had a two-point decline in their average SAT mathematics scores from 2005 to 2006.

[5] Nonresponse for the parental income question was not reported prior to 1992 in *College-Bound Seniors*.

[6] In order to ease interpretation, parental income categories were grouped based on students' responses. Students reported parental income in increments of $5,000–$10,000 (more than 10 categories; see response options in Appendix A); categories were combined resulting in five major categories (less than $30,000, $30,000–$50,000, $50,000–$70,000, $70,000–$100,000, and more than $100,000).

highest income group. While there does appear to be a slight decrease in average scores after the implementation of a new SAT in 1994 for the highest income group (above $70,000), this is simply an artifact of beginning to recategorize the data and including a response option of more than $100,000. The average of the two groups is consistent with previous years' trend data, and all groups experienced a small gain in scores.

As indicated in Table B21, on the SAT Subject Test in Writing the sample started out fairly evenly distributed in terms of income, but as time went on, nonresponse rates increased dramatically, and it appears as though the group is made up of higher-income students ($70,000 or more). Although overall scores are higher on the SAT Subject Test in Writing, the same trends appear. Students in the lowest income group score nearly 100 points lower than those in the highest income group.

Comparison of Subgroup Differences Across Performance Measures

The previous section of this report presented and described the 20-year trends in mean scores on the SAT Reasoning Test and SAT Subject Test in Writing for gender, racial/ethnic, language, and income subgroups. The following section discusses how the trends in subgroup differences compare to other measures, including HSGPA, the ACT Assessment, and NAEP administered to twelfth-graders.

The SAT can provide one measure of academic performance, but there are other measures that are also available to examine subgroup differences. For instance, high school grades are often cited as a useful admission criterion and in fact, the College Board recommends that HSGPA and test scores be used in conjunction with many other criteria (e.g., extracurricular activities) for admissions decisions. HSGPA is self-reported by students when they register to take the SAT and it has been noted to be fairly accurate (Freeberg et al., 1989).

The ACT Assessment is another measure of academic performance. The ACT Assessment is a college admissions test designed to assess students' general educational development and their ability to complete college-level work. Like the SAT, it is also typically taken by juniors or seniors in high school who are interested in attending American colleges and universities. The three-hour multiple-choice tests cover four skill areas: English, mathematics, reading, and science. In addition to providing a composite score and scores for each of four skill tests, the ACT also provides two subscores in English, three subscores in mathematics, and two subscores in reading. The individual tests have scores ranging from 1 to 36 and the composite is the average of all four tests. When students register to take the ACT they complete a student profile section that allows information on student demographics to be obtained along with educational plans, admissions information, etc. Data used in this report were collected from the National ACT High School Profile Report obtained at the ACT Web site (www.act.org).

One final academic measure examined in this paper is the National Assessment of Educational Progress (NAEP). NAEP, sometimes referred to as the "Nation's Report Card," is a nationally representative and continuing assessment of what students in the United States in grades 4, 8, and 12 know and can do in various subject areas including the arts, civics, economics, foreign language, geography, mathematics, reading, science, U.S. history, world history, and writing. These assessments have been conducted since 1969. NAEP administers reading and math assessments for grades 4 and 8 every other year in all states. In grade 12, NAEP must test reading and math at least as often as it has done in the past, or every four years on a nationally representative basis. The NAEP writing assessment is administered to grades 4, 8, and 12 periodically. NAEP does not provide scores for individual students or schools.

Examining other academic measures such as HSGPA, ACT scores, and NAEP scores in conjunction with SAT scores allows observations about subgroup differences for students at approximately similar time points (in their junior or senior year in high school). There are a number of content areas across all tests that allow us to examine the trends. For instance, the SAT, ACT, and NAEP all test math skills. There are some similarities in content for the SAT verbal/critical reading section, ACT Reading Test and NAEP Reading Test. Tracking subgroup performances on all three assessments in addition to high school grades can help shed some light on the achievement gap. Before presenting the data, we explain how subgroup differences are typically measured.

Methods for Examining Subgroup Differences

There are several ways of presenting data on subgroup differences. Raw mean differences are calculated as the average score of one subgroup (e.g., African American students) minus the average score of the total group.[7]

[7] The raw difference may also be calculated as the average score of a focal subgroup (e.g., African American or Hispanic students) minus the average score of a reference subgroup (e.g., white students).

When comparing score differences across more than one subgroup and/or more than one test, the raw mean differences may be misleading due to different variability in scores within subgroups and/or different test scales. A standardized mean difference (also sometimes called an effect size or z-score) is the raw mean difference divided by the standard deviation of scores in the total group. The standardized mean difference is comparable across all subgroups and tests and provides a uniform measure for ease of use and interpretation. In this report, the standardized mean differences for the racial/ethnic, language, and income subgroups were computed by taking the mean score for each subgroup and subtracting the mean score for the total group, then dividing by the total group (pooled) standard deviation. A slightly different method for calculating standardized mean differences was used for the gender comparisons. Because there are only two subgroups within gender, the standardized mean differences were computed by taking the mean score for females minus the mean score for males and dividing by the pooled standard deviation.

Cohen (1988) provided rules of thumb in interpreting effect sizes. He characterized an effect size of .2 "small," an effect size of .5 "medium," and an effect size of .8 "large." However, Cohen emphasized that the interpretation of effect sizes is particular to the data being analyzed, and that personal judgment on the part of the researcher is necessary to assess the practical significance of an effect. This report does not provide information on the statistical significance of subgroup differences and trends over time primarily because identifying a difference as statistically significant does not necessarily mean that the difference is meaningful or practically significant. Valentine and Cooper (2003) state:

> effects of the same size can sometimes be highly significant and at other times not [statistically] significant. Alternatively, effects that don't matter much can be highly statistically significant, while effects that matter a great deal can be statistically not significant. This problem comes about because the test of statistical significance actually confounds two independent pieces of information: the magnitude of the intervention's impact (the effect size) and the size of the sample. (p. 1)

Another approach to examining subgroup differences is to determine whether there is differential validity or differential prediction when using the SAT to predict various measures of college performance, such as freshman grades, cumulative grades, or graduation. Differential validity refers to differences in the magnitude of the correlation for different groups, and differential prediction refers to differences in the best-fitting regression lines or standard errors of estimate between groups of test-takers (Young and Kobrin, 2001). Prediction equations for different subgroups produced through simple or multiple regression analyses are compared to determine if the college performance measures for certain groups are either under- or overpredicted based on a prediction equation that was developed based on the entire population. This report will focus on standardized mean differences rather than differential validity and prediction. A thorough review of the latter can be found in Young and Kobrin (2001).

A distinction is also made between item-level and test-level subgroup differences. One or more items on any given test may be found differentially easier or more difficult for particular subgroups. Differential item functioning (DIF) is a statistical method for examining how test-takers from different subgroups perform on specific test items (Holland and Wainer, 1993). DIF examines whether items that are difficult or easy for one group are the same ones that are difficult or easy for another group of examinees at the same ability level. If analyses show that a test item is hard for one group but easy for another group, it is an indication that the item is performing differently for the two groups and may need to be investigated further before inclusion in future examinations (Morgan, 2005).

Test-level subgroup differences are indicated when average scores on a test differ significantly for different subgroups. Test-level subgroup differences are also referred to as test impact. A test may have several items exhibiting DIF for a particular subgroup, yet the average score for that subgroup on the test as a whole does not indicate adverse test impact. This report focuses on test-level subgroup differences rather than on item-level differences. In the next section we present a series of graphs that display the trend of standardized differences across performance measures for each of the major subgroups over the last 19 years.[8]

Standardized Differences for Females Versus Males Across Performance Measures

Figure 1 displays the trends in the standardized differences for females versus males across several academic performance measures, including HSGPA, SAT, ACT, and twelfth-grade

[8] To ease interpretation of the graphs, the trends for the different performance measures are coded by color, by line weight, and by the shape of the marker. The trends for the SAT are in black, and the trends for the ACT are in gray. The trends for HSGPA are shown with a thick black line and gray square markers, and the trends for NAEP are shown with open markers or "X's." The verbal/reading assessments are represented by diamonds, the math assessments are represented by circles, and the writing assessments are represented with asterisks. Both the reading and writing assessments have dashed lines, while the math assessments have solid lines. So for example, the SAT verbal/critical reading section is represented by a dashed black line with a diamond marker, and the ACT Mathematics Test is represented by a solid gray line with a circle marker. Each graph includes a key to help distinguish the trend lines.

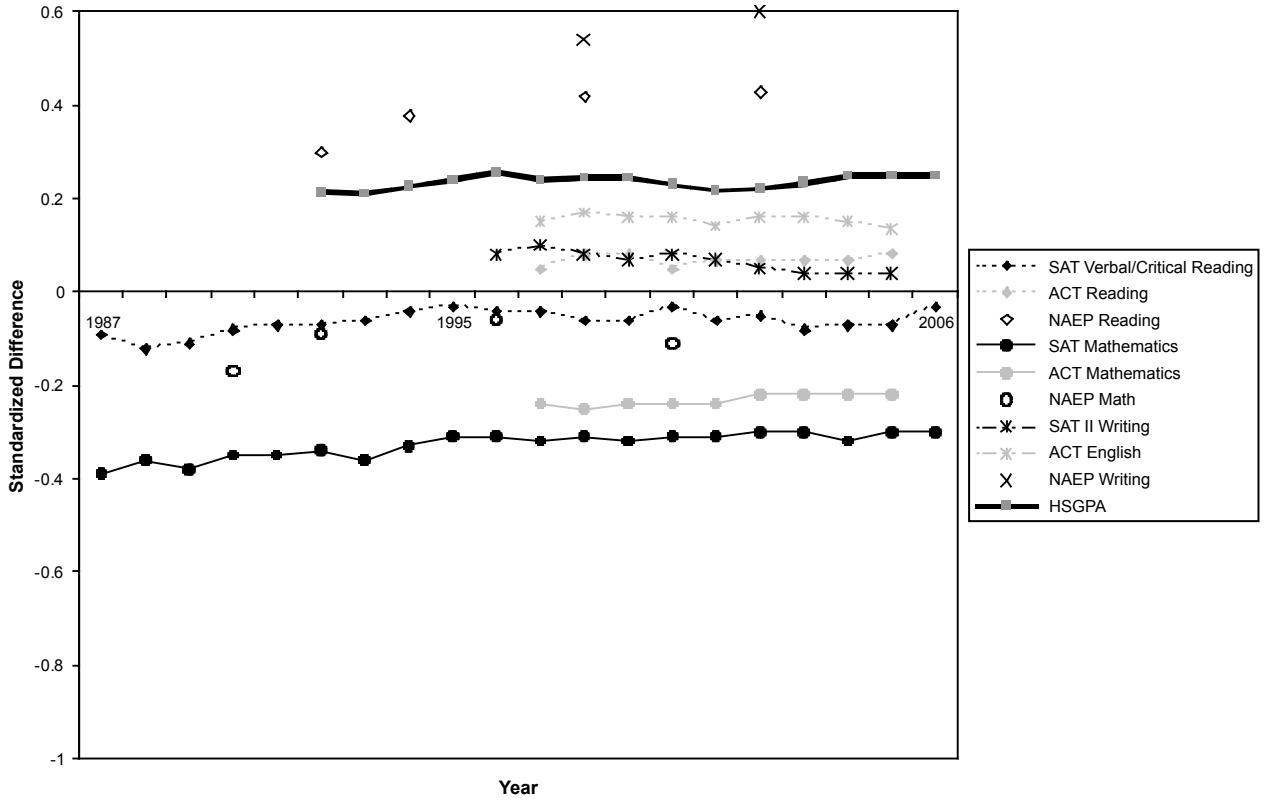


Figure 1. Standardized differences (female minus male) for gender across tests.

NAEP. On the NAEP reading and writing assessments, females scored higher than males by between .3 and .6 of a standard deviation (see Table B22). As mentioned previously, NAEP is a low-stakes assessment with a population that is very different from the college-bound seniors taking the SAT or ACT. The ACT English Test, ACT Reading Test, and SAT Subject Test in Writing/SAT writing section have very similar trends in standardized differences. All show a very slight positive advantage for females over males. The SAT verbal/critical reading section shows a very slight advantage for males, as does the NAEP mathematics assessment. The ACT Math Test and SAT mathematics section both show that males score between .2 and .4 of a standard deviation higher than females. The standardized difference between males and females on the SAT mathematics section decreased gradually from -.39 in 1987 to -.30 in 2006. All of these trends are generally very consistent, showing very little movement over the years, with the exception of NAEP reading and writing, which seem to show an increase in the advantage for females.

Standardized Differences for Racial/Ethnic Subgroups Across Performance Measures

Figures 2 through 6 show the standardized differences for each racial/ethnic group compared to the national SAT population (total group) across the major assessments (SAT verbal/critical reading and mathematics sections; SAT Subject Test in Writing; NAEP reading, writing, and math; and ACT Reading, English, and Mathematics Tests) and HSGPA over the last 20 years. Highlights of the results are given below:

African American test-takers

On all tests—verbal/critical reading, mathematics, and writing—African Americans have scored between .45 and .81 of a standard deviation lower than the total group (see Table B23). NAEP writing and HSGPA show the smallest standardized difference, at around -.45 to -.54, and the SAT mathematics section shows the largest standardized difference, at around -.74 to -.83. In most years, the standardized differences on the SAT verbal/critical reading section are smaller than those on the ACT Reading Test, and the standardized differences on SAT Subject Test in Writing are smaller than those on the ACT English Test.

American Indian/Alaskan Native test-takers

On most of the tests, including the SAT, ACT, and NAEP, American Indian or Alaskan natives have scored between .2 and .5 of a standard deviation lower than the total group (see Table B24). HSGPA has a smaller standardized difference compared to the other measures. In 1996 the standardized difference on the NAEP math assessment was close to -.8. The standardized differences on the ACT are slightly larger than those for the SAT.

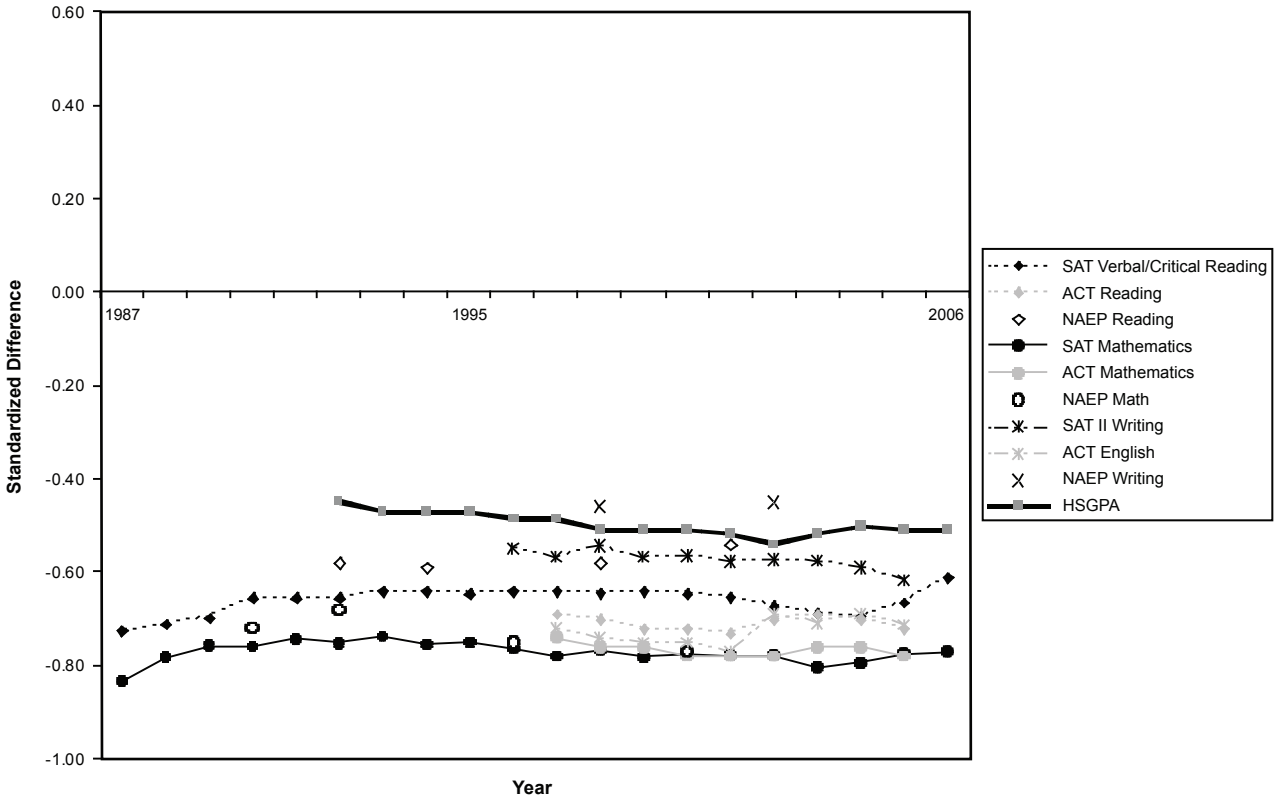


Figure 2. Standardized differences for African Americans across tests (African Americans minus total).

