

Research & Occasional Paper Series: CSHE.12.08

CSHE | Center for Studies in Higher Education

UNIVERSITY OF CALIFORNIA, BERKELEY

<http://cshe.berkeley.edu/>

**BACK TO THE BASICS:
In Defense of Achievement (and Achievement Tests) in College Admissions***

July 2008

Saul Geiser

Center for Studies in Higher Education
University of California, Berkeley

Copyright 2008 Saul Geiser, all rights reserved.

ABSTRACT

Summarizing a decade of research at the University of California, this paper concludes that admissions criteria that tap student mastery of curriculum content, such as high-school grades and performance on achievement tests, are stronger predictors of success in college and are fairer to low-income and minority applicants than tests of general reasoning such as the SAT.

How should we define "readiness" for college, and what should we look for when reviewing applicants for admission? Over most of the last century, there have been two schools of thought.

The traditional view has emphasized academic achievement, as demonstrated by students' high-school grades in college-preparatory subjects. This view reflects an older, reward-and-incentive philosophy of college admissions. Admission to college is seen as a reward for hard work and achievement in high school. Rewarding achievement not only assures colleges of a high-quality student body, in this view, but also has a broader incentive effect on K-12 education, encouraging schools to offer and students to take a rigorous, college-preparatory curriculum.

The alternative view is that students should be judged not simply on what they have learned, but on their ability to learn. The idea that students can and should be judged on ability is closely associated with the SAT and has captivated American college admissions since the SAT was first introduced.

Whereas the older "College Boards" had tested knowledge of college-preparatory subjects, the "Scholastic Aptitude Test," introduced in 1926, purported to measure a student's capacity for learning. The idea of the SAT dovetailed perfectly with the meritocratic ethos of American college admissions. If aptitude for learning could be reliably measured, the SAT could provide a vehicle for social mobility: Colleges could identify promising students from disadvantaged backgrounds who, despite inferior instruction and academic

* Paper prepared for inaugural conference of Center for Enrollment Research, Policy and Practice: "Defining Enrollment in the 21st Century: Understanding Our Students and Our Commitments," University of Southern California, August 4-6, 2008.

performance, were nevertheless deserving of admission. Insofar as the SAT was standardized, it offered a more uniform measure of college readiness than high-school grades, since grading standards vary by school. Above all, the SAT provided a tool for prediction, giving admissions officers a means to distinguish among applicants who were likely to perform well or poorly in college.

All of these claims — equity, uniformity, technical reliability, and prediction — resonated closely with the meritocratic values of leading universities and help explain the SAT's remarkable endurance and growth over the past century. Though both the test and the terminology describing what it is intended to measure have evolved over time — from “aptitude” to “generalized reasoning” and most recently, “critical thinking”— the one constant has been the SAT's claim to tap students' general ability to reason and learn, as distinct from their mastery of specific subject matter.

By the turn of the century, however, the increasing selectivity of college admissions and the intensifying debate over educational equity had cast a harsh light on the SAT and, indeed, on the very notion that ability should trump achievement in assessing readiness for college. On closer scrutiny, many of the claimed advantages of the SAT over traditional measures of academic achievement have been found to be illusory:

- The SAT is a relatively poor predictor of student performance; admissions criteria that tap mastery of curriculum content, such as high-school grades and achievement tests, are more valid indicators of how students are likely to perform in college.
- As an admissions criterion, the SAT has a more adverse impact on poor and minority applicants than high-school grades, class rank, and other measures of academic achievement; admissions criteria that emphasize demonstrated achievement over potential ability are more likely to promote educational equity.

The University of California and the SAT

These are the conclusions of a series of research studies begun over ten years ago at the University of California. After Californians voted to end affirmative action in 1996, the UC system undertook a sweeping review of its admissions policies in an effort to reverse plummeting Latino and African-American enrollments. The aim was not to find a proxy for race — none exists — but to find admissions criteria that were valid indicators of later performance in college yet had a less adverse impact on poor and minority applicants.

The results were surprising. In studies of almost 125,000 students entering UC between 1996 and 2001, my colleagues and I found that the strongest predictor of college success was high-school grades in college-preparatory courses (Geiser with Studley, 2002; Geiser and Santelices, 2007). Grades also had the least adverse effect on admission of low-income and minority applicants. High-school grades are sometimes viewed as a less reliable indicator than the SAT because schools differ in grading practices. But SAT scores are based on a single sitting of three to four hours, whereas high-school GPA is based on repeated sampling of student performance over several years. And college-prep classes present many of the same kinds of academic challenges that students face in college — term papers, quizzes, labs, final exams — so it makes sense that prior performance in such activities would be indicative of later performance.

