A CRITICAL REVIEW OF THE SAT: MENACE OR MILD-MANNERED MEASURE?

Author:
Leslie Stickler

Faculty Sponsor:
Nancy Breland,
Department of Psychology

ABSTRACT

The SAT is a standardized measure of differences in scholastic ability and achievement among undergraduate applicants to U. S. colleges. As a measure of differences, the SAT is exceptionally accurate; SAT scores predict 20 to 30% of the variance in freshman college course grades despite differences among schools, in course difficulty, in professors' grading practices, and in student effort, mood, and adjustment to college. Although the SAT's statistical properties are second to none, the test and its makers are constantly under criticism for alleged biases. However, social and historical analyses reveal that such biases are not an inherent quality of the test; rather, they result from the ways people use the test and from inequities in American culture. Although the SAT may force our attention to the reality that all students are not offered an equal education, redesigning the messenger is not an appropriate solution. Furthermore, the trend toward achievement-based testing is more likely to exacerbate the SAT's reflection of societal inequality than to limit it. Achievement testing privileges those with the resources to prepare, whereas aptitude testing is less susceptible to socioeconomic advantages, such as coaching. The author suggests that the SAT should remain an aptitude test, to balance and supplement the achievement-based criteria of high school GPA and quality of high school courses.

OVERVIEW

The SAT, alternatively known as the Scholastic Aptitude Test, the Scholastic Assessment Test, and the SAT I: Reasoning Test, is a standardized measure of high school achievement and college aptitude. According to the College Board, the non-profit organization that oversees the SAT, the test assesses "skills students have developed over time and skills they need to be successful academically" (College Board, About, 2005). Such a measure allows colleges and universities to compare applicants from a wide variety of educational backgrounds. For example, although a 3.8 GPA at Lawrenceville Academy may not be equivalent to a 3.8 at Trenton Central High School, an SAT score of 2000 is the same everywhere.

The newest incarnation of the SAT consists of 3 hours and 45 minutes of writing, mathematics, and critical reading (the new title for the section formerly known as "verbal") questions spread over 10 sections (College Board, About, 2005). The 25-minute essay always begins the test, while a 10-minute grammar section always wraps it up. In the middle, students face 70 minutes each of mathematics and critical reading and 25 minutes of multiple-choice writing mechanics questions. Additionally, the test contains a 25-minute, unscored experimental section, which may include writing, mathematics, or critical reading questions. Question types for the writing sections include identifying sentence errors, improving sentences, and improving paragraphs (College Board, FAQ, 2005). In the mathematics sections, questions cover arithmetic, algebra, geometry, logic, and data interpretation (graphs), and some trigonometry and pre-calculus. In addition to multiple-choice questions, the mathematics section also incorporates "grid-ins," in which students generate their own answers to problems. The critical reading sections consist of sentence completions and reading comprehension questions for short and long passages.

The SAT is a norm-referenced measure that incorporates both "power" and "speed" components. Educational Testing Service (ETS), the company hired by the College Board to administer the SAT, constructs questions that only approximately 50% of students will be able to answer correctly to make a power test, and that only approximately 80% of students will be able to complete in the allotted time to add a speed factor (College Board, About, 2005).

HISTORY

Although the current formulation of the SAT is less than a year old, the test has been around for over 75 years. In 1901, the phrase "College Boards" was synonymous with the mentally strenuous, extended essay tests administered to Ivy League applicants (Barnes, 2002; Fontecchio, 2005). These entrance tests catered heavily to the elite educational backgrounds of the wealthy, white, Protestant males who took them, with questions on Latin, physics, and history (Barnes, 2002). The organization, College Board, was established primarily to standardize "College Boards" among universities; however, it was used its power as test master to shape educational priorities. For example, by including an essay on *Hamlet*, the College Board virtually guaranteed that every prep school in New England would include *Hamlet* in its curriculum.

