

UC Berkeley

Research and Occasional Papers Series

Title

NORM-REFERENCED TESTS AND RACE-BLIND ADMISSIONS: The Case for Eliminating the SAT and ACT at the University of California by Saul Geiser, UC Berkeley CSHE 15.17 (December 2017)

Permalink

<https://escholarship.org/uc/item/9fw4623g>

Author

Geiser, Saul

Publication Date

2017-12-01

**NORM-REFERENCED TESTS AND RACE-BLIND ADMISSIONS:
The Case for Eliminating the SAT and ACT
at the University of California**

December 2017*

**Saul Geiser
UC Berkeley**

Copyright 2017 Saul Geiser, all rights reserved.

ABSTRACT

Of all college admission criteria, scores on nationally normed tests like the SAT and ACT are most affected by the socioeconomic background of the student. The effect of socioeconomic background on test scores has grown substantially at University of California over the past two decades, and tests have become more of a barrier to admission of disadvantaged students. In 1994, socioeconomic background factors—family income, parents' education, and race/ethnicity—accounted for 25 percent of the variation in test scores among California high school graduates who applied to UC. By 2011, they accounted for 35 percent. More than a third of the variation in SAT/ACT scores is attributable to differences in socioeconomic circumstance. Meanwhile, the predictive value of the tests has declined with the advent of holistic review in UC admissions. Holistic review has expanded the amount and quality of other applicant information, besides test scores, that UC considers in admissions decisions. After taking that information into account, SAT/ACT scores have become largely redundant and uniquely predict less than 2 percent of the variance in student performance at UC. The paper traces the implications of these trends for admissions policy. UC has compensated for the adverse impact of test scores on low income and first-generation college applicants by giving admissions preferences for those students, other qualifications being equal. Proposition 209 prevents UC from doing the same for Latino, African American, and Native American applicants. California is one of eight states to bar consideration of race and ethnicity as a factor in public university admissions. Yet UC data show that race has an independent and growing effect on test scores, after controlling for other socioeconomic factors. The growing correlation between race and test scores over the past 25 years reflects the growing segregation of Latino and black students in California's poorest, lowest-performing schools. Statistically, race has become as important as either family income or parents' education in accounting for test-score differences among UC applicants. Using the SAT and ACT under the constraints of Proposition 209 means accepting adverse impacts on underrepresented minority applicants beyond what can be justified by the limited predictive value of the tests. If UC cannot legally consider race as a socioeconomic disadvantage in admissions, neither should it consider scores on nationally normed tests. Race-blind implies test-blind admissions. The paper concludes with a discussion of options for replacing or eliminating the SAT and ACT in UC admissions.

Keywords: Higher education, college admissions, standardized tests, race and ethnicity

The Growing Social Costs of Admissions Tests

Admissions tests like the SAT and ACT have both costs and benefits for colleges and universities that employ them. Their chief benefit is predictive validity, that is, their purported ability to predict student performance in college and thereby provide a standardized, numerical measure to aid admissions officers in sorting large numbers of applicants. Their chief cost is that test scores are strongly influenced by socioeconomic factors, as shown in Figure 1. The findings are based on a sample of over 1.1 million California high school graduates who applied for freshman admission at UC from 1994 through 2011.

Compared to other admissions criteria like high-school grades or class rank, scores on nationally norm-referenced tests like the SAT and ACT are more highly correlated with student background characteristics like family income, parents' education, and race

* Saul Geiser is a research associate at the Center for Studies in Higher Education at UC Berkeley and former director of admissions research for the UC system.

or ethnicity. To the extent that test scores are emphasized as a selection criterion, they are more of a deterrent to admission of low income, first-generation college, and underrepresented minority applicants.

Figure 1
Correlation of Socioeconomic Factors with SAT/ACT Scores and High School GPA

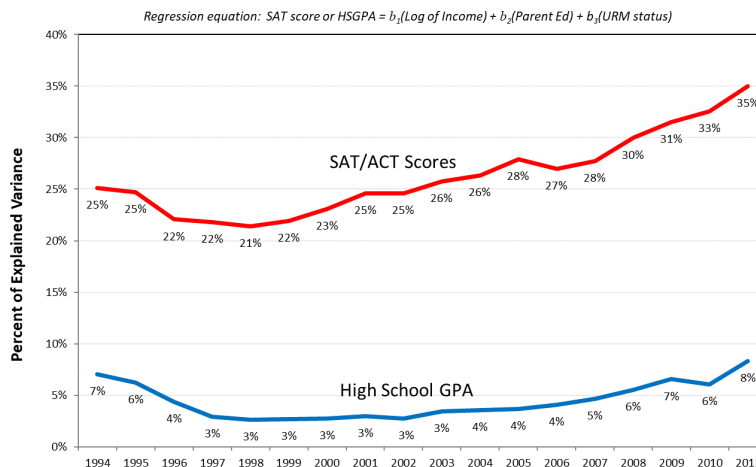
	Family Income	Parents' Education	Underrepresented Minority
SAT/ACT Scores	.36	.45	-.38
High School GPA	.11	.14	-.17

Source: UC Corporate Student data on all California residents who applied for freshman admission between 1994 and 2011 and for whom complete data were available on all covariates. N = 901,905.