Irrespective of the quality or type of school attended, high-school GPA proved the best predictor not only of freshman grades in college, as many studies have shown, but also of long-term college outcomes such as cumulative grade-point average and four-year graduation (Geiser and Santelices, 2007). The predictive superiority of high-school grades was consistently evident across all entering classes, academic disciplines, and campuses in the UC system. Confident in its value as a selection criterion, UC introduced a new policy

in 2001 extending eligibility for admission to the top four percent of graduates from each California high school — based on their high-school GPA.

Advanced Placement courses, on the other hand, were poor predictors of student success in college, while hurting the chances of admission for applicants from schools with limited AP offerings (Geiser and Santelices, 2006). A growing number of students now enroll in AP because they earn “bonus points” simply for attending class, without taking or passing the AP exams. This boosts GPAs and improves admissions profiles. But while AP exam scores were a good indicator, mere enrollment in AP classes bore no relationship to students’ later performance at UC. These findings suggested that the widespread practice of inflating students’ grades for AP participation was unwarranted: “Bonus points” should be awarded only where students demonstrate actual mastery of the subject matter by taking and passing the AP exams.

Most surprising was the relatively poor predictive power of the SAT compared to achievement tests, such as the SAT II subject tests or AP exams, which measure mastery of specific subjects like biology or US history. The SAT’s claim to tap general reasoning abilities, independent of curriculum content, was long thought to give it an advantage in predicting how students will perform in college.

UC has required applicants to take both the SAT I and the SAT II achievement tests since 1968 and so had an extensive database to evaluate that claim. (UC is also a good laboratory for admissions research because it includes a mix of highly selective and less selective campuses.) Our data showed that achievement tests were consistently superior to the SAT in predicting college outcomes, including outcomes for poor and minority students (Geiser with Studley, 2002). After taking account of students’ grades and achievement-test scores, SAT scores added nothing to the prediction. The SAT II writing exam was a relatively strong indicator, a testament to the importance of writing in almost all college majors. But other subject tests, such as the AP exams, proved to be even better indicators (Geiser and Santelices, 2006). Like grades in college-preparatory courses, it makes sense that tested mastery of foundational subjects such as math, science, and history would be predictive of college performance. These findings helped then-UC president Richard Atkinson persuade the College Board to revise the SAT — adding a writing sample, dropping verbal analogies, and phasing in more advanced math — although it is an open question whether the “New SAT” introduced in 2005 has moved far enough in the direction of an achievement test.

In sum, the UC findings challenge the widespread belief in the SAT’s capacity to measure student ability and thus predict success in college. While students do differ in their abilities, it is questionable whether a three or four-hour test can measure such differences with sufficient precision reliably to predict college performance. Traditional measures of academic achievement such as high-school grades and curriculum-based tests are more valid indicators of how students are likely to perform in college.

Achievement Measures and Minority Admissions

The UC findings also challenge the long-held belief in the SAT’s capacity to identify high-ability students from disadvantaged backgrounds and promote greater equity in college admissions. This belief is rooted in the progressive narrative of American higher education and has proven remarkably enduring.

Yet our data showed that the SAT had more of an adverse impact on poor and minority applicants than traditional measures of academic achievement. Compared to high-school GPA, for example, SAT scores were much more closely correlated with students’ socioeconomic characteristics. As a result, the SAT lowered the chances of admission for underrepresented minority applicants, who come disproportionately from disadvantaged backgrounds. When UC applicants were rank-ordered by SAT scores, roughly half as many Latino, African-American, and American Indian students appeared in the top quintile, and twice as many fell into the bottom quintile, as when the same pool was ranked by high-school grades (Geiser and Santelices, 2007).

Nor was the SAT useful for identifying high-ability students from disadvantaged backgrounds, as traditional measures of academic achievement proved to be more effective for that purpose as well. High-school grades and subject tests were the strongest predictors of success at UC even for students from the most disadvantaged schools (Geiser with Studley, 2002). Looking at students with “discrepant scores” — those who scored well on the SAT but poorly on the SAT II subject tests — we found that members of this group came disproportionately from families with higher incomes, performed less well at UC, and were more likely to be white (Geiser and Studley, 2002).

