Different Tests, Same Flaws:
Examining
the SAT I, SAT II, and ACT

by Christina Perez

Spurred in part by University of California (UC) President Richard Atkinson's February 2001 proposal to drop the SAT I for UC applicants, more attention is being paid to other tests such as the SAT II and ACT. Proponents of these alternative exams argue that the SAT I is primarily an aptitude test measuring some vague concept of "inherent ability," while the SAT II and ACT are "achievement tests" more closely tied to what students learn in high school. While the origins of the exams and the rhetoric of their promoters may differ, the SAT I, SAT II and ACT actually exhibit many of the same flaws and shortcomings.

Christina Perez is the university testing reform advocate for FairTest (http://www.fairtest.org). Her work focuses on encouraging colleges and universities to drop the use of SAT and ACT scores in admission decisions, as well as promoting other civil rights issues associated with standardized tests and higher education. She previously worked as a research associate at TERC in Cambridge, MA, where she assisted math curriculum writers and workshop developers in increasing the equity-related content of their workshops for teachers. She holds a dual bachelor's degree in public policy and women's studies from Brown University (RI).

Are Admission Exams Really that Different from One Another?

The roots of the SAT I lie in the racist eugenics movement of the 1920s. Carl Brigham, a psychometrician who worked on the Army Alpha Test (an "IQ" test administered to World War I recruits that was used to justify racial sorting), was hired by ETS in 1925 to develop an intelligence test for college admission. At that time Brigham believed that people of different races could be rank ordered by intelligence level, with African Americans at the lowest point in the pecking order (Lemann 1999). Drawing from his work on the Army Alpha Test, Brigham designed the earliest version of the SAT I, which was at that time taken by just over 8,000 students and promoted to elite colleges primarily in the Northeast. Al-

though it included some mathematical calculations and identification of shapes, this early version of the SAT was focused on word familiarity—a question format popular in intelligence testing. While the SAT I has been modified over the years, the current test remains quite similar in format and content to Brigham's 1925 creation.

Several years earlier, the College Board began administering the College Entrance Examination Board Achievement Tests. Covering subjects such as English, French, Greek, mathematics, botany, and physics, the initial exams were primarily composed of essays that tested mostly rote memorization of discrete facts. Criticism from high schools, charging that the board was trying to dictate what they should teach, led to revisions in the entrance exams (Crouse and Trusheim 1988). In 1916, the College Board offered new comprehensive exams that covered a wider range of material and were used exclusively by highly selective institutions of higher education. These were renamed the SAT II: Subject Tests in the mid-1990s and matched to an existing 200–800 point scale.

Today the College Board tries to distinguish between the two exams by claiming that each measures something unique:

> The SAT [I] measures verbal and mathematical reasoning abilities that students develop over time, both in and out of school, which are related to successful performance in college…SAT II: Subject Tests are designed to measure knowledge, and the ability to apply that knowledge, in specific subject areas. Students take the SAT II: Subject Tests to demonstrate to colleges their mastery of specific subjects." (College Board Web site 2002)

However, the Board's position on what the two exams measure is likely to shift given the changes to the SAT I announced in June 2002. In an attempt to make the SAT I more closely tied to high school curriculum, the verbal analogies section will be replaced with short reading comprehension passages; Algebra II will be added to the math section; and a new "writing" section will be included that is based on the current SAT II Writing Test. The changes are expected to go

The intended purpose of the SAT I, SAT II, and ACT is to help predict first-year college grades. Yet high school grades, class rank, and rigor of courses do a better job of forecasting college performance than any of the tests.

into effect in March 2005 and will be accompanied by a new combined high score of 2400 (the current perfect SAT I score is 1600). Although these changes are being promoted as a move toward making the exam more "achievement" based, this is largely a marketing strategy in response to criticism about the test from places such as the University of California. Most of the test's format and question content will remain unchanged, as will its disconnection from actual classroom learning.