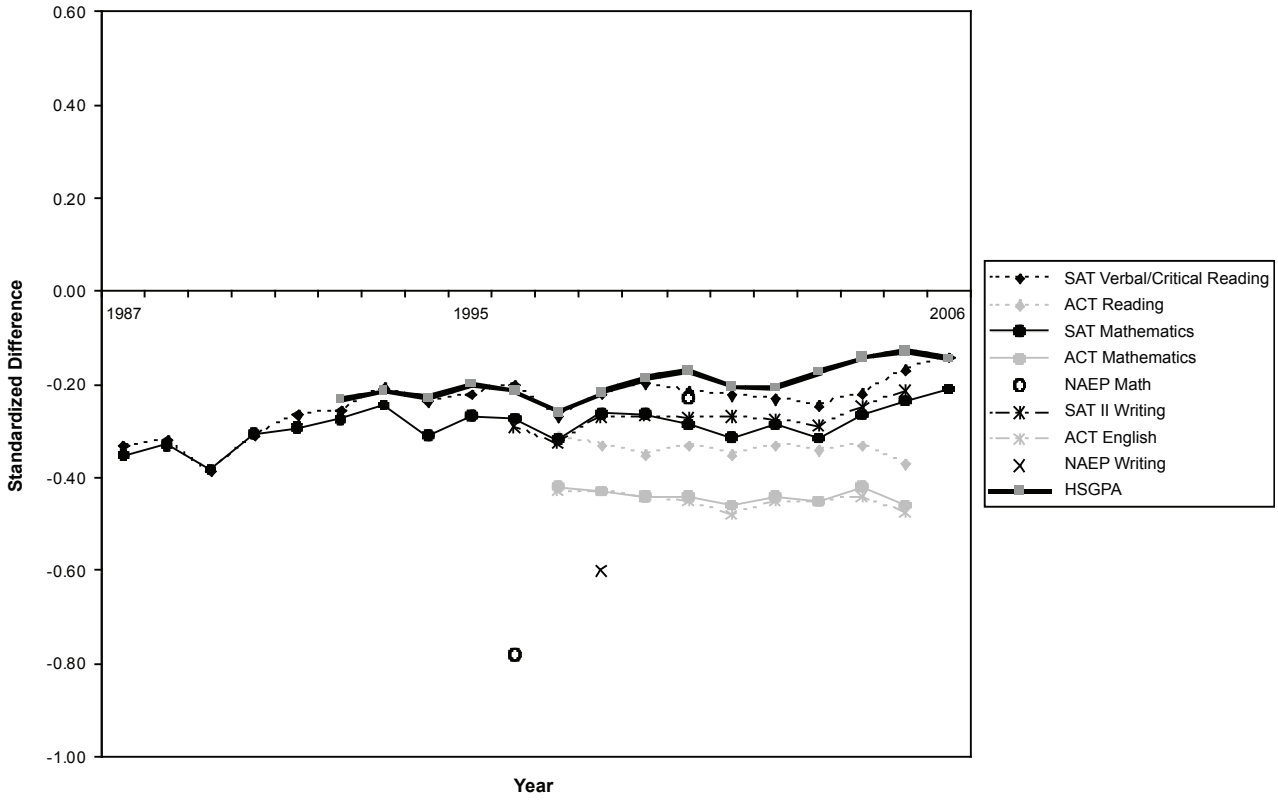


Figure 3. Standardized differences for American Indians or Alaskan Natives across tests (American Indians/Alaskan Natives minus total).

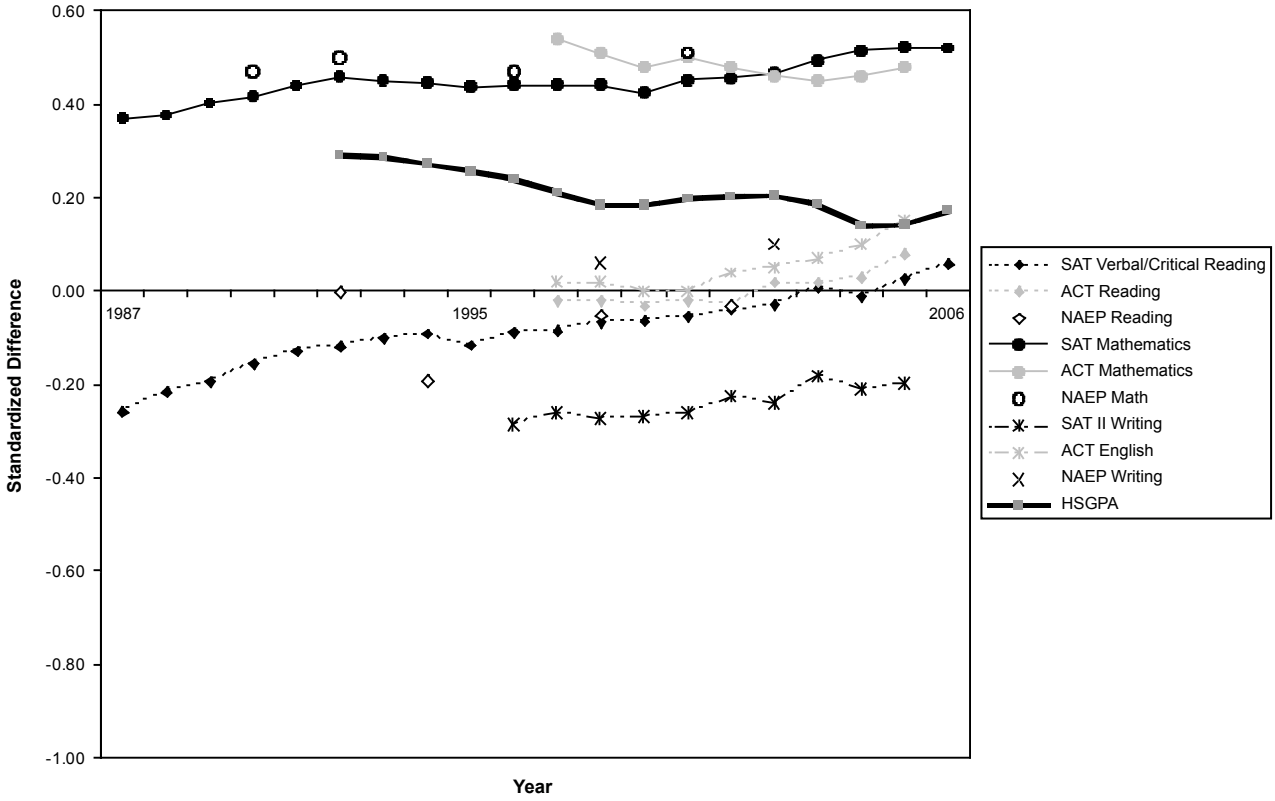


Figure 4. Standardized differences for Asian Americans across tests (Asian Americans minus total).

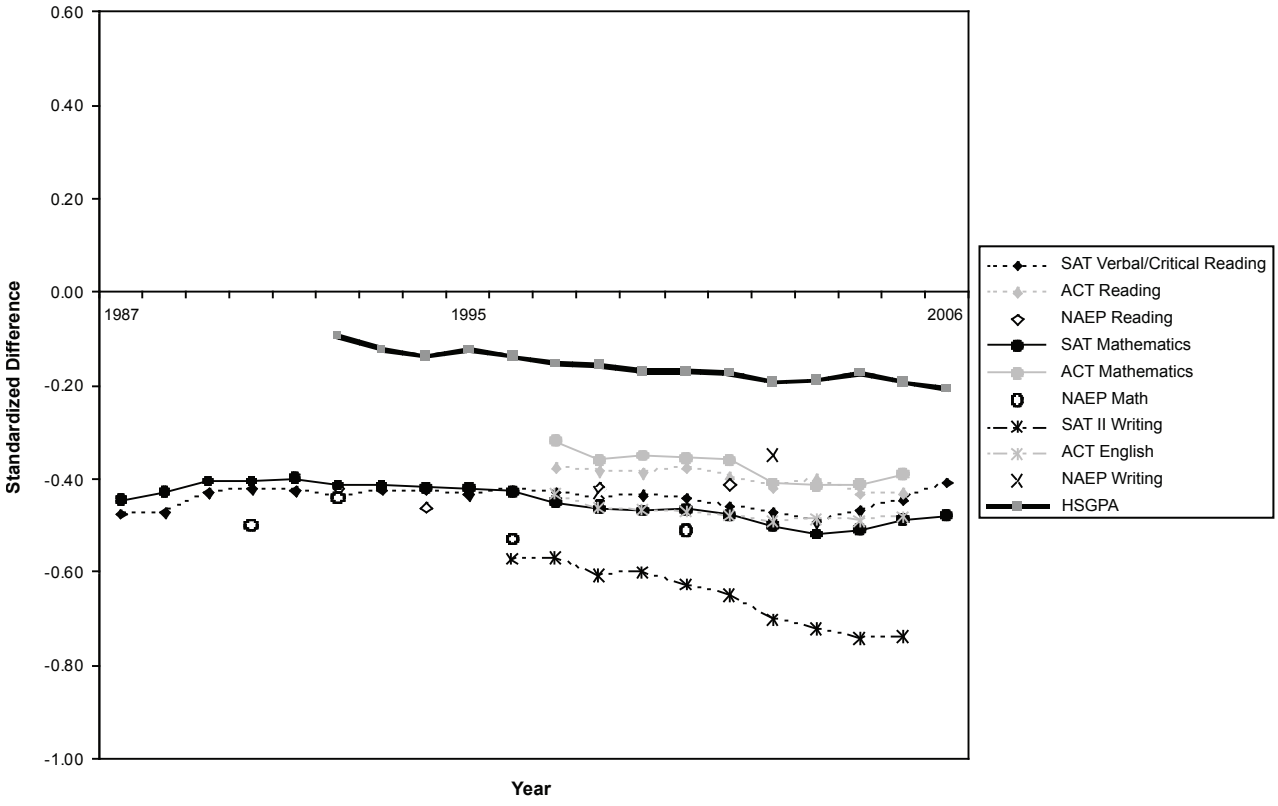


Figure 5. Standardized differences for Hispanics across tests (Hispanics minus total).

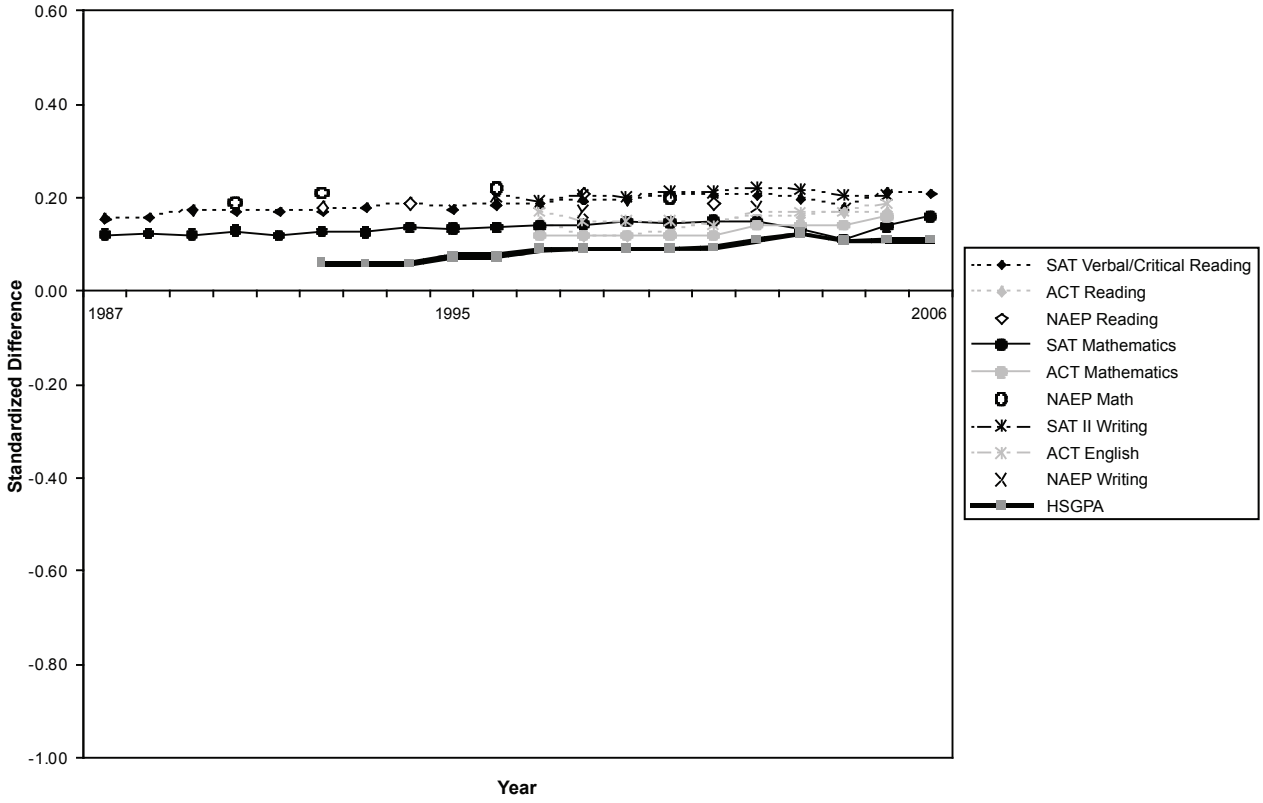


Figure 6. Standardized differences for whites across tests (whites minus total).

Asian American test-takers

On the NAEP math assessment, the ACT Mathematics Test, and SAT mathematics section, Asian American students have consistently scored between .4 and .6 of a standard deviation higher than the total group (see Table B25). The trends are also very consistent across the SAT-V, SAT Subject Test in Writing, ACT Reading and English Tests, and NAEP reading and writing assessments. The performance gap between Asian American students and the national SAT population has narrowed over the years. From 2003–2005, the standardized difference was practically zero, indicating that Asian American students were performing at the national average. Asian Americans now seem to be performing slightly higher than the national average. On the SAT Subject Test in Writing, Asian Americans scored about .3 of a standard deviation lower than the total group, but you can see a gradual decrease in the standardized differences from 1996 to 2005. Interestingly, there has been a decrease in the gap between Asian Americans' HSGPA and that of the national SAT population. Asian Americans' average HSGPA is becoming more similar to the national group. In 1992, Asian American students' average HSGPA was approximately one-third of a standard deviation higher than the total group, while in 2006 it was less than one-fifth of a standard deviation higher than the total group.