In 1926, psychologist Carl Brigham designed the first SAT, the original Scholastic Aptitude Test (Barnes, 2002). Intended as a measure of "innate" intelligence, the first SAT tempered the educational privilege of rich students in New England by identifying less privileged students of exceptional potential. By 1934, Harvard, Princeton, and Yale were all offering scholarships to underprivileged "geniuses," opening up opportunities for higher education to a demographic previously excluded (Barnes, 2002). In the mid-1940s, the College Board dropped its "Boards" altogether in favor of the SAT. Intelligence testing became immensely popular at this time, hailed as a way to streamline education by sorting students according to ability and by identifying "superstudents"; the SAT was its crowning glory (Barnes, 2002).

Critics of intelligence testing and the SAT gained strength in the 1970s, following studies highlighting problems with IQ testing. They noted that many groups, including women and African Americans, scored lower on average than white males, and concluded that the test was biased (Barnes, 2002). Moreover, some found the notion of intelligence as fixed and stable misguided, mistaken, and un-American. SAT scores seemed to contradict the national ideology that hard work garners success; where was the meritocracy in "innate ability"?

The publicity generated by publishing test scores increased both the SAT's power and prestige and its notoriety. The U. S. Department of Education instituted the "Wall Chart," which ranked states against each other and against previous years as, for example, "up three, down two," creating anxiety over negligible differences (Adelman, 2000). *U. S. News & World Report*'s taxonomy of college quality first appeared in 1985; each school's average student SAT score continues to serve as its *coup de grace* (Adelman, 2000). Thus, colleges began recruiting SAT high-scorers to increase their status, leading to more reports of ever-rising scores and the apprehension and resentment of applicants and their families who no longer felt competitive. Eventually, even some college admissions committees began to question the appropriateness of the importance accorded their schools' average SAT scores (Evans, 2005; Miller, 2005).

Although the College Board tried to address public and institutional concerns with studies of the SAT's predictive validity for collegiate GPA, criticisms mounted, culminating in the actions of Richard Atkinson, president of the University of California system, in February of 2001 (Zwick, 2002). Driven by concerns about the face and construct validity of the test, Atkinson advocated the use of SAT II tests in specific subject areas—writing, mathematics, and one of each student's choice—as he regarded these tests as less biased by race and socioeconomic class privileges than the SAT (Zwick, 2002). Atkinson's intuition suggested that these achievement tests would measure effort rather than ability, increasing fairness. He used the University of California system's market share as an SAT consumer to force the College Board's hand in developing a more relevant, achievement-based measure (Barnes, 2002).

Atkinson's scathing remarks about the irrelevance of "aptitude" testing found an ally in Gaston Caperton, head of the College Board (McGrath, 2004). Caperton also believed that there was no place for an "IQ test" in college admissions and called instead for the measurement of "developed skill" (Barnes, 2002). The president of the College Board had an additional agenda of his own, one not far off from the intentions of the authors of the original College Boards: to improve secondary education by changing curricular priorities. Mr. Caperton wanted to see more rigorous mathematics preparation in American high schools, so he encouraged the College Board to add more advanced Algebra and Trigonometry to the test. He lamented that writing was a "shamefully neglected" subject; the revised SAT devoted new

sections to writing production and mechanics (McGrath, 2004). Likewise, both he and Atkinson wanted what they regarded as the "useless" analogies questions off the test and out of the curriculum.

The assumption underlying these changes is that all American high schools are capable of rising to meet these new testing criteria—that administrative or professional limitations alone have kept teachers from challenging students. Whether or not this assumption is true remains to be seen. Perhaps the new, criterion-referenced format will lead to better high school curriculums and more skilled, better prepared applicants of all colors. Or perhaps it will lead to a greater division between students whose districts have the resources to meet Atkinson's and Caperton's challenge and students whose districts do not.