Such differences in the demographic footprint of the SAT and traditional measures of academic achievement are especially problematic where, as in California, affirmative action can no longer be used to offset the SAT’s disparate impact. Given that impact — and given also the SAT’s limited predictive power — it was a straightforward decision to de-emphasize SAT scores in favor of high-school grades and achievement tests as requirements for admission to the UC system, as UC did shortly after affirmative action was phased out in 1998.

Achievement Measures and University Outreach

The advantages of achievement measures and the limitations of the SAT were evident not only in university admissions, but also in UC’s outreach programs to California public schools.

After affirmative action was dismantled, UC recognized early on that simply revising its admissions criteria would not be enough, and that the university would need to provide significant assistance to the state’s deteriorating K-12 school system in order to restore minority enrollments at UC over the long term. To that end, UC massively expanded its outreach programs to disadvantaged students and schools. These programs included not only conventional one-on-one tutoring and mentoring of students, but also “whole school” improvement efforts involving teacher, principal, and curriculum development. At their height, before later state budget cuts, UC outreach programs were serving over 300,000 students and 70,000 teachers and principals, and UC campuses had established school-university partnerships with 300 of the lowest-performing schools in the state.

College admissions criteria can have a profound influence — a “signaling effect,” as Michael Kirst has called it — on such schools. After UC introduced its policy extending eligibility for admission to the top four percent of graduates from each high school, many California schools in poorer districts found themselves pressured by parents to expand offerings of UC-required courses so that their children could qualify for the program.

UC’s experience in low-performing schools showed that curriculum-based achievement tests have significant advantages over the SAT in facilitating educational improvement and school reform:

- Achievement tests help reinforce the teaching and learning of a rigorous academic curriculum. Unlike the SAT, achievement tests assess students on materials that they have studied in the classroom, so that teaching, learning, and assessment are more closely aligned. The reinforcement that achievement tests provide for the college-preparatory curriculum may be even more important in low-performing schools than others: Experience in implementing state curriculum standards (as distinct from No Child Left Behind) suggests that a strategy of setting clear content standards, teaching to the standards, and assessing students against those standards may produce the greatest benefits within the most disadvantaged schools.
- Achievement tests serve an important diagnostic function: Unlike SAT scores, which tell the student only how well they have performed relative to others, achievement-test scores provide feedback on the specific areas of the curriculum where the student is strongest and weakest.

Achievement tests provide a better foundation for self-assessment, both for students as well as for their teachers and schools.

- Most important is the message that achievement tests convey to students. A low SAT score sends the message to students that their performance reflects a lack of ability, rather than factors such as unequal access to good schools and well-trained teachers. Especially for poor and minority students, SAT scores can be damaging to self-esteem and academic aspiration, reinforcing negative societal stereotypes. Achievement tests send a much different message. A low score on an achievement test means simply that the student has not mastered the specified content. This may be due to any number of factors, including inadequate instructional resources, inferior teaching — or lack of hard work on the part of the student. Achievement tests focus attention on determinants of performance that are alterable, in principle, and are thus better suited to the needs of educational improvement and reform.

Persuaded of the advantages that achievement tests offer both for college admissions and for K-12 schools, in 2002 the UC faculty crafted what may be the first comprehensive policy adopted by any major US university on the selection and use of admissions tests. The policy concluded that “achievement-oriented tests are both useful to the University in identifying high-achieving students and philosophically preferable to tests that purport to measure aptitude” (University of California, 2002).

Rewarding Achievement

The UC findings lend strong support for the traditional view that demonstrated achievement, rather than potential ability, should be the decisive consideration in college admissions. Traditional measures of academic achievement are stronger predictors of success in college than tests of general reasoning ability such as the SAT. Compared to the SAT, high-school grades and class rank are also fairer to poor and minority applicants. And curriculum-based achievement tests are best suited to help address the troubling disparities in our K-12 schools.

Restoring the emphasis on high-school record

Rewarding achievement in college admissions means, first and foremost, restoring the emphasis on high-school record. Cumulative performance in high school is the fairest and most meaningful measure of student achievement and the most reliable indicator of success in college. Determination of high-school GPA or class rank should be straightforward, without statistical weighting for enrollment in Advanced Placement classes. Not only is the latter practice unfair to students with limited access to AP offerings, but it is also invalid, according to the UC data, since mere enrollment in AP classes is unrelated to later performance in college. “Bonus points” for AP classes are justified only where students demonstrate actual mastery of the subject matter by passing the relevant AP exams. The key, again, is demonstrated achievement.