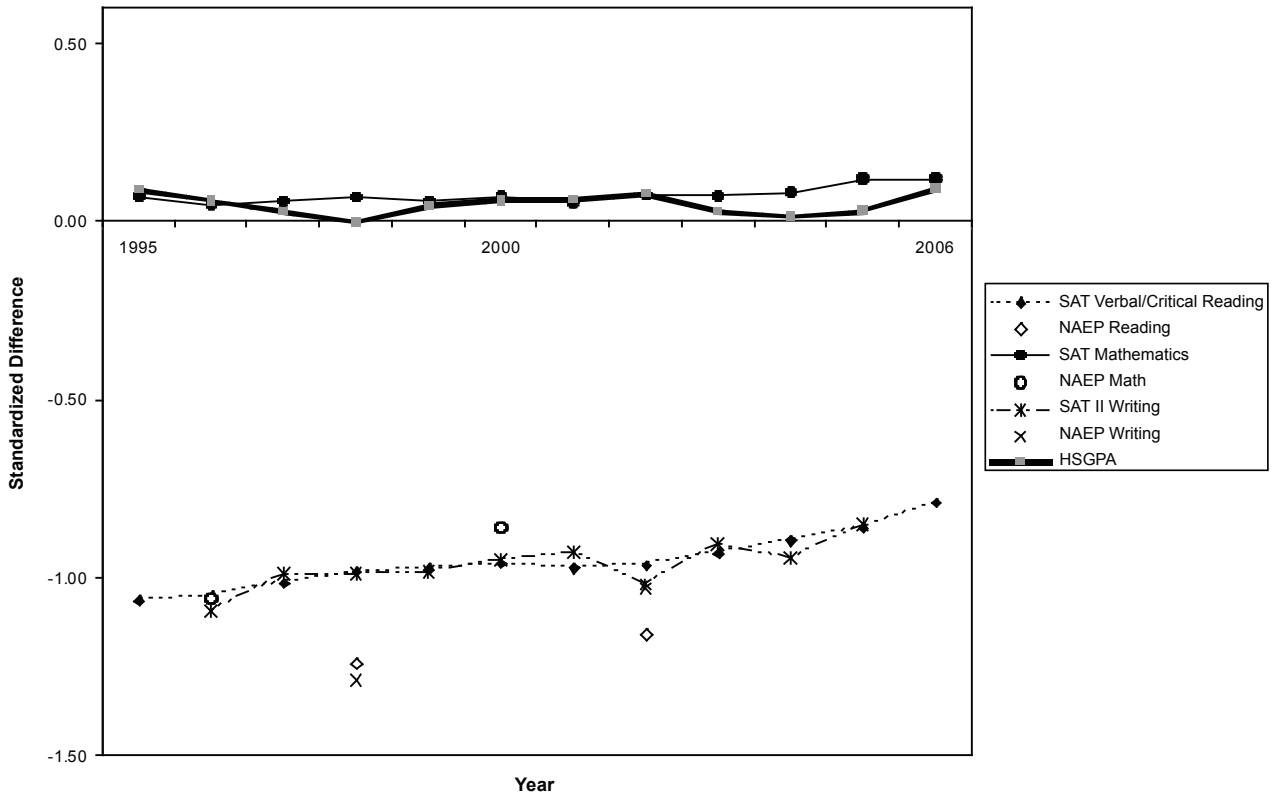
Hispanic test-takers

Hispanic students have scored .4 to .6 of a standard deviation lower than the total group on all measures except for HSGPA (see Table B26).[9] The standardized difference in HSGPA is considerably smaller than the standardized difference on the SAT, ACT, and NAEP. The standardized differences for Hispanic students on the SAT are larger than those found on the corresponding ACT tests. On the SAT Subject Test in Writing, the standardized difference for Hispanic students increased from 1996 to 2004 and did not change in 2005.

White test-takers

The standardized differences across tests for white students shows a very consistent and stable pattern

[9] For SAT data the Hispanic category is composed of the following groups: Mexican/Mexican American; Puerto Rican; and Latin, South, or Central American/Other Hispanic/Latino. For ACT data the Hispanic category is composed of Mexican American; and Puerto Rican/Hispanic for years 2000–2005 and Mexican American; and Puerto Rican/Cuban for 1997–1999. For NAEP data, the Hispanic category is not specified by more detailed subgroups.



Note: This figure is on a different scale than the other figures in this report.
Figure 7. Standardized differences for LEP students across tests (LEP minus total).

across all tests and across all years. The standardized difference for HSGPA is smaller than the other tests, as has been noted for each of the other racial/ethnic groups (see Table B27). The standardized differences on the SAT are slightly higher than those on the ACT.

Standardized Differences for Limited English Proficient (LEP) Test-Takers Across Measures

Figure 7 shows the standardized differences for LEP test-takers across measures. On the SAT, LEP students were defined as those indicating another language as their best language on the SAT Questionnaire. On NAEP, students are identified as LEP only if they are classified by their schools as limited English proficient. ACT data were not available for LEP students. The trends for HSGPA and the SAT mathematics section are comparable, with LEP students scoring just slightly higher than the total group. The trends are also comparable for the SAT verbal/critical reading section and SAT Subject Test in Writing. The standardized differences for LEP students on the NAEP reading and writing assessments are slightly larger than those for the SAT (see Table B28).

Standardized Differences for Low-Income and High-Income Test-Takers Across Measures

Figures 8 and 9 show the standardized differences across measures for students reporting an annual parental income of less than $30,000 a year and for students reporting an annual parental income greater than $100,000 as compared to the total SAT population. On its Web site, ACT reports mean scores by income only for the ACT Composite Score (a sum of the English, Reading, Math, and Science Tests). The standardized difference for students with low parental income in HSGPA is between one-tenth and two-tenths of a standard deviation (see Table B29). The standardized difference for students with low parental income on the SAT mathematics section is very similar to that of the ACT Composite Score, while the standardized difference on the SAT verbal/critical reading section tends to be slightly higher than the ACT Composite Score. The largest standardized differences for students with low parental income occur on the SAT Subject Test in Writing. On the SAT (both the verbal/critical reading and the mathematics sections), SAT Subject Test in Writing, and in HSGPA, there was an increase in the standardized differences for students with low parental income until 2003, when the standardized differences began to decrease.

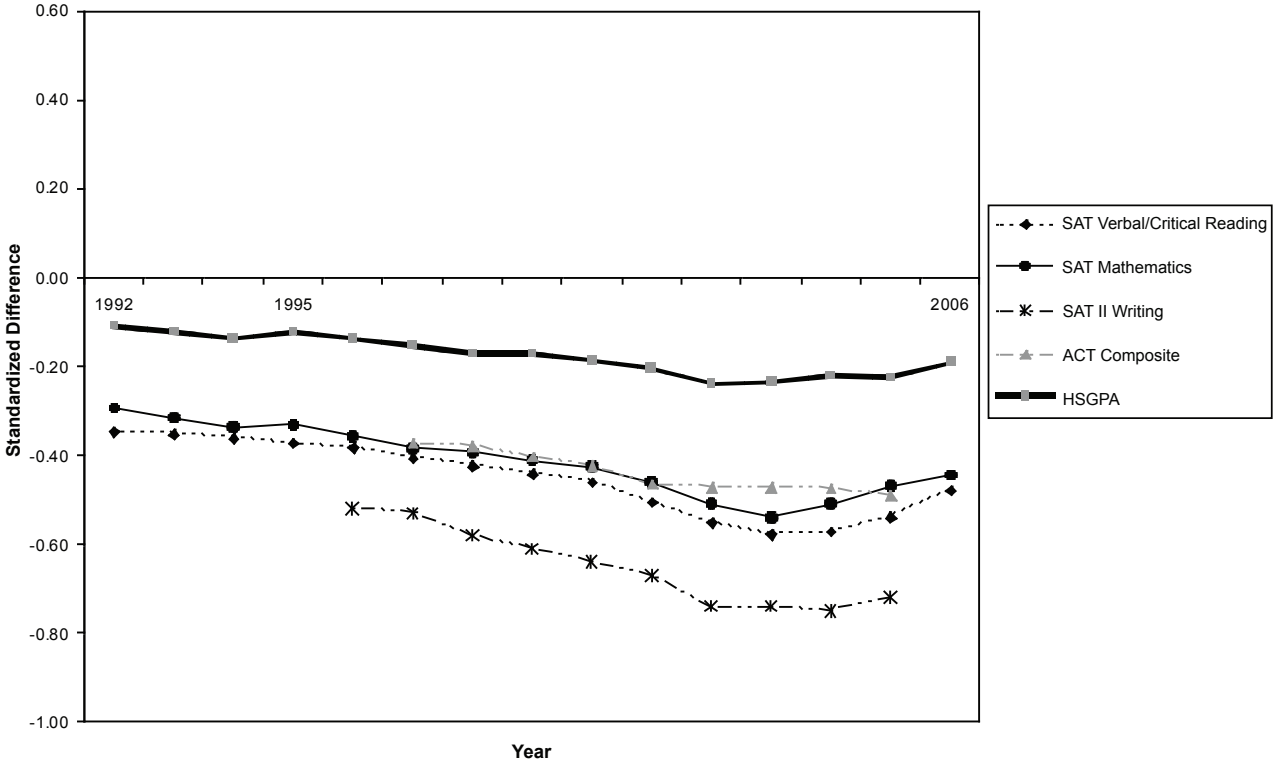


Figure 8. Standardized differences for low-income students* across tests (low income minus total).

*Low-income students are those reporting an annual parental income of less than $30,000.

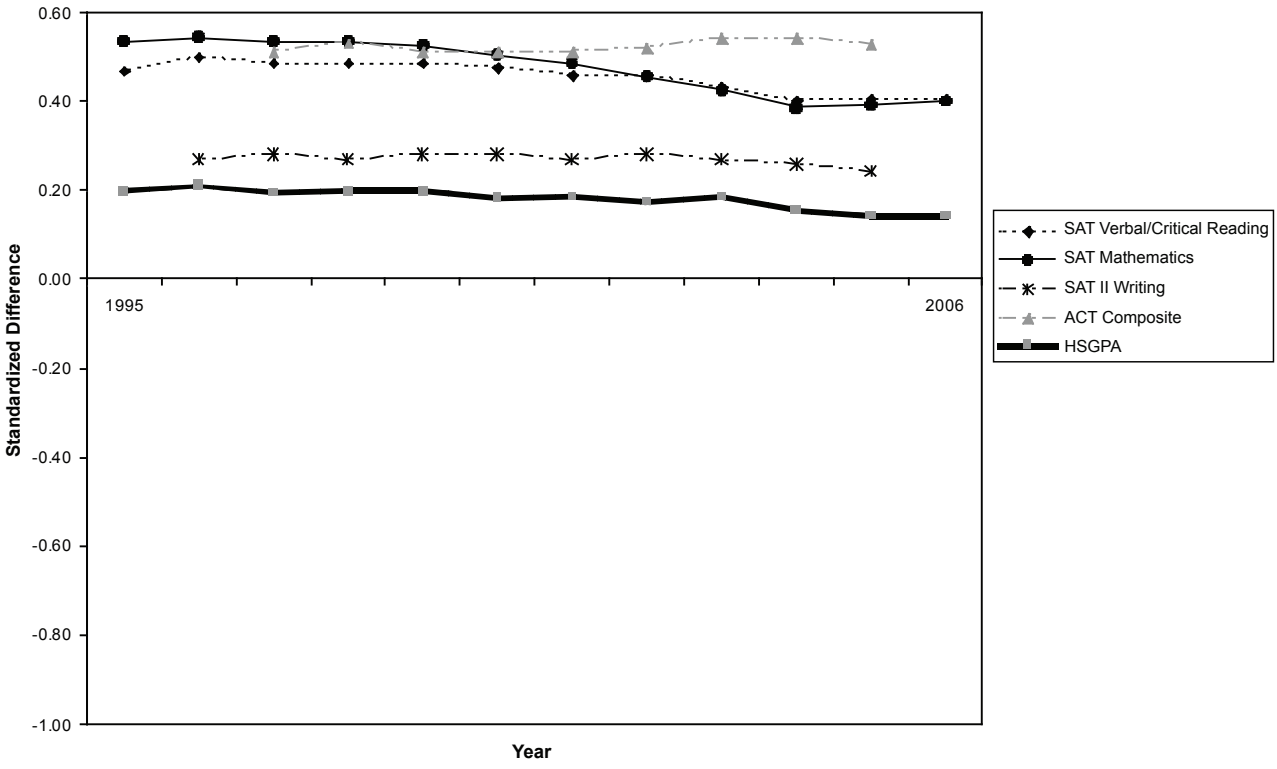


Figure 9. Standardized differences for high-income students* across tests (high income minus total).

*High-income students are those reporting an annual parental income of greater than $100,000.

For test-takers with high parental income,[10] HSGPA has the smallest standardized differences; these students typically earn grades around .2 of a standard deviation higher than the total group (see Table B30). The SAT and ACT trends are very similar, yet from 2002–2005 the standardized difference for students with high parental income on the ACT Composite Score was considerably higher than that of the SAT mathematics and verbal/critical reading sections. There is a slight upward trend for the ACT Composite Score, showing an increase in the advantage of test-takers with high parental income over the total group; conversely, the SAT verbal/critical reading and mathematics sections showed a decrease in standardized difference after 2002.

Summary and Discussion

This report presented 20 years of data on subgroup differences on the SAT and other performance measures. The trend data indicate that subgroup differences have remained generally consistent over nearly the last two decades. Subgroup differences were not affected by the substantial changes to the SAT that were made in 1994. The data for the first cohort taking the new SAT suggest that the performance gap may be narrowing for some subgroups. However, it is still too early to determine whether this is a stable trend, or a one-year fluctuation. Of course, the College Board will continually investigate subgroup performance on the SAT, and this report lays some of the groundwork for monitoring these trends by assembling the performance of students over the last couple of decades.

Furthermore, this report illustrated that the patterns found on the SAT by and large appear across other standardized assessments and performance measures. There were, however, some discrepancies that warrant further discussion and investigation. For instance, the subgroup differences reported in HSGPA are consistently smaller than those for the SAT, ACT, and NAEP. While some may be tempted to infer that HSGPA is a fairer measure, it is important to note that of all of the measures examined in this paper, HSGPA has the largest amount of inconsistency in terms of its measurement (i.e., the lowest reliability). In addition, there is substantial evidence supporting the notion of "grade inflation." For example, Camara et al. (2003) reported that the average HSGPA among SAT takers in 2002 far exceeded the HSGPA reported in 1976 (3.31 versus 3.00), while SAT mathematics scores were only slightly higher (516 versus 507) and SAT verbal scores were lower (506 versus 514). As more college-bound students report GPAs near or above 4.0, HSGPAs lose some of their value in differentiating among students (Camara et al., 2003).

There are a few instances where trends for the SAT differ from the trends for the other measures. For example, while females generally outscore males on verbal/critical reading and writing measures, this is not the case for the SAT verbal/critical reading section. Similarly, while males consistently outperform females across all math measures, the gap is wider on the SAT than on the ACT and NAEP. This is likely due to differences in the demographic composition of students who take the SAT compared to those taking the ACT and NAEP, as well as differences in the content in these three assessments. Willingham et al. (1997) noted that when females or males are overrepresented in a group, the standard mean difference will tend to move in a direction favoring the other gender. Also, Willingham and Cole (1997) remarked that male students' scores tend to spread out more at the top and the bottom than do female students' scores. Therefore, in a group of high-ability students such as those taking the SAT versus a more nationally representative sample such as students taking the NAEP, high-scoring males will tend to outscore high-scoring females.

Another possible explanation for the different subgroup difference trends for the SAT, ACT, and NAEP is differential degree of sample restriction across the three assessments. Willingham et al. (1997) noted that "other things equal, a select group of high-scoring examinees will show a more extreme standard mean difference than the total group for statistical reasons, even if the actual mean difference does not change" (p. 122, Table 3.9). Lewis and Willingham (1995) described a statistical model to predict what standard mean differences could be predicted if the same tests administered to representative twelfth-graders (e.g., NAEP) had been administered to the more select group of students who take college admissions tests. Overall, the predictions matched the actual observed gender differences fairly well, indicating that much of the difference in standard mean differences could be accounted for by the statistical effects of sample restriction. If it were possible to have examined only students who took all three assessments, the pattern of subgroup differences might be very different.

The group with the most variability in standardized differences across measures is the Asian American subgroup. This group consistently outperforms other groups on math assessments, but performs much more similarly to the total test-taking population on verbal and writing measures. The trends show that Asian Americans

[10] Trend data for high-income test-takers (students who report parental incomes of more than $100,000) are presented from 1996 to 2006 because prior to 1996, the SAT Questionnaire item on parental income had a maximum category of $70,000 and higher.

seem to be improving their standing in verbal and writing measures, but at the same time they seem to be losing their advantage in their HSGPAs. Future research might include an examination of this diverging trend of HSGPA and standardized assessment scores for this subgroup. One possibility is that Asian Americans' performance in subjects and/or constructs not assessed in the SAT or ACT have declined, thus resulting in lower HSGPAs but not reflected in the standardized assessment scores. Another possibility is that an increase in grade inflation over the years across all subgroups has diminished the advantage Asian Americans once had in HSGPAs.

Unfortunately, American Indian or Alaskan native, African American, and Hispanic students consistently demonstrate poorer performance across academic measures when compared to white and Asian American students. Camara and Schmidt (1999) noted that a high percentage of students from these groups are living in poverty and have experienced inequities in academic preparation, family support, and critical resources. Educators, researchers, and policymakers must continue to work together to implement school reform, intervention programs, and targeted strategies to improve educational outcomes for disadvantaged minority students.