As the new SAT is still less than a year old, the College Board has yet to put together a statistical analysis of what improvements in fairness, if any, have resulted from the reforms. However, a study in the UK found that using an aptitude test like the SAT instead of the achievement-based "A-levels" currently in place increased the percentage of high scoring students, especially students from underprivileged schools (West & Gibbs, 2004). This suggests that a measure of academic aptitude, rather than achievement, is actually the fairer test for racially diverse and low socioeconomic class groups, contrary to Atkinson's perceptions. Only future studies of the new test will tell.

STANDARDIZATION SAMPLE AND NORMS

As the SAT's long history suggests, its standardization sample and norms have changed over the years. The test that preceded the SAT, the College Boards, was standardized on a sample of 978 predominantly upper and upper-middle class white Anglo-Saxon males from New England; while this sounds hopelessly biased, it did accurately reflect the population of college students at the time (Barnes, 2002). The original SAT, introduced in 1926, was not standardized until 1941; prior to that, the members of each year's cohort were compared to each other only (Dorans, 2002). By 1941, multiple yearly administrations of the SAT became necessary to accommodate more college applicants, creating the problem of non-equivalence between applicants who took the test with more or less able peers during a single year (Dorans, 2002). Thus, the College Board decided that all students who completed the April 1941 SAT would serve as the standardization sample for the verbal section, and those who took the April 1942 SAT would constitute the mathematics section standardization sample; these reference groups were not updated until April 1995 (Dorans, 2002).

Although the 1941 and 1942 standardization group of 10,000 test takers consisted primarily of self-selected, privileged, white males applying to prestigious New England colleges (the College Board's membership base at the time), changes in the population of test takers over the next few decades had to be weighed against the entrenched utility of the test. Scores, especially Verbal scores, declined from the original mean of 500 and population variance increased from the 1940s standard deviation of 100 as the population of college students exploded in the 1950s and 1960s, but SAT users were hesitant to revamp a test that had already been carefully established (Dorans, 2002). By 1990, the mean SAT verbal score was 425 and the mean mathematics score was 475, substantial drift that indicated meaningful population changes and a lack of alignment between the mathematics and verbal scales (Dorans, 2002).

Finally, the College Board re-centered SAT mathematics and verbal scores, using the 1990 cohort of 1,052,000 college-bound seniors as its new reference group (Dorans, 2002). Re-centering the SAT meant both collecting data from a new standardization sample and rescaling the distribution of scores to reset the mean to 500 and the standard deviation to 100 for both sections of the test (Dorans, 2002). These students' average verbal score of 424 was statistically equated to 500, as was the average of 476 on the mathematics section. By measuring the entire population of test takers in 1990, the College Board accurately captured increases in the number and in the racial, class, and gender diversity of college applicants in the new SAT norms.

Information on the standardization sample for the new SAT was not yet available as of Fall 2005.

EVIDENCE OF RELIABILITY

Despite changes in the population of test takers and in education over the last 75 years, the SAT has proven highly reliable. In fact, reliability studies have yielded such consistent results that researchers

focus instead on examining the criterion-related, incremental, and construct validity of the SAT (e. g., Burton, 2001; Camera & Echternacht, 2000; Dorans, Lyu, Pommerlich, & Houston, 1997; Everston & Millsap, 2004; Fleming, 2002; Ting, 2000; Zwick, 2002). As these researchers have found good evidence of the SAT's validity (see next section), it can be inferred that the SAT has met the statistical prerequisite of reliability.

How reliable is the SAT? The College Board website reports that a student's true score is within 30 points of his or her measured score (SEM = 30; College Board, FAQ, 2005). A test with a standard deviation of 100 and a SEM of 30 has an internal consistency coefficient of .91 (a = .91). In a study specifically investigating SAT I score change upon repeated testing, 1,120,563 students in the 1997 college-bound cohort who took the SAT I one to five times in their junior and senior years gained an average of 7 to 13 points on the verbal section and 8 to 16 points on the mathematics section (Nathan & Camara, 1998). Thus, a student who retested at the higher end of both ranges still would not breach the standard error of measurement, indicating high test-retest reliability. Additionally, in a study of the effects of commercial coaching on SAT scores, Powers and Rock (1999) found that the average increase in student scores after intensive coaching was only 30 points on the verbal section and 50 points on the mathematics section, demonstrating that SAT scores remain relatively stable even with intervening practice.