Limitations of the SAT and ACT as achievement tests

Proponents of the SAT often argue that, even if high-school grades are the primary basis for admission to college, reasoning tests provide additional information about applicants that can help admissions officers make better decisions. As a supplement to the high-school record, they assert, SAT scores improve predictions of college performance by a statistically significant increment. SAT scores are also useful as a check on grade inflation, helping to restrain the pressure on grading practices that would result if admissions officers relied on high-school record alone.

Yet curriculum-based achievement tests are a better supplement to the high-school record than the SAT. Subject-specific tests such as the AP exams, SAT II achievement tests, and the New SAT (old SAT II)

writing test add more incremental prediction, beyond what is provided by high-school record alone, than the verbal and mathematical reasoning components of the SAT. And unlike the SAT, subject tests more often function as end-of-course exams and are more closely aligned with what is taught in school, thus helping to reinforce a rigorous, college-preparatory curriculum.

Can the New SAT be considered an achievement test? Many of the changes in the test, such as the incorporation of the SAT II writing exam and the addition of higher-level math, are evidently intended to move the SAT in that direction. But the SAT's provenance as a reasoning test remains evident as well. Unlike the ACT, which develops its test content directly from surveys of high-school curricula, most of the New SAT's content is not developed in this manner and, as a result, is less closely linked to what students actually study in school. It would be difficult to imagine using the SAT as a high-school exit exam in the way that the ACT is used in some states.

Nor does the New SAT exhibit other characteristics of an achievement test. It remains a "norm-referenced" test, designed primarily to compare students against one another, rather than a "criterion-referenced" test, intended to measure students' mastery of curriculum content. It has little diagnostic value in providing feedback to students on specific areas of strength and weakness. And in the first nationwide study of the new test, College Board researchers found that while the writing exam, as expected, was the most predictive of the three individual SAT sections, overall the New SAT was no better at predicting college outcomes than the old SAT (Kobrin, et al., 2008).

The New SAT, in short, is a test at war with itself. Although it has added elements associated with achievement testing, the College Board has been at pains to demonstrate the psychometric continuity between the old and new versions of the test. The New SAT's heritage as a test of general reasoning ability still predominates, and it remains to be seen whether future iterations of the test will evolve more fully into a curriculum-based assessment.

The ACT, on the other hand, comes much nearer to what one would expect of an achievement test. Test content is based on periodic national curriculum surveys as well as review of state standards for K-12 instruction. The test is divided into five subject areas (now including an optional writing test) corresponding to the college-preparatory curriculum. The ACT appears less coachable than the SAT, and the consensus of students who have taken both tests is that the ACT places less of a premium on test-taking skills and more on content mastery. The ACT also has a useful diagnostic component to assist students as early as the eighth grade to get on and stay on track for college.

Like the SAT, however, the ACT remains a norm-referenced test. The ACT is scored in a manner that reproduces the same bell-curve distribution as the SAT, and colleges use ACT scores primarily to compare students against one another rather than to assess curriculum mastery or proficiency.

The ACT is also hampered in its aspiration to serve as the nation's achievement test by the lack of national curriculum standards in the US. The ACT has tried to overcome this problem through curriculum surveys, but the "average" curriculum does not necessarily reflect what students are expected to learn in any given state, district, or school. Nor does the ACT cover each of its subject areas in the same depth as the AP exams or SAT II subject tests. The idea of a single, national curriculum test may be an impossibility in the absence of a national curriculum.

An expanded role for subject tests

Of all nationally administered tests, subject-specific assessments such as the SAT II and AP exams are the best available exemplars of achievement tests. The SAT IIs, now officially renamed the SAT Subject Tests, are offered in 18 subject areas and the AP exams in 33. The SAT Subject Tests are hour-long, multiple-choice exams, while the AP exams take two to three hours and include a combination of multiple-choice,

free-answer, and essay questions. Both are administered at the end of the school year and serve, in effect, as end-of-course examinations. Test-prep services such as the Princeton Review advise students that the most effective way to prepare for subject exams is through classroom coursework, and in a telling departure from its usual services, the Review offers content-intensive courses in mathematics, biology, chemistry, physics, and U.S. history to help students prepare for these tests.