Test-takers of Hispanic descent show a considerable discrepancy between their HSGPAs and their performance on the other measures. The standardized differences in Hispanics' HSGPAs are considerably smaller than the standardized differences on the SAT, ACT, and NAEP. In their examination of students with discrepant SAT scores and HSGPAs, Kobrin, and Milewski (2002) demonstrated that students with a high HSGPA in the presence of low SAT scores will not do any better in college than students with lower HSGPA scores but higher SAT scores. Therefore, the SAT may be a more accurate predictor of college performance than HSGPA for these students. As noted by Kobrin, Milewski, Everson, and Zhou (2003), future research should focus on identifying and measuring certain student- and school-level variables that might influence the discrepancy between HSGPAs and standardized test scores, such as grade inflation, teacher grading standards, and student motivation.

Students who report a best language other than English are another subgroup showing a substantial difference in performance across measures. As was found for Asian American students, students with limited English proficiency generally excel on the SAT mathematics section. However, these students perform nearly one standard deviation lower than the total group on the NAEP math assessment. This discrepancy can be explained by the differences in the population taking the SAT and NAEP assessments and the drastic difference in the stakes associated with these two tests.

Not surprisingly, there appears to be a significant reading and writing performance advantage for students identifying English as their best language, as opposed to English and another language or another language. Research shows that when given a chance to develop their academic English skills, English language learners (ELLs) can be more successful than their native English-speaking peers (Teachers of English to Speakers of Other Languages, 2003). Perhaps a focus on improving and enhancing academic English instruction for ELLs is warranted and in fact necessary to help these students increase and advance their educational opportunities in the United States.

The trends presented for low-income and high-income groups are noteworthy, but are tenuous due to a chronically large nonresponse rate to the income question on the SAT Questionnaire and changes in the meaning of the income categories over the years due to inflation. Interpreting trends for income groups leads one to believe that income alone could drive performance on standardized assessments. However, this obscures the relationship that other factors have in the academic and test performance of these groups. For instance, Camara, Kobrin, and Sathy (2005) revealed one very important factor: the academic rigor of a student's high school curriculum. Additional research is currently being planned to explore the relationship of individual and school-level factors that explain the association of socioeconomic factors with SAT scores and high school grades. Income and academic performance may always be highly related, but it is imperative to understand the mediating factors that can boost a student's preparedness for college despite his or her socioeconomic status.

This paper illustrates that the problem of subgroup differences is not exclusive to the SAT, and that the problem is pervasive across tests and across many different educational outcomes. Many researchers have noted that when minority and low-income students are given opportunities to take more rigorous courses, test score gaps diminish greatly. Future research should focus on the development and evaluation of educational interventions, programs, or strategies aimed at improving achievement outcomes for disadvantaged students. Exploring variables that can increase students' preparedness for college, regardless of certain disadvantages, remains a promising avenue of future research. Also, an examination of individual- and school-level factors contributing to academic performance discrepancies among measures is warranted. In keeping with the mission of the College Board, finding ways to increase access to educational excellence for these students must remain a priority.

References

Burton, N.W., Lewis, C., & Robertson, N. (1988). *Sex differences in SAT scores* (College Board Research Report No. 1988-9). New York: The College Board.

Burton, N.W., & Ramist, L. (2001). *Predicting success in college : SAT studies of classes graduating since 1980* (College Board Research Report No. 2001-2). New York: The College Board.

Cahalan, C., Mandinach, E.B., & Camara, W.J. (2002). *Predictive validity of SAT I: Reasoning Test for test-takers with learning disabilities and extended time accommodations* (College Board Research Report No. 2002-5). New York: The College Board.

Camara, W.J., Kimmel, E., Scheuneman, J., & Sawtell, E.A. (2003). *Whose grades are inflated?* (College Board Research Report No. 2003-4). New York: The College Board.

Camara, W.J., Kobrin, J.L., & Sathy, V. (2005, April). *Is there an SES advantage for the SAT and college success?* Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Canada.

Camara, W.J., & Schmidt, A.E. (1999). *Group differences in standardized testing and social stratification* (College Board Report No. 99-5). New York: The College Board.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Coley, R.J. (2001). *Differences in the gender gap: Comparisons across racial/ethnic groups in education and work* (ETS Policy Information Report). Princeton, NJ: Educational Testing Service.

The College Board. (2005). *College-bound seniors: A profile of SAT program test-takers*. New York: The College Board.

The College Board. (1998). *SAT and gender differences* (College Board Research Summary RS-04). New York: The College Board.

The College Board. (1972). *College-bound seniors: A profile of SAT program test-takers*. New York: The College Board.

Dorans, N.J. (2002). *The recentering of SAT scales and its effects on score distributions and score interpretations* (College Board Research Report 02-11). New York: The College Board.

Duran, R.P., Enright, M.K., & Rock, D.A. (1985). *Language factors and Hispanic freshmen's student profiles* (College Board Research Report 85-03). New York: The College Board.

FairTest. (2001). *The SAT: Questions and answers*. Retrieved March 10, 2005, from www.fairtest.org/facts/satfact.htm.

Fleming, J. (1985). *The usefulness of standardized test scores in the black college environment*. Report to the Southern Education Foundation Department of Research and Government Affairs, May 1985.

Freeberg. N.E., Rock, D.A., & Pollack, J. (1989). *Analysis of the revised student descriptive questionnaire: Phase II predictive validity of academic self-report* (College Board Research Report 89-8). New York: The College Board.

Freedle, R.O. (2003). Correcting the SAT's ethnic and social-class bias: A method for reestimating SAT scores. *Harvard Educational Review*, *73*, 1–43.

Hedges, L.V., & Nowell, A. (1998). Black-white test score convergence since 1965. In C. Jencks & M. Phillips (Eds.), *The black-white test score gap*. (149–81). Washington, DC: The Brookings Institute.

Holland, P.W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kobrin, J.L., & Milewski, G.B. (2002). *Students with discrepant high school GPA and SAT I scores* (College Board Research Note No. RN-15). New York: The College Board.

Kobrin, J.L., Milewski, G.B., Everson, H., & Zhou, Y. (2003, April). *An investigation of school-level factors for students with discrepant high school GPA and SAT scores*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, Illinois.

Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools*. Portsmouth, NH: Heineman.

Koretz, D.M. (2002). Limitations in the use of achievement tests as measures of educators' productivity. *Journal of Human Resources*, *37*, 752–77.

Lewis, C., & Willingham, W.W. (1995). *The effects of sample restriction on gender differences* (ETS Research Report No. RR-95-13). Princeton, NJ: Educational Testing Service.

Linn, R.L. (1990). Admissions testing: Recommended uses, validity, differential prediction, and coaching. *Applied Measurement in Education*, *3*, 297–318.

Linn, R.L. (1986). Bias in the college admissions. *Measures in the college admissions process: A College Board colloquium* (80–86). New York: The College Board.

Morgan, D. (2005). *Testing special populations/subgroup performance*. Unpublished manuscript, The College Board, New York.

Murphy, S.H. (1992). Closing the gender gap: What's behind the differences in test scores, what can be done about it. *The College Board Review*, *163*, 18–25.

National Center for Education Statistics. (2003). *Racial/ethnic differences in the path to a postsecondary credential* (Issue Brief #NCES 2003-005). Washington, DC: U.S. Department of Education Institute of Education Sciences.

Owings, J., McMillen, M., & Burkett, J. (1995). *Making the cut: Who meets highly selective college entrance criteria?* (NCES 95-732). Washington, DC: National Center for Education Statistics. Retrieved February 17, 2005, from http://nces.ed.gov/pubs95/95732.pdf.

Pennock-Román, M. (2002). Relative effects of English proficiency on general admissions tests versus subject tests. *Research in Higher Education*, *43*(5), 601–23.

Pennock-Román, M. (1999). *English proficiency and differences among racial and ethnic groups in mean SAT and GRE scores: A longitudinal analysis* (ETS Research Report No.1999-17). Princeton, NJ: Educational Testing Service.

Pennock-Román, M. (1990). *Test validity and language background: A study of Hispanic American students at six universities*. New York: The College Board.

Phillips, M., Brooks-Gunn, J., Duncan, G.L., Klebanov, P., & Crane, J. (1998). Family background, parenting practices, and the black–white test score gap. In C. Jencks & M. Phillips (Eds.), *The black–white test score gap* (103–45). Washington, DC: The Brookings Institute.

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1993). *Student group differences in predicting college grades: Sex, language and ethnic groups* (College Board Research Report No. 1993-1). New York: The College Board.

Rosner, J. (2003, April 14). On white preferences. *The Nation*, *276*, 24.

Sackett, P.R., Hardison, C.M., & Cullen, M.J. (2004). On interpreting stereotype threat as accounting for African American-White differences on cognitive tests. *American Psychologist*, *59*, 7–13.

Sacks, P. (1997). Standardized testing: Meritocracy's crooked yardstick. *Change*, *29*, 24–31.

Scheuneman, J.D., & Slaughter, C. (1991). *Issues of test bias, item bias, and group differences and what to do while waiting for the answers*. Unpublished manuscript, Educational Testing Service.

Steele, C.M. (1998). Stereotyping and its threat are real. *American Psychologist*, *53*(6), 680–81.

Steele, C.M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613–29.

Steele, C.M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797–811.

Stewart, D.S. (June 22, 1999). *Standardized testing and social stratification*. Invited Address at the Macalester College Forum on Higher Education, St. Paul, MN.

Stricker, L.J. (1998). *Inquiring about examinees' ethnicity and sex: Effects on AP Calculus AB Examination performance* (College Board Report No. 1998-1). New York: The College Board.

Stricker, L.J., Rock, D.A., & Burton, N.W. (1991). *Sex differences in SAT predictions of college grades* (College Board Research Report No. 1991-2). New York: The College Board.

Teachers of English to Speakers of Other Languages. (2003). *TESOL position paper on high-stakes testing for K–12 English language learners in the United States of America*. Alexandria, VA: Teachers of English to Speakers of Other Languages, Inc. (ERIC Document Reproduction Service No. ED477566).

Valentine, J., & Cooper, H. (2003). *Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes*. Washington, DC: What Works Clearinghouse.

Wendler, C., Ninneman, A., & Feigenbaum, M. (2001). *Evaluating the appropriateness of the SAT I: Reasoning Test for seventh and eighth graders* (College Board Research Note RN-12). New York: The College Board.

Wiesen, J.P. (2005). *Possible reasons for the black-white mean score differences seen with many cognitive ability tests: Informal notes to file*. Retrieved March 10, 2005, from http://personnelselection.com/Adimpact27.pdf.

Wilder, G., & Powell, K. (1989). *Sex differences in test performance: A survey of the literature* (College Board Research Report No. 1989-3). New York: The College Board.

Willingham, W.W., & Cole, N.S. (1997). *Gender and fair assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Willingham, W.W., Cole, N.S., Lewis, C., & Leung, S.W. (1997). Test performance. In W.W. Willingham & N.S. Cole (Eds.), *Gender and fair assessment* (55–126). Mahwah, NJ: Lawrence Erlbaum Associates.

Young, J.W. (2004). Differential validity and prediction: Race and sex differences in college admissions testing. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (289–302). New York: RoutledgeFalmer.

Young, J.W., & Fisler, J.L. (2000). Sex differences on the SAT: An analysis of demographic and educational variables. *Research in Higher Education*, *41*(3), 401–416.

Young, J.W., & Kobrin, J.L. (2001). *Differential validity, differential prediction, and college admission testing: A comprehensive review and analysis* (College Board Research Report No. 2001-6). New York: The College Board.

Zwick, R. (2002a). *Fair game? The use of standardized admissions tests in higher education*. New York: RoutledgeFalmer.

Zwick, R. (2002b). Is the SAT a 'wealth test'? *Phi Delta Kappan*, *84*(4), 307–11.

Zwick, R. (2004). *Rethinking the SAT: The future of standardized testing in university admissions*. New York: RoutledgeFalmer.

APPENDIX A: Student Descriptive Questionnaire Questions

Ethnicity Question

How do you describe yourself? (Mark only one.)

Responses before 1994:[11]

a. American Indian or Alaskan Native
b. Asian, Asian American, or Pacific Islander
c. Black or African American
d. Mexican American or Chicano
e. Puerto Rican
f. Latin American, South American, Central American, or other Hispanic
g. White
h. Other

Responses after 1994:

a. American Indian or Alaskan Native
b. Asian, Asian American, or Pacific Islander
c. African American or Black

Hispanic or Latino background:

d. Mexican or Mexican American
e. Puerto Rican
f. Latin American, South American, Central American, or other Hispanic or Latino
g. White
h. Other

Responses after 2005:

a. American Indian or Alaskan Native
b. Asian, Asian American, or Pacific Islander
c. Black or African American
d. Mexican or Mexican American
e. Puerto Rican
f. Other Hispanic, Latino, or Latin American
g. White
h. Other

English As a First Language Question

What language did you learn to speak first?

a. English
b. English and another language
c. Another language

English As Best Language Question (added in 1994)

What language do you know best?

a. English
b. English and another language about the same
c. Another language

Income Question

What was the approximate combined income of your parents before taxes last year? Include taxable and nontaxable income from all sources.

Responses before 1994:

a. Less than $10,000
b. About $10,000 to $15,000
c. About $15,000 to $20,000
d. About $20,000 to $25,000
e. About $25,000 to $30,000
f. About $30,000 to $35,000
g. About $35,000 to $40,000
h. About $40,000 to $50,000
i. About $50,000 to $60,000
j. About $60,000 to $70,000
k. More than $70,000

Responses after 1994:

a. Less than $10,000
b. About $10,000 to $15,000
c. About $15,000 to $20,000
d. About $20,000 to $25,000
e. About $25,000 to $30,000
f. About $30,000 to $35,000
g. About $35,000 to $40,000
h. About $40,000 to $50,000
i. About $50,000 to $60,000
j. About $60,000 to $70,000
k. About $70,000 to $80,000
l. About $80,000 to $100,000
m. More than $100,000

[11] Revisions were made to the Student Descriptive Questionnaire in 1994, but because cohort data are not released until students are seniors, these changes do not appear in reported data until *College-Bound Seniors 1995*.

APPENDIX B: Tables

Table B1

Descriptive Statistics for SAT Verbal/Critical Reading and Mathematics Totals: 1987 to 2006

	SAT Verbal/Critical Reading Totals			*SAT Mathematics Totals*		
Year	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*
1987	1,080,426	507	109	1,080,426	501	108
1988	1,134,364	505	107	1,134,364	501	106
1989	1,088,223	504	109	1,088,223	502	107
1990	1,025,523	500	110	1,025,523	501	108
1991	1,032,685	499	110	1,032,685	500	109
1992	1,034,131	500	110	1,034,131	501	109
1993	1,044,465	500	111	1,044,465	503	111
1994	1,050,386	499	111	1,050,386	504	110
1995	1,067,993	504	113	1,067,993	506	112
1996	1,084,725	505	110	1,084,725	508	112
1997	1,127,021	505	111	1,127,021	511	112
1998	1,172,779	505	111	1,172,779	512	112
1999	1,220,130	505	111	1,220,130	511	114
2000	1,260,278	505	111	1,260,278	514	113
2001	1,276,320	506	111	1,276,320	514	113
2002	1,327,831	504	111	1,327,831	516	114
2003	1,406,324	507	111	1,406,324	519	115
2004	1,419,007	508	112	1,419,007	518	114
2005	1,475,623	508	113	1,475,623	520	115
2006	1,465,744	503	113	1,465,744	518	115

Note: For 1987 to 1995, individual student scores were converted to the recentered scale and then the mean was recomputed. From 1996 to 1999, nearly all students received scores on the recentered scale. Any score on the original scale was converted to the recentered scale prior to computing the mean. From 2000 to 2006, all scores were reported on the recentered scale. Percentage values may not add to 100 percent due to rounding.

Table B2

Descriptive Statistics for SAT Verbal/Critical Reading by Gender: 1987 to 2006

	Gender							
	Female				*Male*			
Year	*N*	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*
1987	560,100	52	502	107	520,326	48	512	110
1988	590,299	52	499	106	544,065	48	512	108
1989	566,994	52	498	107	521,229	48	510	110
1990	535,103	52	496	108	490,420	48	505	111
1991	539,433	52	495	109	493,252	48	503	111
1992	542,383	52	496	109	491,748	48	504	111
1993	549,379	53	497	110	495,086	47	504	113
1994	557,323	53	497	110	493,063	47	501	113
1995	571,977	54	502	111	496,016	46	505	114
1996	580,127	53	503	109	504,598	47	507	112
1997	606,683	54	503	110	520,338	46	507	112
1998	630,817	54	502	109	541,962	46	509	112
1999	657,219	54	502	110	562,911	46	509	113
2000	676,947	54	504	110	583,331	46	507	112
2001	683,954	54	502	109	592,366	46	509	112
2002	711,630	54	502	110	616,201	46	507	112
2003	753,718	54	503	110	652,606	46	512	112
2004	758,737	53	504	111	660,270	47	512	113
2005	789,325	53	505	112	686,298	47	513	114
2006	785,019	54	502	111	680,725	46	505	114

Note: For 1987 to 1995, individual student scores were converted to the recentered scale and then the mean was recomputed. From 1996 to 1999, nearly all students received scores on the recentered scale. Any score on the original scale was converted to the recentered scale prior to computing the mean. From 2000 to 2006, all scores were reported on the recentered scale. Percentage values may not add to 100 percent due to rounding.