EVIDENCE OF VALIDITY

Amid the turmoil of criticism and controversy, the makers of the SAT have relied upon evidence of the test's validity to counter claims that the SAT is outdated, unfair, and measures only "how to take a multiple-choice test" (Barnes, 2002). These studies fall into two main groups: studies of the SAT's ability to predict success in college, and studies of the SAT's construct validity. The former group provides a consistent and resounding "Yes!" to the question of whether or not the SAT is a useful tool in the college admissions process, whereas the latter tends to confirm that the SAT measures a construct that corresponds predictably with other measures of academic achievement, such as high school GPA and ACT scores (Burton, 2001; Dorans et al., 2002; Camara & Echternacht, 2000; Everston & Millsap, 2004; Powers & Rock, 1999; Zwick, 2002).

In the category of predictive validity studies, Nancy Burton's (2001) meta-analysis of studies of classes graduating between 1980 and 2000 found that combined SAT scores accurately predicted many measures of success in college, from freshman GPA to acceptance to graduate programs. Additionally, Burton confirmed that high school GPA and SAT scores were consistent and equally precise predictors of college success for women, African American students, and differently-abled students; that is, although SAT scores under-predicted these groups' college GPAs, high school GPA under-predicted their performance to the same extent (Burton, 2001). Other desirable college outcomes, such as leadership and nonacademic accomplishments, were not as strongly predicted by high school GPA or by SAT score, providing evidence of divergent validity and emphasizing that neither of these measures of academic ability can or should stand alone as admissions selection criteria.

In another review of the SAT's ability to predict success in college, Camara and Echternacht (2000) verified the SAT's utility. Despite intracollegiate restriction of range (because admitted students tend to be similarly high-achieving) and inconsistent corrections for college class difficulty and grading standards, the authors found that SAT consistently served as a moderate to robust predictor of college success (Camara & Echternacht, 2000). Reanalyzing data from others' studies of college students test scores and grades, Camara and Echternacht (2000) demonstrated that adequately and consistently correcting for these types of variability approximately doubled already significant validity coefficients; the adjusted coefficients ranged from .49 for African American and Hispanic males to .63 for Asian American males. Use of SAT scores added a substantial increment to the prediction equation (r = .10), and the best prediction of college success was made by combining high school GPA and SAT mathematics and verbal scores (Camara & Echternacht, 2000). These numbers did not predict persistence to graduation as accurately as they did GPA, but this is unsurprising because persistence is influenced more strongly by environmental and personality factors.

Additional predictive validity studies focused on the SAT's ability to predict college success for specific populations (Fleming, 2002; Ting, 2000). Ting (2000) found that SAT mathematics scores and

students' realistic self-appraisal contributed most significantly to freshman GPA for Asian American students. SAT mathematics scores may have been useful predictors whereas verbal scores were not because Asian Americans tend to value applied science more highly than humanities and to pursue technical degrees (Ting, 2000). In a study of the factors that enhance SAT prediction for African American students, Fleming (2002) found that the SAT best predicted undergraduate GPA for black males at traditionally black colleges. When only these students were considered, the predictive validity of the SAT for African American students rose from .46 to .57, suggesting that nonacademic, adjustment factors of black students in white colleges, rather than test bias, account for the SAT's under-prediction of minority students' GPAs (Fleming, 2002).

Other researchers have established the SAT's construct validity by studying the extent to which SAT scores vary according to theoretical predictions (Everston & Millsap, 2004; Powers & Rock, 1999; Zwick, 2002). For example, Everston and Millsap (2004) found that educational opportunities and experiences both inside and outside the classroom correlated with SAT scores; the fact students who are better educated perform better on the SAT supports the claim that the test measures learned knowledge and skill rather than wealth or prowess with the multiple-choice format. Additionally, latent variable models indicated that educational experiences inside the classroom, through extracurricular activities, and at home moderated the relationship between socioeconomic status and SAT achievement (Everston & Millsap, 2004).