In 1959 ACT, Inc. came into the picture as the manufacturer of a third major admission test. The ACT, created by E.F. Lindquist (who also designed the Iowa Test of Basic Skills) and Ted McCarrel, was initially intended to compete with the SAT I's foothold in college admission. According to the test's manufacturer, the ACT is distinguished from the SAT I by its link to classroom curriculum: "The ACT Assessment tests are curriculum based. The ACT Assessment is not an aptitude or an IQ test. Instead, the questions on the ACT are directly related to what is learned in high school courses in English, mathematics, and science" (ACT, Inc. Web site 2002). Students are rated with four subscores (Reading, English, Mathematics, and Science Reasoning), as well as a Composite score (1-36 scale).

In fact, students' scores on all three exams tend to be very similar. According to the College Board's own research the correlation between SAT I and SAT II scores hovers around .84, while the correlation between SAT I and ACT scores ranges from .89 to .92 (College Entrance Examination Board 1999a). In other words, students receive very similar scores on each of the exams, calling into question how reliable test-maker claims are that the exams measure something unique. Indeed, both the College Board and ACT produce concordance tables which allow the translation of SAT scores into ACT equivalents, and vice versa.

One reason for the strong relationship among SAT I, SAT II and ACT scores may be their similarity in format. All three are timed, multiple-choice tests normed on national samples of students. Since students are tested under similar conditions and in a similar way, the exams tap a narrow range of skills and cater to one kind of learning and test-taking style. If a student does not do well at timed, multiple-choice exams, his or her true abilities may not be reflected in the final score. For example, research has shown that the "speeded" nature of the exams tends to put females and students whose first language isn't English at a disadvantage (Kessel & Linn 1996; Schmitt, et al 1991). Such research may explain why females earn higher grades in high school and college but receive lower SAT and ACT scores.

Similarity in test question content is another explanation for the high correlation between scores. Items on the SAT I-Verbal, SAT II Literature, and ACT Reading tests all ask students to read a short essay and answer multiple-choice questions based on the passage's content. The SAT II Writing (multiple-choice section) and ACT Reading tests both contain English "grammar" items in which test-takers need to select the correct words or phrases to complete a sentence. A similar component will also be added to the "new" SAT I, making the three tests resemble one another even more than they currently do. For example, a SAT II Writing question asks test-takers to indicate which, if any, of the underlined words or phrases contains a grammatical error:

Alexis has discovered that she can express her creativity more freely through her sketches and not in her photography. *No error*. (College Entrance Examination Board 1994)

On a similar ACT question, test-takers must indicate the correct word choice for the following sentence:

Japanese students observe a rigorous annually schedule. (ACT 1997)

The answer choices consist of the following: a) no change, b) an annual rigorously, c) an annual rigorous, or d) a rigorous annual.

The SAT I-Math, ACT Mathematics, and SAT II Math IC tests cover very similar content and format (although the SAT I does not include trigonometry or statistics). Sample questions from each of the three exams illustrate this point (College Entrance Examination Board 1994 and 1995; ACT 1997):

How Well Do Admission Exams Predict Success in College?

The intended purpose of the SAT I, SAT II, and ACT is to help predict first-year college grades. Yet high school grades, class rank, and rigor of courses do a better job of forecasting college performance than any of the tests. College Board claims that the SAT I explains approximately 27 percent of the variation in first-year grades, while SAT II exams on average explain around 16 percent of the differences (College Entrance Examination Board 1999b; Ramist et al 2001). Although the addition of a third component on the SAT I may slightly improve the test's predictive power, the change will likely be small at best and will not be equal across demographic groups. The ACT has a similar predictive validity: 17 percent of explained variance (ACT 1998). High school grades—the single best predictor of first-year performance according to both test manufacturers—account for nearly 30 percent of the difference in freshman grades, even in an era of grade inflation and variability in quality among high schools.