Table B3

Descriptive Statistics for SAT Mathematics by Gender: 1987 to 2006

	Gender							
	Female				Male			
Year	N	%	Mean	SD	N	%	Mean	SD
1987	560,100	52	481	102	520,326	48	523	111
1988	590,299	52	483	100	544,065	48	521	108
1989	566,994	52	482	101	521,229	48	523	110
1990	535,103	52	483	103	490,420	48	521	111
1991	539,433	52	482	102	493,252	48	520	112
1992	542,383	52	484	103	491,748	48	521	112
1993	549,379	53	484	105	495,086	47	524	114
1994	557,323	53	487	105	493,063	47	523	113
1995	571,977	54	490	107	496,016	46	525	114
1996	580,127	53	492	107	504,598	47	527	115
1997	606,683	54	494	108	520,338	46	530	114
1998	630,817	54	496	108	541,962	46	531	114
1999	657,219	54	495	110	562,911	46	531	115
2000	676,947	54	498	109	583,331	46	533	115
2001	683,954	54	498	109	592,366	46	533	115
2002	711,630	54	500	110	616,201	46	534	116
2003	753,718	54	503	111	652,606	46	537	116
2004	758,737	53	501	110	660,270	47	537	116
2005	789,325	53	504	111	686,298	47	538	116
2006	785,019	54	502	111	680,725	46	536	117

Note: For 1987 to 1995, individual student scores were converted to the recentered scale and then the mean was recomputed. From 1996 to 1999, nearly all students received scores on the recentered scale. Any score on the original scale was converted to the recentered scale prior to computing the mean. From 2000 to 2006, all scores were reported on the recentered scale. Percentage values may not add to 100 percent due to rounding.

Table B4

Descriptive Statistics for SAT Subject Test in Writing by Gender: 1996 to 2005

	Gender								Total		
	Female				Male						
Year	N	%	Mean	SD	N	%	Mean	SD	N	Mean	SD
1996	107,675	54	570	102	90,706	46	562	103	198,381	566	103
1997	111,132	55	573	103	92,330	45	563	104	203,462	568	104
1998	112,537	54	573	105	94,433	46	564	106	206,970	569	106
1999	116,549	54	597	107	97,701	46	589	107	214,250	593	107
2000	118,106	54	600	106	99,073	46	591	106	217,179	596	106
2001	123,449	55	598	106	102,275	45	591	106	225,724	595	106
2002	127,468	55	602	107	103,790	45	597	107	231,258	600	107
2003	134,174	55	598	108	109,468	45	594	108	243,642	596	108
2004	127,737	55	606	107	105,132	45	602	108	232,869	604	107
2005	132,069	55	607	108	108,295	45	603	107	240,364	605	107

Note: Although trend information was not reported until 1996, the SAT Subject Test in Writing was introduced in 1994, resulting in a cohort of scores reported in 1995, but the scores were subsequently rescaled and 1996 is the first year reported. There was another adjustment to the scale in May 1998 that resulted in an average increase of approximately 23 points in 1999. Percentage values may not add to 100 percent due to rounding.

Table B5

Percentage of Nonrespondents on the Total SAT Questionnaire and for Ethnicity, First Language, Best Language, and Income: 1987 to 2006

Year	*SAT Questionnaire*	*Ethnicity*	*First Language*	*Best Language*	*Income*
1987	5	7	7	--	14
1988	5	7	7	--	14
1989	5	8	7	--	15
1990	5	8	7	--	15
1991	5	8	7	--	14
1992	6	8	7	--	15
1993	6	9	8	--	16
1994	6	9	9	--	17
1995	6	9	8	10	17
1996	6	9	8	10	17
1997	6	9	9	11	18
1998	6	10	10	12	20
1999	6	12	11	13	22
2000	6	15	14	16	26
2001	5	16	17	16	32
2002	5	19	21	17	40
2003	17	25	26	23	47
2004	11	19	16	16	42
2005	4	10	7	7	33
2006	1	9	6	6	35

Note: The category of Best Language was added to the SAT Questionnaire in 1995. Percentage values may not add to 100 percent due to rounding.

Table B6

Percentage of Students in Racial/Ethnic Subgroups: 1987 to 2006

Year	*African American*	*American Indian/ Alaskan Native*	*Asian American/ Pacific Islander*	*Hispanic*	*Mexican/ Mexican American*	*Puerto Rican*	*Latin, South, or Central American/ Other Hispanic/ Latino*	*White*	*Other*	*No Response*
1987	8	1	5	5	2	1	2	73	1	7
1988	9	1	6	5	2	1	2	72	1	7
1989	9	2	6	5	2	1	2	69	1	8
1990	9	1	7	6	3	1	2	68	1	8
1991	10	1	7	6	3	1	2	67	2	8
1992	10	1	8	7	3	1	3	66	2	8
1993	10	1	8	7	3	1	3	64	2	9
1994	10	1	8	7	3	1	3	63	2	9
1995	10	1	8	7	3	1	3	63	2	9
1996	10	1	8	8	3	1	3	63	3	9
1997	10	1	8	8	4	1	3	62	3	9
1998	10	1	8	8	3	1	3	60	3	10
1999	10	1	8	8	4	1	3	59	3	12
2000	9	1	8	8	4	1	3	57	3	15
2001	9	1	8	8	4	1	3	55	3	16
2002	9	1	8	8	4	1	3	53	3	19
2003	9	1	7	8	4	1	3	48	3	25
2004	10	1	8	9	4	1	3	51	3	19
2005	10	1	9	10	5	1	4	56	4	10
2006	10	1	9	10	4	1	5	56	4	9

Note: The Hispanic category is composed of the following groups: Mexican/Mexican American; Puerto Rican; and Latin, South, or Central American/Other Hispanic/Latino. Percentage values may not add to 100 percent due to rounding.

Table B7

Descriptive Statistics for SAT Verbal/Critical Reading by Racial/Ethnic Subgroups: 1987 to 2006

	N	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*
Year	*African American*			*American Indian/Alaskan Native*			*Asian American/Pacific Islander*			*Hispanic*		
1987	88,037	428	99	10,107	471	95	58,216	479	134	49,913	455	--
1988	97,483	429	100	12,330	471	93	64,102	482	131	54,432	455	--
1989	96,615	428	100	18,005	462	91	68,254	483	131	57,864	457	--
1990	94,311	428	99	10,466	466	96	71,792	483	132	61,081	454	--
1991	100,209	427	99	7,843	470	100	76,703	485	133	66,251	452	--
1992	99,126	428	100	7,412	472	102	78,387	487	135	69,193	452	106
1993	102,939	429	101	7,488	477	104	78,693	489	135	73,420	453	106
1994	102,679	428	100	8,150	473	103	81,097	489	135	77,828	452	104
1995	103,872	431	104	8,936	479	105	81,514	491	130	80,092	455	108
1996	106,573	434	99	8,737	483	103	84,319	496	127	81,985	458	105
1997	110,462	434	101	10,677	475	102	89,236	496	126	86,068	457	106
1998	114,912	434	99	10,159	480	102	94,066	498	125	90,412	456	105
1999	119,394	434	100	8,261	484	105	96,108	498	126	94,667	457	104
2000	119,591	434	100	7,658	482	107	96,717	499	124	97,872	457	104
2001	120,506	433	100	7,622	481	108	102,312	501	124	101,172	455	104
2002	122,684	430	99	7,506	479	107	103,242	501	124	104,155	452	103
2003	125,657	431	99	7,452	480	106	100,970	508	123	107,492	453	103
2004	137,953	430	99	8,219	483	108	112,542	507	121	122,380	456	102
2005	153,132	433	99	8,916	489	108	134,996	511	121	144,196	458	103
2006	150,643	434	98	9,301	487	106	138,303	510	122	151,761	457	102
	White			*Other*			*No Response*					
1987	788,613	524	100	13,102	480	126	72,438	--	--			
1988	813,116	522	99	14,094	485	124	78,807	--	--			
1989	752,257	523	100	13,454	490	124	81,774	--	--			
1990	694,976	519	100	14,632	484	126	78,265	--	--			
1991	687,231	518	100	16,300	486	125	78,148	--	--			
1992	680,806	519	100	17,771	491	126	81,436	478	124			
1993	670,965	520	101	19,614	497	125	91,346	479	124			
1994	662,107	520	101	22,198	500	124	96,327	479	124			
1995	674,343	524	102	25,113	506	123	94,123	482	125			
1996	681,053	526	101	28,099	511	120	93,959	486	123			
1997	693,736	526	101	31,050	512	120	105,792	489	123			
1998	704,462	526	101	35,762	511	118	123,006	490	123			
1999	717,632	527	101	38,130	511	119	145,938	492	122			
2000	712,105	528	100	38,634	508	119	187,701	495	120			
2001	703,724	529	100	38,680	503	118	202,304	497	119			
2002	698,659	527	100	38,967	502	117	252,618	501	118			
2003	670,260	529	100	39,146	501	116	355,347	510	114			
2004	719,753	528	100	46,615	494	116	271,545	522	120			
2005	824,776	532	101	58,167	495	117	151,440	511	136			
2006	825,921	527	102	54,469	494	116	135,346	487	135			

Note: For 1987 to 1995, individual student scores were converted to the recentered scale and then the mean was recomputed. From 1996 to 1999, nearly all students received scores on the recentered scale. Any score on the original scale was converted to the recentered scale prior to computing the mean. From 2000 to 2006, all scores were reported on the recentered scale. The Hispanic category is composed of the following groups: Mexican/Mexican American; Puerto Rican; and Latin, South, or Central American/Other Hispanic/Latino. Standard deviations for the Hispanic group prior to 1992 could not be calculated because only the aggregate data for the Hispanic subgroups were available. The mean scores for nonrespondents prior to 1992 are not known because they were not reported in *College-Bound Seniors*.

Table B8

Descriptive Statistics for SAT Verbal/Critical Reading by Hispanic Subgroups: 1987 to 2006

	Mexican/Mexican American			*Puerto Rican*			*Latin, South, or Central American/Other Hispanic/Latino*		
Year	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*
1987	20,512	457	99	10,173	436	105	18,633	464	109
1988	22,722	459	102	11,497	431	103	20,213	463	110
1989	25,207	459	100	11,415	437	105	21,242	466	111
1990	26,073	457	99	11,400	435	106	23,608	459	111
1991	28,602	454	99	12,065	436	106	25,584	458	110
1992	30,336	449	101	12,091	442	106	26,766	459	111
1993	32,355	451	102	12,645	443	106	28,420	460	111
1994	35,397	448	100	13,036	444	103	29,395	460	109
1995	36,323	452	104	13,056	446	108	30,713	464	111
1996	36,689	455	101	13,103	452	104	32,193	465	109
1997	39,737	451	102	13,208	454	104	33,123	466	109
1998	41,028	453	102	13,635	452	104	35,749	461	108
1999	43,160	453	102	13,986	455	103	37,521	463	108
2000	44,921	453	101	14,147	456	103	38,804	461	108
2001	46,849	451	101	14,074	457	104	40,249	460	106
2002	48,255	446	101	14,273	455	103	41,627	458	106
2003	50,375	448	101	14,569	456	102	42,548	457	106
2004	57,739	451	99	16,449	457	102	48,192	461	105
2005	66,968	453	100	19,402	460	103	57,826	463	106
2006	64,019	454	99	19,008	459	104	68,734	458	105

Note: For 1987 to 1995, individual student scores were converted to the recentered scale and then the mean was recomputed. From 1996 to 1999, nearly all students received scores on the recentered scale. Any score on the original scale was converted to the recentered scale prior to computing the mean. From 2000 to 2006, all scores were reported on the recentered scale.

Table B9

Descriptive Statistics for SAT Mathematics by Racial/Ethnic Subgroups: 1987 to 2006

	N	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*
Year	*African American*			*American Indian/Alaskan Native*			*Asian American/Pacific Islander*			*Hispanic*		
1987	88,037	411	97	10,107	463	99	58,216	541	116	49,913	453	--
1988	97,483	418	95	12,330	466	95	64,102	541	114	54,432	456	--
1989	96,615	421	94	18,005	461	93	68,254	545	116	57,864	459	--
1990	94,311	419	96	10,466	468	98	71,792	546	117	61,081	457	--
1991	100,209	419	94	7,843	468	99	76,703	548	118	66,251	457	--
1992	99,126	419	96	7,412	471	102	78,387	551	119	69,193	456	102
1993	102,939	421	98	7,488	476	104	78,693	553	120	73,420	457	104
1994	102,679	421	100	8,150	470	103	81,097	553	119	77,828	458	103
1995	103,872	422	96	8,936	476	103	81,514	555	119	80,092	459	102
1996	106,573	422	96	8,737	477	104	84,319	558	120	81,985	460	103
1997	110,462	423	97	10,677	475	101	89,236	560	120	86,068	460	103
1998	114,912	426	97	10,159	483	103	94,066	562	120	90,412	460	102
1999	119,394	422	99	8,261	481	106	96,108	560	123	94,667	458	103
2000	119,591	426	98	7,658	481	106	96,717	565	122	97,872	461	103
2001	120,506	426	99	7,622	479	106	102,312	566	122	101,172	460	103
2002	122,684	427	99	7,506	483	106	103,242	569	124	104,155	459	104
2003	125,657	426	99	7,452	482	106	100,970	575	124	107,492	459	103
2004	137,953	427	99	8,219	488	107	112,542	577	123	122,380	460	101
2005	153,132	431	99	8,916	493	106	134,996	580	121	144,196	464	102
2006	150,643	429	100	9,301	494	106	138,303	578	122	151,761	463	103
	White			*Other*			*No Response*					
1987	788,613	514	102	13,102	482	117	72,438	--	--			
1988	813,116	514	99	14,094	487	113	78,807	--	--			
1989	752,257	515	101	13,454	493	115	81,774	--	--			
1990	694,976	515	101	14,632	492	117	78,265	--	--			
1991	687,231	513	101	16,300	492	117	78,148	--	--			
1992	680,806	515	101	17,771	498	117	81,436	485	121			
1993	670,965	517	103	19,614	501	118	91,346	489	123			
1994	662,107	519	101	22,198	504	116	96,327	489	121			
1995	674,343	521	104	25,113	510	115	94,123	492	121			
1996	681,053	523	104	28,099	512	116	93,959	494	121			
1997	693,736	526	103	31,050	514	115	105,792	502	122			
1998	704,462	528	103	35,762	514	114	123,006	503	121			
1999	717,632	528	104	38,130	513	117	145,938	505	122			
2000	712,105	530	103	38,634	515	116	187,701	509	119			
2001	703,724	531	103	38,680	512	116	202,304	510	119			
2002	698,659	533	103	38,967	514	116	252,618	516	118			
2003	670,260	534	104	39,146	513	116	355,347	525	117			
2004	719,753	531	102	46,615	508	114	271,545	535	119			
2005	824,776	536	102	58,167	513	114	151,440	525	131			
2006	825,921	536	103	54,469	513	115	135,346	506	128			

Note: For 1987 to 1995, individual student scores were converted to the recentered scale and then the mean was recomputed. From 1996 to 1999, nearly all students received scores on the recentered scale. Any score on the original scale was converted to the recentered scale prior to computing the mean. From 2000 to 2006, all scores were reported on the recentered scale. The Hispanic category is composed of the following groups: Mexican/Mexican American; Puerto Rican; and Latin, South, or Central American/Other Hispanic/Latino. Standard deviations for the Hispanic group prior to 1992 could not be calculated because only the aggregate data for the Hispanic subgroups were available. The mean scores for nonrespondents prior to 1992 are not known because they were not reported in *College-Bound Seniors*.