The previously mentioned results of Powers's and Rock's (1999) coaching study also provide evidence of the SAT's construct validity—that what the test measures is not subject to fluctuations from "mere coaching." While the perception that coaching substantially increases test scores has been used in arguments against the validity of the SAT, the falsity of this perception provides evidence that the SAT is valid (Zwick, 2002). Theoretically, SAT scores should be stable as long as knowledge and critical thinking skills remain stable; one would not expect short-term coaching substantially to change how students think about problems.

Responding to public concern and misconceptions about the SAT's construct validity, Zwick (2002) analyzed the effect of student socioeconomic status on multiple measures of student achievement and college potential, including ACT scores, SAT II scores, high school exit examination results, National Assessment of Educational Progress (NAEP) surveys, high school GPAs, and teacher ratings. She concluded that, although the SAT does measure wealth to the extent that wealthier students earn higher SAT scores, *all* of the other measures of previous achievement are wealth tests as well (Zwick, 2002). Thus, the public appears to be asking the wrong question of the SAT and blaming it for revealing unpleasant truths; what American students and their families really want to know is: how important should past achievement be for college admissions when not all high school students have the same resources and opportunities to achieve? (Zwick, 2002).

SCORE REPORT INFORMATION

SAT scores are reported on a normalized scale that ranges from 200 to 800 (College Board, About, 2005). The 1990 SAT recentering reset the mean SAT score to 500 starting in 1995, with a standard deviation of 100, so that the possible distribution of scores from 200 (z = -3) to 800 (z = +3) corresponds to percentile ranks (College Board, FAQ, 2005). However, the actual mean score has increased over the past 10 years, so that the national mean Verbal score for 2005 was 508 and the mean Mathematics, 520 (College Board, Press Release, 2005). A combined SAT score of 1028 in 2005 indicated that a student performed as well or better than 50% of his or her national college-bound cohort. State-wide percentile rankings are also reported for test takers and also vary slightly according to mean cohort performance from year to year (College Board, FAQ, 2005). Preliminary reports on the new SAT indicate that students perform slightly better on the revision; the mean Mathematics score was 537, Critical Reading score was 519, and mean Writing score was 516. Detailed equivalency tables established during the 1990-1995 renorming of the SAT provide information about statistically equating and converting each student's raw scores into these standardized scores (Dorans, 2002).

Raw scores for each section of the SAT are computed by summing the number of items in those sections that the student answered correctly, and then subtracting 1/4 of the number of items answered

incorrectly (College Board, FAQ, 2005). Subtracting a fraction of the questions answered incorrectly helps to control for guessing, which would decrease the test's reliability and predictive validity if left uncorrected. Omitted questions neither add nor subtract points from a student's raw score (College Board, FAQ, 2005). Thus, if a student completes 52 of the 67 Critical Reading questions on the new SAT, and answers 45 of the questions correctly and 7 of the questions incorrectly, his or her raw score will equal 45 – (1/4)7, or 43.25. This raw score will be rounded to the nearest whole number and statistically equated with raw scores on other forms of the SAT, then converted to its corresponding standardized score and percentile rank (College Board, FAQ, 2005).

COMPARISON TO OTHER COLLEGE ADMISSIONS TESTS

As mentioned in the section discussing the test's validity, the SAT correlates with many other measures of academic ability and achievement, including ACT scores (Dorans et al., 1997; Zwick, 2002). That is, students who score highly on the SAT also tend to score highly on the ACT ($r = .92$), and both scores are good predictors of college success (Dorans et al., 1997). However, in terms of what it assesses, the SAT's rival is its opposite (Barnes, 2002). Whereas the SAT focuses on critical thinking skills and academic potential, the ACT assesses student mastery of the high school curriculum: English, mathematics, reading, and science reasoning (Dorans et al., 1997). Prior to the revision of the new SAT, the two tests exemplified the rivalry between aptitude and achievement testing for college selection purposes. With the revision to make the new SAT more achievement-based, the SAT and ACT have become much more similar. Moreover, the ACT has followed the new SAT's lead by adding an optional essay and writing section (Barnes, 2002; McGrath, 2004).