Independent studies confirm these findings. *The Case Against the SAT* by James Crouse and Dale Trusheim demonstrates the SAT I's poor utility in forecasting both short- and long-term academic success. The authors compared two admission strategies, one using just the high school record and the other using high school record and SAT I scores. More than 90 percent of the admission decisions were the same under both strategies. However, for the 10 percent of the applicant pool in which the two strategies led to different admission decisions, the SAT-based approach led to a far greater number of rejections of otherwise academically qualified African-American and low-income applicants (Crouse and Trusheim 1988).

SAT I	SAT II	ACT
If $(y + 2)^2 = (y - 2)^2$, what is the value of y?	If $f(x) + 1 - x^3$, then $f(-1) =$	If $(x + k)^2 = x^2 + 22x + k^2$ for real numbers x, then $k =$?
a) 0 b) 1 c) 2 d) 4 e) 6	a) −2 b) −1 c) 0 d) 1 e) 6	a) 11 b) 22 c) 44 d) 88 e) 176

Test-maker claims about the radically different content and purposes of the exams do not hold up upon a comparison of test questions.

Researchers at the University of Pennsylvania looked at the power of high school class rank and SAT I and SAT II scores in predicting cumulative college grade point average (GPA). They found that the SAT I was by far the weakest predictor, explaining only 4 percent of the variation in college grades, while SAT II scores accounted for 6.8 percent of the differences in academic performance (Baron and Norman 1992). By far the most useful tool proved to be class rank, which predicted 9.3 percent of the variation in cumulative GPAs. Combining SAT I scores and class rank inched this figure up to 11.3 percent, still leaving almost 90 percent of the variation in grades unexplained.

Chicago State University researchers looked at the ability of the ACT to predict college academic performance. For the vast majority of the university's graduates who had mid-range ACT scores, the test explained only 3.6 percent of the differences in cumulative college GPA (Paszczyk 1994). In fact, the exam over-predicted the performance of students graduating in 1992, who had the highest average ACT score among the classes in the research study yet the poorest academic performance over four years at the university.

Research on the exams' ability to foretell long-term success shows even weaker predictive power. Crouse and Trusheim demonstrated that using the high school record alone to forecast who would complete a bachelor's degree resulted in "correct" admission decisions 73.4 percent of the time, while using the SAT I and high school GPA was accurate in 72.2 percent of the cases (Crouse and Trusheim 1988).

Drawing from a national database of nearly 10,000 students, one study sponsored by the U.S. Department of Education considered the value of high school grades, class rank, test scores, and rigor of courses in predicting attainment of a bachelor's degree. High school curriculum offered a higher correlation with bachelor's degree attainment (.54) than either SAT scores (.48) or class rank/GPA (.44) (Adelman 1999). Significantly, high school curriculum was an even stronger predictor for African-American and Latino students than for students overall, indicating the value of employing broader admission criteria than just a test score/high school GPA index when trying to forecast the college performance of underrepresented minorities. Researcher Clifford Adelman concluded: "[T]he intensity and quality of curriculum is a cumulative investment of years of effort by schools, teachers, and students, and provides momentum into higher education and beyond. It obviously pays off. The effects of grades and tests diminish in time, but the stuff of learning does not go away."

> The new writing section of the SAT I will not alleviate that exam's "coachability," and may , in fact, increase the ability of test prep courses to boost students' scores.

The Impact of Test Coaching

Further diminishing their power to predict academic performance in college, the SAT I, SAT II, and ACT are all susceptible to coaching. While there is a great deal of debate surrounding *how* coachable exams are, numerous studies have demonstrated that student scores on admission exams can be significantly boosted through rigorous coaching. For example, the Princeton Review test preparation company guarantees a 100+ point increase in SAT I scores, and a 4+ point increase in ACT scores (equivalent to approximately 150 points on the SAT I) for students enrolling in one of their test prep classes.

The predictability of the exams' formats and their narrow range of content are cited by test prep companies as the primary reasons why students' scores can be boosted through coaching. The Princeton Review Web site states:

> The ACT tests the same information the same way, year after year…By reviewing the very specific knowledge that the people who write the ACT think is important, and by learning good test-taking strategies, it should be possible to increase your ACT score significantly" (2002).