Table B10

Descriptive Statistics for SAT Mathematics by Hispanic Subgroups: 1987 to 2006

	Mexican/Mexican American			*Puerto Rican*			*Latin, South, or Central American/Other Hispanic/Latino*		
Year	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*
1987	20,512	455	98	10,173	432	103	18,633	462	107
1988	22,722	460	95	11,497	434	101	20,213	463	105
1989	25,207	462	94	11,415	438	101	21,242	466	104
1990	26,073	460	96	11,400	437	101	23,608	464	105
1991	28,602	459	96	12,065	439	101	25,584	462	105
1992	30,336	457	98	12,091	438	103	26,766	463	106
1993	32,355	459	99	12,645	440	103	28,420	463	108
1994	35,397	458	98	13,036	442	104	29,395	464	106
1995	36,323	458	98	13,056	444	102	30,713	468	106
1996	36,689	459	99	13,103	445	101	32,193	466	107
1997	39,737	458	99	13,208	447	103	33,123	468	107
1998	41,028	460	98	13,635	447	101	35,749	466	106
1999	43,160	456	100	13,986	448	102	37,521	464	107
2000	44,921	460	100	14,147	451	102	38,804	467	106
2001	46,849	458	100	14,074	451	103	40,249	465	106
2002	48,255	457	101	14,273	451	104	41,627	464	107
2003	50,375	457	101	14,569	453	102	42,548	464	107
2004	57,739	458	98	16,449	452	102	48,192	465	105
2005	66,968	463	99	19,402	457	102	57,826	469	105
2006	64,019	465	100	19,008	456	104	68,734	463	106

Note: For 1987 to 1995, individual student scores were converted to the recentered scale and then the mean was recomputed. From 1996 to 1999, nearly all students received scores on the recentered scale. Any score on the original scale was converted to the recentered scale prior to computing the mean. From 2000 to 2006, all scores were reported on the recentered scale.

Table B11

Percentages of Students Taking the SAT Subject Test in Writing by Racial/Ethnic Subgroups: 1996 to 2005

Year	*African American*	*American Indian/ Alaskan Native*	*Asian American/ Pacific Islander*	*Hispanic*	*White*	*Other*	*No Response*
1996	5	1	17	7	57	4	10
1997	4	1	18	7	56	4	11
1998	4	1	18	7	54	4	12
1999	4	0	18	7	53	4	14
2000	4	0	17	7	50	4	16
2001	4	0	18	8	49	4	17
2002	4	0	17	8	47	4	20
2003	4	0	16	8	42	3	26
2004	4	0	16	8	42	4	26
2005	4	0	19	9	47	4	14

Note: Although trend information is not reported until 1996, the SAT Subject Test in Writing was introduced in 1994 resulting in a cohort of scores reported in 1995, but the scores were subsequently rescaled and 1996 is the first year reported. The Hispanic category is composed of the following groups: Mexican/Mexican American; Puerto Rican; and Latin, South, or Central American/Other Hispanic/Latino. Percentage values may not add to 100 percent due to rounding.

Table B12

Descriptive Statistics for SAT Subject Test in Writing by Racial/Ethnic Subgroups: 1996 to 2005

	N	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*
Year	*African American*			*American Indian/Alaskan Native*			*Asian American/Pacific Islander*			*Hispanic*		
1996	9,172	510	96	1,155	537	94	34,313	537	110	14,413	508	98
1997	9,044	509	98	1,140	534	95	35,950	541	110	14,636	509	99
1998	9,182	512	99	1,132	540	98	37,492	540	112	15,023	505	102
1999	9,041	532	102	967	564	98	38,077	564	115	15,208	529	103
2000	9,201	536	102	914	567	100	37,696	568	113	15,443	529	104
2001	9,505	534	102	874	567	104	40,888	571	112	17,319	526	103
2002	9,299	538	101	886	570	102	40,100	574	114	17,434	525	104
2003	9,342	534	101	866	565	107	39,069	576	114	18,406	518	105
2004	9,400	541	102	889	577	103	37,751	581	116	19,121	524	102
2005	10,690	539	101	1,023	582	103	46,547	584	113	22,629	526	103
	White			*Other*			*No Response*					
1996	112,770	587	93	7,496	566	103	19,062	570	107			
1997	113,001	588	95	7,915	569	102	21,776	576	108			
1998	110,863	590	95	8,796	570	106	24,482	578	110			
1999	112,744	615	97	9,228	593	107	28,985	601	111			
2000	109,142	618	95	9,124	593	106	35,659	602	109			
2001	110,592	618	95	8,982	592	106	37,564	603	108			
2002	108,370	624	95	8,908	595	108	46,261	608	107			
2003	102,909	619	96	8,520	591	108	64,530	602	107			
2004	97,313	626	95	8,634	591	108	59,761	620	105			
2005	114,104	627	94	10,713	590	107	34,658	639	104			

Note: Although trend information is not reported until 1996, the SAT Subject Test in Writing was introduced in 1994 resulting in a cohort of scores reported in 1995, but the scores were subsequently rescaled and 1996 is the first year reported. There was another adjustment to the scale in May 1998 that resulted in an average increase of approximately 23 points in 1999. The Hispanic category is composed of the following groups: Mexican/Mexican American; Puerto Rican; and Latin, South, or Central American/Other Hispanic/Latino.

Table B13

Descriptive Statistics for SAT Verbal/Critical Reading by First Language Subgroups: 1987 to 2006

Year	N	%	Mean	SD	N	%	Mean	SD
	English Only				*English and Another Language*			
1987	866,164	80	516	104	84,569	8	480	110
1988	912,038	80	514	103	90,698	8	480	110
1989	859,884	79	513	104	91,303	8	481	112
1990	802,784	78	509	105	81,668	8	480	114
1991	804,108	78	508	105	80,767	8	478	114
1992	801,360	77	509	105	80,006	8	477	116
1993	798,882	76	510	106	81,703	8	477	116
1994	792,494	75	510	105	86,315	8	477	116
1995	810,311	76	514	107	87,416	8	479	115
1996	822,776	76	515	105	90,568	8	483	113
1997	845,543	75	515	106	96,917	9	483	114
1998	867,090	74	516	106	102,083	9	483	114
1999	885,836	73	516	106	106,128	9	484	114
2000	883,230	70	516	106	108,415	9	484	114
2001	861,006	67	517	106	108,875	9	484	114
2002	845,352	64	515	106	109,289	8	481	114
2003	827,018	59	515	106	119,821	9	481	114
2004	938,638	66	515	106	153,885	11	483	112
2005	1,071,262	73	519	106	177,449	12	486	114
2006	1,062,751	73	515	107	191,741	13	485	114
	Another Language				*No Response*			
1987	51,638	5	436	121	78,055	7	--	--
1988	55,688	5	438	122	75,940	7	--	--
1989	58,117	5	440	121	78,919	7	--	--
1990	66,031	6	441	123	75,040	7	--	--
1991	73,268	7	443	123	74,542	7	--	--
1992	75,564	7	443	126	77,201	7	475	123
1993	77,650	7	444	125	86,230	8	476	123
1994	80,615	8	442	124	90,962	9	475	122
1995	81,276	8	447	124	88,990	8	479	124
1996	83,356	8	452	121	88,025	8	482	121
1997	86,062	8	452	120	98,499	9	485	122
1998	89,826	8	453	119	113,780	10	485	121
1999	92,129	8	454	120	136,037	11	487	121
2000	92,958	7	455	119	175,675	14	490	118
2001	94,194	7	455	120	212,245	17	494	116
2002	92,753	7	453	120	280,437	21	500	115
2003	90,535	6	456	120	368,950	26	511	112
2004	102,961	7	457	118	223,523	16	517	120
2005	123,057	8	462	120	103,855	7	490	141
2006	122,658	8	467	120	88,594	6	455	134

Note: For 1987 to 1995, individual student scores were converted to the recentered scale and then the mean was recomputed. From 1996 to 1999, nearly all students received scores on the recentered scale. Any score on the original scale was converted to the recentered scale prior to computing the mean. From 2000 to 2006, all scores were reported on the recentered scale. The mean scores for nonrespondents prior to 1992 are not known because they were not reported in *College-Bound Seniors*. Percentage values may not add to 100 percent due to rounding.

Table B14

Descriptive Statistics for SAT Mathematics by First Language Subgroups: 1987 to 2006

	N	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*
Year	*English Only*				*English and Another Language*			
1987	866,164	80	504	106	84,569	8	480	111
1988	912,038	80	504	103	90,698	8	482	109
1989	859,884	79	505	104	91,303	8	486	111
1990	802,784	78	504	106	81,668	8	488	114
1991	804,108	78	502	106	80,767	8	487	115
1992	801,360	77	504	105	80,006	8	487	117
1993	798,882	76	505	107	81,703	8	489	118
1994	792,494	75	507	106	86,315	8	489	117
1995	810,311	76	509	108	87,416	8	492	118
1996	822,776	76	511	109	90,568	8	495	119
1997	845,543	75	513	108	96,917	9	498	120
1998	867,090	74	515	108	102,083	9	501	120
1999	885,836	73	514	110	106,128	9	500	122
2000	883,230	70	516	109	108,415	9	504	122
2001	861,006	67	517	109	108,875	9	504	123
2002	845,352	64	517	110	109,289	8	504	124
2003	827,018	59	517	111	119,821	9	504	125
2004	938,638	66	516	108	153,885	11	507	122
2005	1,071,262	73	521	109	177,449	12	513	122
2006	1,062,751	73	521	110	191,741	13	511	123
	Another Language				*No Response*			
1987	51,638	5	506	120	78,055	7	--	--
1988	55,688	5	507	118	75,940	7	--	--
1989	58,117	5	512	120	78,919	7	--	--
1990	66,031	6	511	121	75,040	7	--	--
1991	73,268	7	511	122	74,542	7	--	--
1992	75,564	7	513	124	77,201	7	483	120
1993	77,650	7	513	126	86,230	8	486	122
1994	80,615	8	511	125	90,962	9	486	120
1995	81,276	8	512	124	88,990	8	489	120
1996	83,356	8	513	125	88,025	8	492	121
1997	86,062	8	515	125	98,499	9	500	122
1998	89,826	8	516	126	113,780	10	500	121
1999	92,129	8	513	128	136,037	11	502	121
2000	92,958	7	517	127	175,675	14	507	119
2001	94,194	7	516	129	212,245	17	508	117
2002	92,753	7	515	130	280,437	21	517	117
2003	90,535	6	515	132	368,950	26	528	116
2004	102,961	7	515	130	223,523	16	536	123
2005	123,057	8	521	129	103,855	7	518	140
2006	122,658	8	523	131	88,594	6	486	132

Note: For 1987 to 1995, individual student scores were converted to the recentered scale and then the mean was recomputed. From 1996 to 1999, nearly all students received scores on the recentered scale. Any score on the original scale was converted to the recentered scale prior to computing the mean. From 2000 to 2006, all scores were reported on the recentered scale. The mean scores for nonrespondents prior to 1992 are not known because they were not reported in *College-Bound Seniors*. Percentage values may not add to 100 percent due to rounding.

Table B15

Descriptive Statistics for SAT Subject Test in Writing by First Language Subgroups: 1996 to 2005

	N	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*
Year	*English Only*				*English and Another Language*			
1996	133,729	67	580	96	22,199	11	546	106
1997	134,466	66	581	98	23,735	12	549	107
1998	134,030	65	583	99	24,924	12	547	109
1999	136,024	63	607	100	25,888	12	571	111
2000	132,970	61	610	99	26,119	12	573	110
2001	134,047	59	610	99	27,515	12	573	110
2002	128,504	56	615	99	26,090	11	573	112
2003	122,375	50	610	101	27,011	11	569	114
2004	125,653	54	616	99	31,717	14	573	112
2005	147,995	62	618	99	38,165	16	576	111
	Another Language				*No Response*			
1996	25,014	13	510	109	17,439	9	566	107
1997	25,263	12	517	110	19,998	10	573	108
1998	25,997	13	514	112	22,019	11	574	110
1999	26,161	12	539	115	26,177	12	597	111
2000	25,905	12	544	114	32,185	15	598	109
2001	27,560	12	546	115	36,602	16	597	108
2002	26,313	11	545	117	50,351	22	602	107
2003	25,445	10	547	118	68,811	28	599	106
2004	25,785	11	551	119	49,714	21	620	105
2005	31,374	13	556	116	22,830	9	643	106

Note: Although trend information is not reported until 1996, the SAT Subject Test in Writing was introduced in 1994 resulting in a cohort of scores reported in 1995, but the scores were subsequently rescaled and 1996 is the first year reported. There was another adjustment to the scale in May 1998 that resulted in an average increase of approximately 23 points in 1999. Percentage values may not add to 100 percent due to rounding.

Table B16

Descriptive Statistics for SAT Verbal/Critical Reading by Best Language Subgroups: 1995 to 2006

	N	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*
Year	*English Only*				*English and Another Language About the Same*			
1995	866,009	81	514	108	64,910	6	453	114
1996	879,139	81	516	106	66,310	6	455	112
1997	906,652	80	515	107	69,658	6	454	112
1998	930,789	79	516	106	73,570	6	453	111
1999	953,686	78	516	106	76,612	6	454	112
2000	956,317	76	516	106	78,049	6	455	111
2001	964,768	76	517	106	85,778	7	458	112
2002	977,565	74	515	106	92,374	7	460	114
2003	959,691	68	516	107	97,891	7	464	114
2004	1,054,399	74	515	107	104,315	7	462	113
2005	1,210,581	82	518	107	118,328	8	463	113
2006	1,220,257	83	514	108	114,176	8	463	112
	Another Language				*No Response*			
1995	26,434	2	384	111	110,640	10	476	121
1996	26,006	2	389	106	113,270	10	480	118
1997	25,794	2	392	105	124,917	11	483	119
1998	26,843	2	396	105	141,577	12	483	119
1999	26,813	2	397	108	163,019	13	484	119
2000	26,666	2	399	107	199,246	16	487	117
2001	27,482	2	398	106	198,292	16	486	117
2002	27,360	2	397	106	230,532	17	489	116
2003	27,252	2	404	108	321,490	23	502	113
2004	32,567	2	408	107	227,726	16	510	120
2005	38,938	3	411	111	107,776	7	481	139
2006	39,091	3	414	109	92,220	6	453	132

Note: For 1995, individual student scores were converted to the recentered scale and then the mean was recomputed. From 1996 to 1999, nearly all students received scores on the recentered scale. Any score on the original scale was converted to the recentered scale prior to computing the mean. From 2000 to 2006, all scores were reported on the recentered scale. A question on best language was added to the SAT Questionnaire in 1995. Percentage values may not add to 100 percent due to rounding.

Table B17

Descriptive Statistics for SAT Mathematics by Best Language Subgroups: 1995 to 2006

Year	N	%	Mean	SD	N	%	Mean	SD
	English Only				*English and Another Language About the Same*			
1995	866,009	81	511	109	64,910	6	481	118
1996	879,139	81	513	110	66,310	6	481	119
1997	906,652	80	515	109	69,658	6	484	120
1998	930,789	79	517	109	73,570	6	485	120
1999	953,686	78	516	111	76,612	6	483	122
2000	956,317	76	518	110	78,049	6	489	122
2001	964,768	76	519	110	85,778	7	492	123
2002	977,565	74	520	111	92,374	7	496	126
2003	959,691	68	521	111	97,891	7	499	127
2004	1,054,399	74	518	110	104,315	7	496	126
2005	1,210,581	82	522	110	118,328	8	499	124
2006	1,220,257	83	522	111	114,176	8	495	124
	Another Language				*No Response*			
1995	26,434	2	514	129	110,640	10	482	118
1996	26,006	2	513	128	113,270	10	484	118
1997	25,794	2	517	129	124,917	11	492	120
1998	26,843	2	520	131	141,577	12	492	119
1999	26,813	2	518	135	163,019	13	495	120
2000	26,666	2	522	134	199,246	16	501	118
2001	27,482	2	520	136	198,292	16	499	118
2002	27,360	2	524	138	230,532	17	505	119
2003	27,252	2	527	143	321,490	23	518	118
2004	32,567	2	527	140	227,726	16	528	123
2005	38,938	3	534	139	107,776	7	507	140
2006	39,091	3	532	139	92,220	6	483	131

Note: For 1995, individual student scores were converted to the recentered scale and then the mean was recomputed. From 1996 to 1999, nearly all students received scores on the recentered scale. Any score on the original scale was converted to the recentered scale prior to computing the mean. From 2000 to 2006, all scores were reported on the recentered scale. A question on best language was added to the SAT Questionnaire in 1995. Percentage values may not add to 100 percent due to rounding.