The SAT also has been compared to the SAT II: Subject Tests, which assess specific content-area mastery rather than general proficiencies, and the Graduate Record Examination (GRE), the SAT for graduate school applicants (Barnes, 2002; Burton, 2001; ETS, 2005; McGrath, 2004; Zwick, 2002). The SAT II tests gained attention when Richard Atkinson suggested that they replace the general SAT for University of California admissions decisions because the SAT II measured "well-defined and readily described" criteria (Zwick, 2002). However, further analysis revealed that the increases in low socioeconomic class Hispanic students' scores, assumed to reflect more fair testing, really resulted from the use of a third SAT II test of the student's choice—in the case of Hispanic Californian students, the SAT II Spanish test (Zwick, 2002). The GRE correlates highly with the SAT because it is a more difficult, unrevised, and unrecentered version of the SAT, replete with the feared antonym, analogy, and quantitative comparison questions (Burton, 2001; ETS, 2005). Additionally, the GRE includes a two-essay Analytical Writing section that resembles an advanced version of the new SAT essay and is administered in a computer-adaptive format, rather than the paper-based format of the SAT.

SUMMARY AND RECOMMENDATIONS

The SAT serves as a standardized measure of differences in scholastic ability and/or achievement among applicants to U. S. undergraduate colleges and universities. In order to detect maximum differences between college-bound students, the SAT is a norm-referenced measure on which relatively few students excel. For the purpose of predicting how well students will perform in their college courses, the SAT is not only appropriate but exceptionally accurate. SAT scores typically predict 20% to 30% of the variance in freshman college course grades despite school differences and intra-collegiate restriction of range, course difficulty differences, grading differences, differences in students' effort, differences in student adjustment and mood, and differences in life events (e. g., illness of a parent or financial need to work; Camara & Echternacht, 2000).

The SAT's statistical properties are remarkable; yet the test and its makers are frequently criticized for the SAT's biases. However, these biases are not inherent in the test, but result from the ways in which the test is used and the biases of American culture. It is not the SAT's or ETS's fault that people track negligible differences in aggregate SAT scores as if they determined a municipality's or an ethnic group's worth. It is not the SAT's or ETS's fault that students who attend more crowded and less academically rigorous schools (and often are members of minority groups) do not perform as well on the

test. Indeed, the SAT may be performing a service by reminding people of the vast social, economic, and educational inequalities in this country. Redesigning the messenger in an attempt to obscure the truth is not an appropriate solution. It may seem easier to "fix" the test than society, but "fixing" the test is a waste of time when the problems stem from vast but often ignored social inequities, not analogy or quantitative comparison questions.

Furthermore, the current trend toward achievement-based testing is more likely to exacerbate the SAT's reflection of social inequity than to attenuate it. If there is an appreciable difference in score distributions according to group membership, that difference will probably be an increase in the score gap between wealthy whites from more privileged school districts and poor, racially diverse students from impoverished districts without adequate staffing or resources. I have already seen this trend, working with students whose parents can afford Kaplan's fee of almost $200 an hour for private tutoring: the new writing section is the most "coachable" on the test. I teach students the rules of grammar and a convenient template for a five-paragraph personal essay and their scores increase about 100 points. Is the message that the new SAT should send that students will be held accountable for the depth and breadth of educational opportunities their parents can afford? Just as the switch from achievement-based "College Boards" to SATs opened up opportunities for underprivileged students, the move back towards an achievement-oriented SAT will likely increase privileged students' advantage.