Kaplan, the largest test prep company, makes similar claims about the SAT I: "The SAT I lends itself to preparation because it is a predictable test concentrating on a set of specific reasoning skills that can be learned." (Kaplan Web site 2002). The SAT II exams are equally susceptible to coaching. David Owen writes in *None of the Above*: "As John Katzman [president of the Princeton Review] said…ETS' Achievement Tests [SAT IIs] are even easier to coach than the SAT [I]—not because, as some have said, their subject matter is more finite but because the method by which they are put together is more transparent" (Owen 1999). The new writing section of the SAT I will not alleviate that exam's "coachability," and may, in fact, increase the ability of test prep courses to boost students' scores. The narrow criteria used to grade essays lends itself to coaching, as evidenced by the success test prep companies have in preparing students for the SAT II Writing Test.

Score gains that come with test coaching exacerbate the inequities already present with college admission exams. These courses, which can cost $800 or more, skew scores in favor of higher-income test-takers, who already tend to do well on the exam (Stockwell et al 1991). Because college admission officials do not know who has been coached and who has not, they cannot fairly compare two applicants' scores.

Will Substituting One Exam for Another Increase Racial Diversity?

Some critics of UC President Atkinson's proposal to drop the SAT I and rely instead on tests of "academic achievement" such as the SAT II or ACT argue that the measure is simply a way to skirt the state's ban on affirmative action. Yet similar racial/ethnic score gaps exist on the SAT II and ACT. The following chart illustrates the deep divides between students of different ethnic groups on all three admission exams:

Ethnic Group Score Gaps (Points above or below White Average)

SAT I Verbal + Math		SAT II Writing		SAT II Math IIC		ACT Composite (SAT equivalent)	
African American	-198	African American	-82	African American	-69	African American	-4.9 (-197)
American Indian	-95	American Indian	-51	American Indian	-37	American Indian	-3.0 (-120)
Asian/Asian American	+6	Asian/Asian American	-30	Asian/Asian American	+16	Asian/Asian American	-0.1 (-4)
Mexican American	-145	Mexican American	-106	Mexican American	-63	Mexican American	-3.3 (-132)
Other Latino	-130						
Puerto Rican	-151	Puerto Rican	-67	Puerto Rican	-39	Puerto Rican	-2.4 (-100)

Source: College Entrance Examination Board, *2001 College-Bound Seniors: A Profile of SAT Program Test-takers* and written correspondence with the College Board; ACT, Inc., *The High School Profile Report Normative Data 2001.*

The score gaps on the SAT I and ACT Composite are almost identical and will not likely change with the introduction of the "new" SAT I in March 2005. Many colleges that require SAT II exams ask applicants to submit scores on three or more subject tests. When the score differences on the most commonly required SAT II Writing and Math IIC are added together with a third exam (most of which have score gaps similar to those on the Writing and Math tests), the gaps will match, if not exceed, those present with the SAT I and ACT. In other words, replacing the "old" SAT I with the "new" SAT I or with the ACT or multiple SAT II exams will do little to increase racial diversity on college campuses if test scores are heavily emphasized.

> All three exams have a weak ability to predict academic performance in college, while high school grades, class rank, and rigor of classes taken are better predictors of success.

Is There a Better Way?

Despite differences in test-maker rhetoric, the SAT I, SAT II, and ACT contain similar flaws and shortcomings. All three exams have a weak ability to predict academic performance in college, while high school grades, class rank, and rigor of classes taken are better predictors of success. All three exams are highly coachable, advantaging students who can afford to spend $800 or more on test preparation classes. All three exams have a similar format, placing groups such as females and English as a Second Language learners at a disadvantage since they do not tend to perform as well on timed, multiple-choice exams. All three exams show large gaps in scores between students of different racial and economic groups, leading to racial and class bias in admission and financial aid formulas that utilize rigid test score requirements. Finally, all three exams place the financial and time burden on students rather than universities, making them low-investment sources of information for colleges but high-investment hurdles for students.