Table B18

Descriptive Statistics for SAT Subject Test in Writing by Best Language Subgroups: 1996 to 2005

	N	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*
Year	*English Only*				*English and Another Language About the Same*			
1996	157,545	79	576	98	15,555	8	509	106
1997	158,974	78	577	99	16,505	8	515	108
1998	159,926	77	578	101	16,892	8	512	109
1999	162,530	76	603	102	17,341	8	537	112
2000	159,605	73	605	101	17,713	8	541	112
2001	167,175	74	605	101	20,848	9	544	111
2002	166,391	72	611	102	22,500	10	548	113
2003	161,738	66	606	103	23,943	10	547	114
2004	158,919	68	612	103	22,321	10	550	115
2005	186,815	78	613	102	25,128	10	550	113
	Another Language				*No Response*			
1996	5,656	3	454	96	19,625	10	564	106
1997	5,567	3	466	101	22,416	11	570	108
1998	5,822	3	465	100	24,330	12	572	109
1999	5,856	3	487	106	28,523	13	595	110
2000	5,649	3	495	107	34,212	16	596	109
2001	6,072	3	497	105	31,629	14	596	109
2002	5,901	3	491	103	36,466	16	600	108
2003	5,747	2	498	108	52,214	21	597	107
2004	6,159	3	502	109	45,470	20	616	106
2005	7,531	3	514	108	20,890	9	640	107

Note: Although trend information is not reported until 1996, the SAT Subject Test in Writing was introduced in 1994 resulting in a cohort of scores reported in 1995, but the scores were subsequently rescaled and 1996 is the first year reported. There was another adjustment to the scale in May 1998 that resulted in an average increase of approximately 23 points in 1999. Percentage values may not add to 100 percent due to rounding.

Table B19

Descriptive Statistics for SAT Verbal/Critical Reading by Income Subgroups: 1992 to 2006

	N	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*
Year	*<$30,000*				*$30,000–50,000*				*$50,000–70,000*			
1992	272,011	26	462	110	269,703	26	501	102	164,322	16	521	101
1993	273,537	26	461	110	261,435	25	501	102	164,036	16	521	102
1994	259,123	25	459	110	248,832	24	499	102	167,343	16	518	101
1995	248,139	23	462	112	242,897	23	502	104	175,470	16	520	103
1996	239,894	22	463	109	238,716	22	502	102	181,025	17	520	101
1997	237,140	21	460	110	236,895	21	501	103	185,854	16	518	102
1998	234,883	20	458	109	232,316	20	500	102	189,542	16	517	101
1999	226,816	19	456	109	225,661	18	499	103	190,664	16	516	101
2000	212,407	17	454	108	211,476	17	496	102	183,397	15	514	101
2001	192,000	15	450	108	188,329	15	493	102	166,640	13	512	100
2002	178,855	13	443	107	167,677	13	487	103	148,946	11	508	101
2003	169,203	12	443	107	152,094	11	487	103	132,994	9	507	101
2004	188,200	13	444	107	164,378	12	485	103	145,555	10	504	101
2005	212,344	14	447	108	185,994	13	487	103	173,453	12	508	101
2006	183,317	13	449	106	167,284	11	485	103	160,992	11	503	101
	$70,000–100,000				*>$100,000*				*No Response*			
1992	176,081	17	546	102	--	--	--	--	152,014	15	488	119
1993	180,901	17	548	103	--	--	--	--	164,556	16	489	120
1994	200,262	19	545	104	--	--	--	--	174,826	17	488	120
1995	130,133	12	538	104	90,261	9	557	104	181,093	17	493	121
1996	138,532	13	537	101	100,429	9	560	102	185,378	17	496	119
1997	149,819	13	535	101	111,252	10	559	103	206,051	18	498	119
1998	160,764	14	534	102	122,383	10	559	102	232,891	20	499	119
1999	170,851	14	533	101	136,658	11	559	102	269,478	22	499	119
2000	173,389	14	531	100	146,319	12	558	101	333,289	26	503	117
2001	167,954	13	529	101	152,191	12	557	101	409,206	32	506	115
2002	157,080	12	525	101	149,004	11	555	101	526,269	40	510	113
2003	145,341	10	524	101	141,885	10	555	101	664,807	47	514	111
2004	165,362	12	522	101	161,809	11	553	102	593,703	42	519	113
2005	208,594	14	524	101	207,199	14	554	101	488,039	33	517	119
2006	209,312	14	518	101	225,869	15	549	103	518,970	35	503	119

Note: For 1992 to 1995, individual student scores were converted to the recentered scale and then the mean was recomputed. From 1996 to 1999, nearly all students received scores on the recentered scale. Any score on the original scale was converted to the recentered scale prior to computing the mean. From 2000 to 2006, all scores were reported on the recentered scale. Income categories have been collapsed for reporting purposes. The category of >$100,000 was added to the SAT Questionnaire in 1994. Percentage values may not add to 100 percent due to rounding.

Table B20

Descriptive Statistics for SAT Mathematics by Income Subgroups: 1992 to 2006

	N	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*
Year	*<$30,000*				*$30,000–50,000*				*$50,000–70,000*			
1992	272,011	26	469	108	269,703	26	499	102	164,322	16	519	102
1993	273,537	26	468	110	261,435	25	501	104	164,036	16	520	103
1994	259,123	25	467	109	248,832	24	500	103	167,343	16	519	102
1995	248,139	23	469	109	242,897	23	501	104	175,470	16	520	105
1996	239,894	22	468	109	238,716	22	501	105	181,025	17	520	105
1997	237,140	21	468	110	236,895	21	502	104	185,854	16	521	104
1998	234,883	20	468	110	232,316	20	502	104	189,542	16	521	104
1999	226,816	19	464	111	225,661	18	499	106	190,664	16	519	104
2000	212,407	17	466	111	211,476	17	498	105	183,397	15	518	104
2001	192,000	15	462	111	188,329	15	495	105	166,640	13	515	104
2002	178,855	13	458	112	167,677	13	492	106	148,946	11	513	104
2003	169,203	12	457	113	152,094	11	490	107	132,994	9	511	105
2004	188,200	13	460	113	164,378	12	488	105	145,555	10	507	103
2005	212,344	14	466	114	185,994	13	493	105	173,453	12	512	103
2006	183,317	13	467	114	167,284	11	494	107	160,992	11	512	104
	$70,000–100,000				*>$100,000*				*No Response*			
1992	176,081	17	548	103	--	--	--	--	152,014	15	491	116
1993	180,901	17	552	104	--	--	--	--	164,556	16	494	119
1994	200,262	19	549	103	--	--	--	--	174,826	17	494	117
1995	130,133	12	540	106	90,261	9	566	104	181,093	17	499	118
1996	138,532	13	539	106	100,429	9	569	105	185,378	17	500	118
1997	149,819	13	540	104	111,252	10	571	104	206,051	18	506	119
1998	160,764	14	540	104	122,383	10	572	104	232,891	20	508	119
1999	170,851	14	538	104	136,658	11	571	104	269,478	22	509	120
2000	173,389	14	537	104	146,319	12	571	104	333,289	26	515	118
2001	167,954	13	534	104	152,191	12	569	104	409,206	32	518	117
2002	157,080	12	532	105	149,004	11	568	105	526,269	40	524	116
2003	145,341	10	530	105	141,885	10	568	106	664,807	47	529	115
2004	165,362	12	525	103	161,809	11	562	105	593,703	42	533	116
2005	208,594	14	529	103	207,199	14	565	104	488,039	33	532	120
2006	209,312	14	529	104	225,869	15	564	105	518,970	35	521	120

Note: For 1992 to 1995, individual student scores were converted to the recentered scale and then the mean was recomputed. From 1996 to 1999, nearly all students received scores on the recentered scale. Any score on the original scale was converted to the recentered scale prior to computing the mean. From 2000 to 2006, all scores were reported on the recentered scale. Income categories have been collapsed for reporting purposes. The category of >$100,000 was added to the SAT Questionnaire in 1994. Percentage values may not add to 100 percent due to rounding.

Table B21

Descriptive Statistics for SAT Subject Test in Writing by Income Subgroups: 1996 to 2005

	N	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*
Year	*<$30,000*				*$30,000–50,000*				*$50,000–70,000*			
1996	29,616	15	513	106	31,298	16	554	100	29,977	15	571	97
1997	28,855	14	513	106	30,236	15	553	100	29,351	14	570	98
1998	28,279	14	508	107	28,657	14	553	103	28,219	14	569	100
1999	27,288	13	528	110	27,225	13	574	104	27,661	13	593	101
2000	25,201	12	528	109	24,410	11	573	104	24,953	11	594	101
2001	23,659	10	524	109	21,705	10	570	105	22,378	10	592	101
2002	20,805	9	520	108	18,236	8	568	107	18,517	8	595	102
2003	19,204	8	516	109	16,490	7	562	108	15,732	6	589	104
2004	18,975	8	523	109	15,739	7	565	108	15,017	6	591	104
2005	23,417	10	528	109	19,074	8	562	107	18,879	8	590	102
	$70,000–100,000				*>$100,000*				*No Response*			
1996	31,859	16	581	95	37,944	19	594	94	37,153	19	574	104
1997	33,159	16	581	97	40,646	20	597	97	41,208	20	578	105
1998	33,340	16	582	99	43,132	21	598	97	45,343	22	580	106
1999	33,767	16	606	99	46,234	22	623	98	52,075	24	603	108
2000	32,008	15	608	98	46,993	22	626	96	63,614	29	604	106
2001	30,411	13	606	99	48,177	21	624	96	79,394	35	603	105
2002	26,066	11	610	99	44,882	19	630	96	102,752	44	607	105
2003	22,819	9	603	100	41,290	17	625	97	128,107	53	602	106
2004	22,240	10	610	101	40,628	17	632	96	120,270	52	613	105
2005	28,605	12	607	100	51,536	21	631	65	98,853	41	621	105

Note: Although trend information is not reported until 1996, the SAT Subject Test in Writing was introduced in 1994 resulting in a cohort of scores reported in 1995, but the scores were subsequently rescaled and 1996 is the first year reported. There was another adjustment to the scale in May 1998 that resulted in an average increase of approximately 23 points in 1999. Income categories have been collapsed for reporting purposes. Percentage values may not add to 100 percent due to rounding.

Table B22

Standardized Differences Across Measures by Gender (Female Minus Male): 1987 to 2006

	HS Grades	*Reading*			*Writing*			*Math*		
Year	*HSGPA*	*SAT Verbal/ Critical Reading*	*ACT Reading*	*NAEP Reading*	*SAT Subject Test in Writing*	*ACT English*	*NAEP Writing*	*SAT Math*	*ACT Math*	*NAEP Math*
1987	--	-0.09	--	--	--	--	--	-0.39	--	--
1988	--	-0.12	--	--	--	--	--	-0.36	--	--
1989	--	-0.11	--	--	--	--	--	-0.38	--	--
1990	--	-0.08	--	--	--	--	--	-0.35	--	-0.17
1991	--	-0.07	--	--	--	--	--	-0.35	--	--
1992	0.22	-0.07	--	0.30	--	--	--	-0.34	--	-0.09
1993	0.21	-0.06	--	--	--	--	--	-0.36	--	--
1994	0.23	-0.04	--	0.38	--	--	--	-0.33	--	--
1995	0.24	-0.03	--	--	--	--	--	-0.31	--	--
1996	0.26	-0.04	--	--	0.08	--	--	-0.31	--	-0.06
1997	0.24	-0.04	0.05	--	0.10	0.15	--	-0.32	-0.24	--
1998	0.25	-0.06	0.08	0.42	0.08	0.17	0.54	-0.31	-0.25	--
1999	0.25	-0.06	0.08	--	0.07	0.16	--	-0.32	-0.24	--
2000	0.23	-0.03	0.05	--	0.08	0.16	--	-0.31	-0.24	-0.11
2001	0.22	-0.06	0.07	--	0.07	0.14	--	-0.31	-0.24	--
2002	0.22	-0.05	0.07	0.43	0.05	0.16	0.60	-0.30	-0.22	--
2003	0.23	-0.08	0.07	--	0.04	0.16	--	-0.30	-0.22	--
2004	0.25	-0.07	0.07	--	0.04	0.15	--	-0.32	-0.22	--
2005	0.25	-0.07	0.08	--	0.04	0.14	--	-0.30	-0.22	--
2006	0.25	-0.03	--	--	--	--	--	-0.30	--	--

Note: HSGPA data from the SAT Questionnaire were available for analysis by subgroup after 1991. Standardized differences for the ACT are reported for the years for which data were available on www.act.org. Although trend information is not reported until 1996, the SAT Subject Test in Writing was introduced in 1994 resulting in a cohort of scores reported in 1995, but the scores were subsequently rescaled and 1996 is the first year reported. There was another adjustment to the scale in May 1998 that resulted in an average increase of approximately 23 points in 1999. In grade 12, NAEP reading and math assessments are given at least as often as they had been in the past, or every four years on a nationally representative basis. The NAEP writing assessment is administered to grade 12 periodically. Standardized differences for NAEP are reported for the years for which data were available on http://nces.ed.gov/nationsreportcard/.

Table B23

Standardized Differences Across Measures by Ethnicity, 1987 to 2006: African Americans Minus Total

	HS Grades	*Reading*			*Writing*			*Mathematics*		
Year	*HSGPA*	*SAT Verbal/ Critical Reading*	*ACT Reading*	*NAEP Reading*	*SAT Subject Test in Writing*	*ACT English*	*NAEP Writing*	*SAT Math*	*ACT Math*	*NAEP Math*
1987	--	-0.72	--	--	--	--	--	-0.83	--	--
1988	--	-0.71	--	--	--	--	--	-0.78	--	--
1989	--	-0.70	--	--	--	--	--	-0.76	--	--
1990	--	-0.65	--	--	--	--	--	-0.76	--	-0.72
1991	--	-0.65	--	--	--	--	--	-0.74	--	--
1992	-0.45	-0.65	--	-0.58	--	--	--	-0.75	--	-0.68
1993	-0.47	-0.64	--	--	--	--	--	-0.74	--	--
1994	-0.47	-0.64	--	-0.59	--	--	--	-0.75	--	--
1995	-0.47	-0.65	--	--	--	--	--	-0.75	--	--
1996	-0.48	-0.65	--	--	-0.54	--	--	-0.77	--	-0.75
1997	-0.48	-0.64	-0.69	--	-0.57	-0.72	--	-0.79	-0.74	--
1998	-0.51	-0.64	-0.70	-0.58	-0.54	-0.74	-0.46	-0.77	-0.76	--
1999	-0.51	-0.64	-0.72	--	-0.57	-0.75	--	-0.78	-0.76	--
2000	-0.51	-0.64	-0.72	--	-0.57	-0.75	--	-0.78	-0.78	-0.77
2001	-0.52	-0.66	-0.73	-0.54	-0.58	-0.77	--	-0.78	-0.78	--
2002	-0.54	-0.67	-0.70	--	-0.58	-0.69	-0.45	-0.78	-0.78	--
2003	-0.52	-0.68	-0.69	--	-0.57	-0.71	--	-0.81	-0.76	--
2004	-0.50	-0.70	-0.70	--	-0.59	-0.69	--	-0.80	-0.76	--
2005	-0.51	-0.66	-0.72	--	-0.62	-0.71	--	-0.77	-0.78	--
2006	-0.51	-0.61	--	--	--	--	--	-0.77	--	--

Note: HSGPA data from the SAT Questionnaire were available for analysis by subgroup after 1991. Standardized differences for the ACT are reported for the years for which data were available on www.act.org. Although trend information is not reported until 1996, the SAT Subject Test in Writing was introduced in 1994 resulting in a cohort of scores reported in 1995, but the scores were subsequently rescaled and 1996 is the first year reported. There was another adjustment to the scale in May 1998 that resulted in an average increase of approximately 23 points in 1999. In grade 12, NAEP reading and math assessments are given at least as often as they had been in the past, or every four years on a nationally representative basis. The NAEP writing assessment is administered to grade 12 periodically. Standardized differences for NAEP are reported for the years for which data were available on http://nces.ed.gov/nationsreportcard/.