So what is the solution if, as Zwick claims, all measures of past achievement are wealth tests and, as I suggest, achievement testing privileges affluent, white students even more than aptitude testing does? For one, the SAT should be reformatted to make it as much an aptitude test as possible to offset the achievement-based criteria of high school GPA and the quality of high school courses. Additionally, underprivileged students' special attributes, such as life experience or bilingualism, should be considered along with objective criteria as valuable assets to the admissions portfolio and to the college environment. Finally, college admissions committees and media sources should stop placing excessive emphasis on a single number. Students are much more than the sum of their SAT scores.

REFERENCES

Adelman, C. (Winter 2000). The SAT trap: Why do we make so much of one 3-hour test? *American Educator,* 38-41.

Barnes, J. E. (2002, November 11). The SAT revolution. *US News & World Report, 133,* 50-59.

Burton, N. (2001). *Predicting success in college: SAT studies of classes graduating since 1980* (College Board Research Report No. 2001-2). Princeton, NJ: Educational Testing Service.

Camara, W. J., & Echternacht, G. (2000). *The SAT I and high school grades: Utility in predicting success in college* (College Board Research Notes RN-10). Princeton, NJ: Educational Testing Service.

College Board. (2005). SAT: About the SAT Reasoning Test. Retrieved October 26, 2005, from http://www.collegeboard.com/prof/counselors/tests/sat/about/about_sat.html.

College Board. (2005). SAT FAQ. Retrieved October 26, 2005, from http://www.collegeboard.com/highered/ra/sat/sat_scorefaq.html#rea6.

Dorans, N. (2002). *The recentering of SAT scores and its effect on score distributions and score interpretations* (College Board Research Report No. 2002-11). Princeton, NJ: Educational Testing Service.

Dorans, N., Lyu, C. F., Pommerich, M., & Houston, W. M. (1997). Concordance between ACT assessment and recentered SAT I sum scores. *College & University, 73,* 24-33.

ETS. (2005). GRE FAQ. Retrieved November 6, 2005, from http://www.gre.org/portal/site/ets/menuitem.1488512ecfd5b8849a77b13bc3921509/?vgnextoid=302433c7f00c5010VgnVCM1000022f95190RCRD&vgnextchannel=7196e3b5f64f4010VgnVCM1000022f95190RCR

Evans, B. (2005, February 3). Ohio State U. welcomes its brightest class ever. *The OSU Lantern*. Retrieved October 26, 2005, from LexusNexus database.

Everston, H. T., & Millsap, R. E. (2004). Beyond individual differences: Exploring school effects on SAT scores. *Educational Psychologist, 39,* 157-172.

Fleming, J. (2002). Who will succeed in college? When the SAT predicts black students' performance. *The Review of Higher Education, 25,* 281-296.

Fontecchio, M. (2005, October 26). In 1901, SAT had questions about Latin. *The Patriot Ledger*. Retrieved
    October 26, 2005, from LexusNexus database.

McGrath, C. (2004, November 7). Writing to the test. *The New York Times.* Retrieved October 26, 2005
    from LexusNexus database.

Miller, J. (2005, April 23). Commentary about the new SAT—and getting into college. *The Santa Fe New
    Mexican*. Retrieved October 26, 2005, from LexusNexus database.

Nathan, J. S., & Camara, W. J. (1998). *Score change when retaking the SAT: I Reasoning Test* (College Board
    Research Report No. 1998-05). Princeton, NJ: Educational Testing Service.

Powers, D. E., & Rock, D. A. (1999). Effects of coaching on SAT I: Reasoning Test scores. *Journal of
    Educational Measurement, 36,* 93-118.

Ting, S. R. (2000). Predicting Asian Americans' academic performance in the first year of college: An
    approach combining SAT scores and non-cognitive variables. *Journal of College Student Development,
    41,* 442-449.

West, A., & Gibbs, R. (2004). Selecting undergraduate students: What can the UK learn from the
    American SAT? *Higher Education Quarterly, 58,* 63-67.

Zwick. R. (2002). Is the SAT a 'wealth test'? *Phi Delta Kappan, 84,* 307-311.