The SAT Subject Tests and AP exams do have limitations. Scoring on both tests is norm-referenced, despite the fact that colleges more often treat them as proficiency tests (especially the AP exams, which are used for college placement as well as admissions). And the AP program has come under fire from some educators who charge that, by “teaching to the test,” AP classes too often restrict the high-school curriculum and prevent students from exploring the material in depth. A number of elite college-prep academies have dropped AP for this reason.

Nevertheless, subject tests proved most effective of all nationally available tests in predicting student performance at UC. The AP exams, in particular, were remarkably strong indicators, second only to high-school GPA in predicting student outcomes (Geiser and Santelices, 2006). The SAT II writing test, now incorporated in the New SAT, was the next-strongest indicator, followed closely by the elective SAT II subject test that UC requires of all applicants. (Under this requirement, students submit SAT II scores in a subject area of their choosing.) Taken together, subject tests performed significantly better than either the SAT reasoning tests or the ACT (Geiser with Studley, 2002; Geiser and Santelices, 2006).

Subject tests also have advantages for students. They provide students an opportunity to demonstrate knowledge of subjects in which they excel, and can assist them in gaining admission to particular college majors. And insofar as students can choose which exams to take, subject tests can be fairer and less burdensome. The elective SAT II subject test had the lowest correlation of any test with students' socioeconomic status, while remaining a relatively strong indicator of their performance at UC (Geiser with Studley, 2002).

That students should be able to choose the tests they take for admission may seem anomalous to those accustomed to viewing the SAT or ACT as national “yardsticks” for assessing student readiness for college. But the real anomaly may be the idea that all students should take one test, or that one test is suitable for all students. In the final analysis, admissions tests must be judged on results. If readiness for college is operationally defined by pre-admissions measures that are most directly related to performance in college, then a selection of subject tests—including tests selected by students—is superior to either of the generic, national assessments.

Back to the basics

Emphasis on curriculum mastery can help restore a measure of rationality to the overheated world of “high-stakes” college admissions. Norm-referenced admissions tests, such as the SAT, which are designed primarily to compare students against one another, only add fuel to the fire. Such tests also do a disservice to poor and minority applicants. Even where these students achieve real gains in academic preparation, as measured on standards-based assessments, they fail to improve their relative standing on admissions tests, as the bar keeps rising and the competition for college pushes test scores for all applicants ever higher. Small differences in test scores too often lead to denial of admission for students from less-privileged circumstances, when in fact such differences have little, if any, validity as indicators of how they will perform once in college.

College admissions may never be a perfectly fair and rational process, but it can be fairer and more rational than it is today if we judge students on what really matters — demonstrated achievement, as reflected in the high-school record and performance on subject tests. Our first consideration should not be how an applicant compares with others, but whether he or she demonstrates sufficient mastery of college-preparatory

subjects to benefit from and succeed in college. When we apply that standard, as admissions officers well know, we will find that we have many more qualified candidates than places available, and our candidate pool will be more diverse. Then begins the true work of admissions in applying institutional selection criteria — special talents and abilities, leadership and community service, opportunity to learn, socioeconomic disadvantage, and, where permissible, race — to build an entering class that reflects our institutional values and commitments.

REFERENCES

Geiser, S. with R. Studley. (2002). "UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California." *Educational Assessment*, vol. 8, no. 1, pp. 1-26.

Geiser, S. and R. Studley. (2002). "Research addendum: Additional findings on UC and the SAT." University of California, Office of the President: <http://www.ucop.edu/news/sat/resaddendfinal.pdf>.

Geiser, S. and M. Santelices. (2006). "The role of advanced placement and honors courses in college admissions." Pp. 75 -114 in P. Gandara, G. Orfield and C. Horn, (eds.), *Expanding Opportunity in Higher Education: Leveraging Promise*. Albany, NY: SUNY Press.

Geiser, S. and M. Santelices. (2007). "Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes." Center for Studies in Higher Education, UC Berkeley: <http://cshe.berkeley.edu/publications/publications.php?id=265>.

Kobrin, J., Patterson, B., Shaw, E., Mattern, K., and S. Barbuti. (2008). "Validity of the SAT for predicting first-year college grade point average." College Board Research Report No. 2008-5. New York: College Board.

University of California, Board of Admissions and Relations with Schools. (2002). "The use of admissions tests by the University of California." University of California, Office of the President: <http://www.universityofcalifornia.edu/senate/committees/boars/admissionstests.pdf>.