Table B24

Standardized Differences Across Measures by Ethnicity, 1987 to 2006: American Indians/Alaskan Natives Minus Total

	HS Grades	*Reading*			*Writing*			*Mathematics*		
Year	*HSGPA*	*SAT Verbal/ Critical Reading*	*ACT Reading*	*NAEP Reading*	*SAT Subject Test in Writing*	*ACT English*	*NAEP Writing*	*SAT Math*	*ACT Math*	*NAEP Math*
1987	--	-0.33	--	--	--	--	--	-0.35	--	--
1988	--	-0.32	--	--	--	--	--	-0.33	--	--
1989	--	-0.39	--	--	--	--	--	-0.38	--	--
1990	--	-0.31	--	--	--	--	--	-0.31	--	--
1991	--	-0.26	--	--	--	--	--	-0.29	--	--
1992	-0.23	-0.25	--	--	--	--	--	-0.28	--	--
1993	-0.21	-0.21	--	--	--	--	--	-0.24	--	--
1994	-0.23	-0.23	--	--	--	--	--	-0.31	--	--
1995	-0.20	-0.22	--	--	--	--	--	-0.27	--	--
1996	-0.21	-0.20	--	--	-0.28	--	--	-0.28	--	-0.78
1997	-0.26	-0.27	-0.31	--	-0.33	-0.43	--	-0.32	-0.42	--
1998	-0.22	-0.23	-0.33	--	-0.27	-0.43	-0.60	-0.26	-0.43	--
1999	-0.18	-0.19	-0.35	--	-0.27	-0.44	--	-0.26	-0.44	--
2000	-0.17	-0.21	-0.33	--	-0.27	-0.45	--	-0.29	-0.44	-0.23
2001	-0.20	-0.23	-0.35	--	-0.26	-0.48	--	-0.31	-0.46	--
2002	-0.21	-0.23	-0.33	--	-0.28	-0.45	--	-0.29	-0.44	--
2003	-0.17	-0.24	-0.34	--	-0.29	-0.45	--	-0.32	-0.45	--
2004	-0.14	-0.22	-0.33	--	-0.25	-0.44	--	-0.26	-0.42	--
2005	-0.13	-0.17	-0.37	--	-0.21	-0.47	--	-0.23	-0.46	--
2006	-0.14	-0.14	--	--	--	--	--	-0.21	--	--

Note: HSGPA data from the SAT Questionnaire were available for analysis by subgroup after 1991. Standardized differences for the ACT are reported for the years for which data were available on www.act.org. Although trend information is not reported until 1996, the SAT Subject Test in Writing was introduced in 1994 resulting in a cohort of scores reported in 1995, but the scores were subsequently rescaled and 1996 is the first year reported. There was another adjustment to the scale in May 1998 that resulted in an average increase of approximately 23 points in 1999. In grade 12, NAEP reading and math assessments are given at least as often as they had been in the past, or every four years on a nationally representative basis. The NAEP writing assessment is administered to grade 12 periodically. Standardized differences for NAEP are reported for the years for which data were available on http://nces.ed.gov/nationsreportcard/. Data were not available for NAEP Reading because the sample size was not large enough.

Table B25

Standardized Differences Across Measures by Ethnicity, 1987 to 2006: Asian Americans/Pacific Islanders Minus Total

	HS Grades	*Reading*			*Writing*			*Mathematics*		
Year	*HSGPA*	*SAT Verbal/ Critical Reading*	*ACT Reading*	*NAEP Reading*	*SAT Subject Test in Writing*	*ACT English*	*NAEP Writing*	*SAT Math*	*ACT Math*	*NAEP Math*
1987	--	-0.26	--	--	--	--	--	0.37	--	--
1988	--	-0.21	--	--	--	--	--	0.38	--	--
1989	--	-0.19	--	--	--	--	--	0.40	--	--
1990	--	-0.15	--	--	--	--	--	0.42	--	0.47
1991	--	-0.13	--	--	--	--	--	0.44	--	--
1992	0.29	-0.12	--	0.00	--	--	--	0.46	--	0.50
1993	0.29	-0.10	--	--	--	--	--	0.45	--	--
1994	0.27	-0.09	--	-0.19	--	--	--	0.45	--	--
1995	0.26	-0.12	--	--	--	--	--	0.44	--	--
1996	0.24	-0.08	--	--	-0.28	--	--	0.45	--	0.47
1997	0.21	-0.08	-0.02	--	-0.26	0.02	--	0.44	0.54	--
1998	0.18	-0.06	-0.02	-0.05	-0.27	0.02	0.06	0.45	0.51	--
1999	0.18	-0.06	-0.03	--	-0.27	0.00	--	0.43	0.48	--
2000	0.20	-0.05	-0.02	--	-0.26	0.00	--	0.45	0.50	0.51
2001	0.20	-0.05	-0.03	-0.03	-0.23	0.04	--	0.46	0.48	--
2002	0.21	-0.03	0.02	--	-0.24	0.05	0.10	0.46	0.46	--
2003	0.19	0.01	0.02	--	-0.19	0.07	--	0.49	0.45	--
2004	0.14	-0.01	0.03	--	-0.21	0.10	--	0.52	0.46	--
2005	0.14	0.03	0.08	--	-0.20	0.15	--	0.52	0.48	--
2006	0.17	0.06	--	--	--	--	--	0.52	--	--

Note: HSGPA data from the SAT Questionnaire were available for analysis by subgroup after 1991. Standardized differences for the ACT are reported for the years for which data were available on www.act.org. Although trend information is not reported until 1996, the SAT Subject Test in Writing was introduced in 1994 resulting in a cohort of scores reported in 1995, but the scores were subsequently rescaled and 1996 is the first year reported. There was another adjustment to the scale in May 1998 that resulted in an average increase of approximately 23 points in 1999. In grade 12, NAEP reading and math assessments are given at least as often as they had been in the past, or every four years on a nationally representative basis. The NAEP writing assessment is administered to grade 12 periodically. Standardized differences for NAEP are reported for the years for which data were available on http://nces.ed.gov/nationsreportcard/.

Table B26

Standardized Differences Across Measures by Ethnicity, 1987 to 2006: Hispanics Minus Total

	HS Grades	*Reading*			*Writing*			*Mathematics*		
Year	*HSGPA*	*SAT Verbal/ Critical Reading*	*ACT Reading*	*NAEP Reading*	*SAT Subject Test in Writing*	*ACT English*	*NAEP Writing*	*SAT Math*	*ACT Math*	*NAEP Math*
1987	--	-0.47	--	--	--	--	--	-0.45	--	--
1988	--	-0.47	--	--	--	--	--	-0.43	--	--
1989	--	-0.43	--	--	--	--	--	-0.40	--	--
1990	--	-0.42	--	--	--	--	--	-0.41	--	-0.50
1991	--	-0.42	--	--	--	--	--	-0.40	--	--
1992	-0.09	-0.44	--	-0.42	--	--	--	-0.41	--	-0.44
1993	-0.12	-0.42	--	--	--	--	--	-0.41	--	--
1994	-0.14	-0.42	--	-0.46	--	--	--	-0.42	--	--
1995	-0.12	-0.43	--	--	--	--	--	-0.42	--	--
1996	-0.14	-0.43	--	--	-0.56	--	--	-0.43	--	-0.53
1997	-0.15	-0.43	-0.38	--	-0.57	-0.43	--	-0.46	-0.32	--
1998	-0.15	-0.44	-0.38	-0.42	-0.60	-0.46	-0.43	-0.46	-0.36	--
1999	-0.17	-0.43	-0.39	--	-0.60	-0.47	--	-0.46	-0.35	--
2000	-0.17	-0.43	-0.38	--	-0.63	-0.47	--	-0.47	-0.35	-0.51
2001	-0.17	-0.45	-0.40	-0.41	-0.65	-0.48	--	-0.48	-0.36	--
2002	-0.19	-0.47	-0.42	--	-0.70	-0.49	-0.35	-0.50	-0.41	--
2003	-0.19	-0.49	-0.40	--	-0.72	-0.49	--	-0.52	-0.41	--
2004	-0.17	-0.46	-0.43	--	-0.75	-0.49	--	-0.51	-0.41	--
2005	-0.19	-0.44	-0.43	--	-0.74	-0.48	--	-0.49	-0.39	--
2006	-0.21	-0.41	--	--	--	--	--	-0.48	--	--

Note: For SAT data, the Hispanic category is composed of the following groups: Mexican/Mexican American; Puerto Rican; and Latin, South, or Central American/Other Hispanic/Latino. For ACT data, the Hispanic category is composed of Mexican American; and Puerto Rican/ Hispanic for years 2000 to 2006 and Mexican American; and Puerto Rican/Cuban for 1997 to 1999. For NAEP data, the Hispanic category is not specified by more detailed subgroups.

HSGPA data from the SAT Questionnaire were available for analysis by subgroup after 1991. Standardized differences for the ACT are reported for the years for which data were available on www.act.org. Although trend information is not reported until 1996, the SAT Subject Test in Writing was introduced in 1994 resulting in a cohort of scores reported in 1995, but the scores were subsequently rescaled and 1996 is the first year reported. There was another adjustment to the scale in May 1998 that resulted in an average increase of approximately 23 points in 1999. In grade 12, NAEP reading and math assessments are given at least as often as they had been in the past, or every four years on a nationally representative basis. The NAEP writing assessment is administered to grade 12 periodically. Standardized differences for NAEP are reported for the years for which data were available on http://nces.ed.gov/nationsreportcard/.

Table B27

Standardized Differences Across Measures by Ethnicity, 1987 to 2006: Whites Minus Total

	HS Grades	*Reading*			*Writing*			*Mathematics*		
Year	*HSGPA*	*SAT Verbal/ Critical Reading*	*ACT Reading*	*NAEP Reading*	*SAT Subject Test in Writing*	*ACT English*	*NAEP Writing*	*SAT Math*	*ACT Math*	*NAEP Math*
1987	--	0.16	--	--	--	--	--	0.12	--	--
1988	--	0.16	--	--	--	--	--	0.12	--	--
1989	--	0.17	--	--	--	--	--	0.12	--	--
1990	--	0.17	--	--	--	--	--	0.13	--	0.19
1991	--	0.17	--	--	--	--	--	0.12	--	--
1992	0.06	0.17	--	0.18	--	--	--	0.13	--	0.21
1993	0.06	0.18	--	--	--	--	--	0.13	--	--
1994	0.06	0.19	--	0.19	--	--	--	0.14	--	--
1995	0.08	0.18	--	--	--	--	--	0.13	--	--
1996	0.08	0.19	--	--	0.20	--	--	0.13	--	0.22
1997	0.09	0.19	0.15	--	0.19	0.17	--	0.13	0.12	--
1998	0.09	0.19	0.12	0.21	0.20	0.15	0.17	0.14	0.12	--
1999	0.09	0.20	0.12	--	0.21	0.15	--	0.15	0.12	--
2000	0.09	0.21	0.13	--	0.21	0.15	--	0.14	0.12	0.20
2001	0.09	0.21	0.15	0.19	0.22	0.14	--	0.15	0.12	--
2002	0.11	0.21	0.16	--	0.22	0.17	0.18	0.15	0.14	--
2003	0.13	0.20	0.16	--	0.21	0.17	--	0.13	0.14	--
2004	0.11	0.18	0.17	--	0.21	0.17	--	0.11	0.14	--
2005	0.11	0.21	0.17	--	0.21	0.19	--	0.14	0.16	--
2006	0.11	0.21	--	--	--	--	--	0.16	--	--

Note: HSGPA data from the SAT Questionnaire were available for analysis by subgroup after 1991. Standardized differences for the ACT are reported for the years for which data were available on www.act.org. Although trend information is not reported until 1996, the SAT Subject Test in Writing was introduced in 1994 resulting in a cohort of scores reported in 1995, but the scores were subsequently rescaled and 1996 is the first year reported. There was another adjustment to the scale in May 1998 that resulted in an average increase of approximately 23 points in 1999. In grade 12, NAEP reading and math assessments are given at least as often as they had been in the past, or every four years on a nationally representative basis. The NAEP writing assessment is administered to grade 12 periodically. Standardized differences for NAEP are reported for the years for which data were available on http://nces.ed.gov/nationsreportcard/.

Table B28

Standardized Differences for Limited English Proficient (LEP) Test-Takers Across Measures, 1995 to 2006: LEP Minus Total

	HS Grades	*Reading*		*Writing*		*Mathematics*	
Year	*HSGPA*	*SAT Verbal/ Critical Reading*	*NAEP Reading*	*SAT Subject Test in Writing*	*NAEP Writing*	*SAT Math*	*NAEP Math*
1995	0.09	-1.06	--	--	--	0.07	--
1996	0.06	-1.05	--	-1.10	--	0.05	-1.06
1997	0.03	-1.01	--	-0.99	--	0.06	--
1998	0.00	-0.98	-1.24	-0.99	-1.29	0.07	--
1999	0.05	-0.97	--	-0.98	--	0.06	--
2000	0.06	-0.96	--	-0.95	--	0.07	-0.86
2001	0.06	-0.97	--	-0.93	--	0.06	--
2002	0.08	-0.96	-1.16	-1.02	-1.03	0.07	--
2003	0.03	-0.93	--	-0.91	--	0.07	--
2004	0.02	-0.89	--	-0.94	--	0.08	--
2005	0.03	-0.86	--	-0.85	--	0.12	--
2006	0.10	-0.79	--	--	--	0.12	--

Note: LEP for the SAT Verbal, Math, and Subject Test in Writing is determined by the student indicating that their best language is another language on the SAT Questionnaire. ACT does not collect language subgroup information. Although trend information is not reported until 1996, the SAT Subject Test in Writing was introduced in 1994 resulting in a cohort of scores reported in 1995, but the scores were subsequently rescaled and 1996 is the first year reported. There was another adjustment to the scale in May 1998 that resulted in an average increase of approximately 23 points in 1999. In grade 12, NAEP reading and math assessments are given at least as often as they had been in the past, or every four years on a nationally representative basis. The NAEP writing assessment is administered to grade 12 periodically. Standardized differences for NAEP are reported for the years for which data were available on http://nces.ed.gov/nationsreportcard/.

Table B29

Standardized Differences for Low-Income Students (<$30K) Across Measures, 1992 to 2006: Low Income Minus Total

	HS Grades	*SAT*			*ACT*
Year	*HSGPA*	*SAT Verbal/Critical Reading*	*SAT Math*	*SAT Subject Test in Writing*	*ACT Composite*
1992	-0.11	-0.35	-0.29	--	--
1993	-0.12	-0.35	-0.32	--	--
1994	-0.14	-0.36	-0.34	--	--
1995	-0.12	-0.37	-0.33	--	--
1996	-0.14	-0.38	-0.36	-0.52	--
1997	-0.15	-0.41	-0.38	-0.53	-0.37
1998	-0.17	-0.42	-0.39	-0.58	-0.38
1999	-0.17	-0.44	-0.41	-0.61	-0.40
2000	-0.18	-0.46	-0.42	-0.64	-0.42
2001	-0.20	-0.50	-0.46	-0.67	-0.46
2002	-0.24	-0.55	-0.51	-0.74	-0.47
2003	-0.23	-0.58	-0.54	-0.74	-0.47
2004	-0.22	-0.57	-0.51	-0.75	-0.47
2005	-0.22	-0.54	-0.47	-0.72	-0.49
2006	-0.19	-0.48	-0.44	--	--

Note: Although trend information is not reported until 1996, the SAT Subject Test in Writing was introduced in 1994 resulting in a cohort of scores reported in 1995, but the scores were subsequently rescaled and 1996 is the first year reported. There was another adjustment to the scale in May 1998 that resulted in an average increase of approximately 23 points in 1999. Standardized differences for the ACT are reported for the years for which data were available on www.act.org. NAEP does not report data by parental income.

Table B30

Standardized Differences for High-Income Students (>$100K) Across Measures, 1995 to 2006: High Income Minus Total

	HS Grades	*SAT*			*ACT*
Year	*HSGPA*	*SAT Verbal/Critical Reading*	*SAT Math*	*SAT Subject Test in Writing*	*ACT Composite*
1995	0.20	0.47	0.54	--	--
1996	0.21	0.50	0.54	0.27	--
1997	0.20	0.49	0.54	0.28	0.51
1998	0.20	0.49	0.54	0.27	0.53
1999	0.20	0.49	0.53	0.28	0.51
2000	0.18	0.48	0.50	0.28	0.51
2001	0.19	0.46	0.49	0.27	0.51
2002	0.17	0.46	0.46	0.28	0.52
2003	0.19	0.43	0.43	0.27	0.54
2004	0.16	0.40	0.39	0.26	0.54
2005	0.14	0.41	0.39	0.24	0.53
2006	0.14	0.41	0.40	--	--

Note: The income category of >100K was added to the SAT Questionnaire in 1994. Although trend information is not reported until 1996, the SAT Subject Test in Writing was introduced in 1994 resulting in a cohort of scores reported in 1995, but the scores were subsequently rescaled and 1996 is the first year reported. There was another adjustment to the scale in May 1998 that resulted in an average increase of approximately 23 points in 1999. Standardized differences for the ACT are reported for the years for which data were available on www.act.org. NAEP does not report data by parental income.


CollegeBoard
connect to college success™