These findings lead to one central conclusion: *substituting one university admission exam for another benefits neither students nor schools.* Colleges that are genuinely committed to increasing the representation of African Americans, Latinos, and Native Americans must look into other alternatives.

Test-score optional policies are one significant step toward a sound and equitable admission process. Nearly 400 bachelor-degree granting institutions now admit a substantial number of freshman applicants without regard to SAT or ACT scores (see http://www.fairtest.org/univ/optional.htm for a list). Regardless of size, selectivity, geography, and student body composition, "test-score optional" institutions report widespread success in enrolling talented pools of diverse students.

The debate about which test score(s) to require doesn't acknowledge the larger issue present in reassessing college admission practices—all current standardized admission exams act as gatekeepers to students of color and low-income students. Until university officials and test-makers are willing to confront this fact, shifting admission policies from one test score requirement to another will simply uphold the faulty paradigm that "test scores equal merit" and maintain the narrow pipeline through which traditionally underrepresented groups struggle to pass.

ACT, Inc. 1998. ACT Prediction Research Summary Tables. Iowa City, IA: ACT, Inc.

ACT, Inc. 1997. *Getting Into the ACT: Official Guide to the ACT Assessment*. San Diego: Harcourt Brace & Company.

ACT, Inc. 2002. Web site: http://www.act.org/aap/faq/general.html.

Adelman, C. 1999. *Answers in the Tool Box: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment*. Washington, DC: U.S. Department of Education.

Baron, J. and Norman, M.F. 1992. SATs, Achievement Tests, and High School Class Rank as Predictors of College Performance. *Educational and Psychology Measurement* 52: 1047-1055.

College Board. 2002. Web site. http://www.collegeboard.com/sat/satguide/html/satinfo.html.

College Entrance Examination Board. 1994. *The Official Guide to SAT II: Subject Tests*. New York: College Entrance Examination Board.

College Entrance Examination Board. 1995. *Real SATs*. New York: College Entrance Examination Board.

College Entrance Examination Board. 1999a. *Concordance Between SAT I and ACT Scores For Individual Students*. New York: College Entrance Examination Board.

College Entrance Examination Board. 1999b. *Counselor's Handbook for the SAT Program 1999-2000*. New York: College Entrance Examination Board

Crouse, J., and Trusheim, D. 1988. *The Case Against the SAT*. Chicago: The University of Chicago Press.

Kaplan, Inc. 2002. Web site: http://www.kaplan.com.

Kessel, C. & Linn, M. 1996. "Grades or Scores: Predicting Future College Mathematics Performance," *Educational Measurement: Issues and Practice*, Vol. 15, N.4, Winter.

Owen, D. with M. Doerr. 1999. *None of the Above: The Truth Behind the SATs*. New York: Rowman & Littlefield Publishers, Inc.

Paszczyk, S. 1994. A Comparative Analysis of ACT Scores and Final GPA's of Chicago State University Undergraduate Students. ERIC Document no. 370 519.

Princeton Review. 2002. Web site: http://www.review.com.

Ramist, L., C. Lewis, and L. McCamley-Jenkins. 2001. *Using Achievement Tests/SA- II: Subject Tests to Demonstrate Achievement and Predict College Grades: Sex, Language, Ethnic, and Parental Education Groups*. New York: College Entrance Examination Board.

Schmitt, A.P., Dorans, N.J., Crone, C.R., & Maneckshana, B.T., 1991. *Differential Speededness and Item Omit Patterns on the SAT*. Research Report 91-50, Princeton, NJ: Educational Testing Service.

Stockwell, S., R. Schaeffer, and J. Lowenstein. 1991. *The SAT Coaching Cover-Up*, Cambridge, National Center for Fair & Open Testing.