At institutions that continue to employ the SAT and ACT, this cost/benefit trade-off is of course well-known. While many smaller colleges have gone “test optional,” most highly selective universities have not. At those institutions, a *détente* has emerged that accepts SAT/ACT scores as an indicator of individual differences in readiness for college, yet acknowledges that test scores are affected by disparities in students’ opportunity to learn. Rather than eliminate tests, colleges and universities make accommodations in how tests are used. Almost all selective institutions now consider SAT/ACT scores as only one of several factors in their “holistic” admissions process, and not necessarily the most important factor. Many institutions also give admissions preferences to applicants from socioeconomic categories most adversely affected when test scores are emphasized as a selection criterion. These and other accommodations serve to offset the adverse statistical effect of SAT/ACT scores on admission of low income, first-generation college, and underrepresented minority applicants.

Yet the balance of costs and benefits has changed over time, and the uneasy *détente* over testing has become more difficult to sustain. This is evident at the University of California. Over the past two decades, the correlation between test scores and socioeconomic factors has increased sharply among UC applicants, as shown in Figure 2:

Figure 2
Variance in SAT/ACT Scores and High School GPA Explained by Family Income, Parents' Education, and Race/Ethnicity, 1994 to 2011



Source: UC Corporate Student System data on all California residents who applied for freshman admission from 1994 through 2011 for whom complete data were available on all covariates. N = 901,905.

In 1994, family income, parental education, and race/ethnicity together accounted for 25 percent of the variance in test scores among UC applicants, compared to 7 percent of the variance in high school grades. By 2011, those same factors accounted for over 35 percent of the variance in SAT/ACT scores, compared to 8 percent for high school GPA. What this means is that over a third of the variation in students’ SAT/ACT scores is attributable to differences in socioeconomic circumstance. It also means that

test scores have become more of a deterrent to admission of poor, first-generation college, and Latino and African American applicants at UC.¹

UC as a Test Case for “Race-Blind” Admissions

To be sure, UC has taken extraordinary steps to offset the adverse impacts of test scores and to expand admission of students from disadvantaged backgrounds. In addition to holistic review, UC introduced Eligibility in the Local Context in 2001, extending eligibility for admission to the top 4 percent of students in each California high school based on their grades in UC-required coursework, irrespective of test scores. ELC was expanded to include the top 9 percent of students in each school in 2012. UC also gives admissions preferences for low-income and first-generation college students, other qualifications being equal. A common denominator in these reforms is their de-emphasis on test scores as a selection factor in favor of students’ high-school record and socioeconomic circumstances.

UC’s reforms have been closely watched for their national implications. California is one of eight states to bar consideration of race as a factor in university admissions, either through state referenda, legislative action, or gubernatorial edict. The other 42 states allow public colleges and universities to take applicants’ race into account as one among several admissions factors under US Supreme Court guidelines reiterated most recently in *Fisher II*.

California is viewed as a test case for “race blind” admissions. Some argue that race-neutral or “class-based” affirmative action—admissions preferences for low-income and first-generation college applicants—can be as effective as race-conscious affirmative action in boosting underrepresented minority enrollments indirectly and without the political controversy those programs have attracted. Some advocates have characterized UC’s reforms as a promising step in that direction (see, e.g., Kahlenberg, 2014a).

Admissions reforms undertaken at UC in the aftermath of Proposition 209 have substantially improved the number and proportion of low income and first-generation college students. By 2014, 42 percent of UC undergraduates were eligible for Pell grants (federal financial aid for low-income students), and 44 percent were first-generation college students. Both figures are far higher than other comparable universities, public or private. Four UC campuses now individually enroll more Pell-eligible students than the entire Ivy League combined. If class-based affirmative action can be an effective proxy for race-conscious affirmative action, the effect should be evident in California.

But the racial impact has been limited. Underrepresented minority enrollments—Black, Latino, and American Indian students—plummeted after Proposition 209 took effect at UC in 1998 and did not return to pre-209 levels until 2005. Today, underrepresented groups account for about 28 percent of all new UC freshmen, up from about 20 percent before affirmative action was ended.

Almost all that growth, however, is the result of underlying demographic changes in California. The racial/ethnic composition of the state has changed dramatically in the past quarter century, especially among younger Californians and particularly Latinos. White enrollments in the state’s public schools declined in both absolute and relative terms between 1993 and 2012, Asian enrollments held steady as a proportion of California’s school population, and black enrollments increased numerically but decreased proportionately. The number of Chicano and Latino students increased by 68 percent, becoming a majority of K-12 public school enrollments (Orfield & Ee, 2014). That demographic shift accounts for most of the growth in underrepresented minority enrollments at UC since Proposition 209 took effect. In fact, as measured by the gap between their share of new UC freshmen and their share of all college-age Californians, underrepresentation of these groups has worsened at UC since then.²

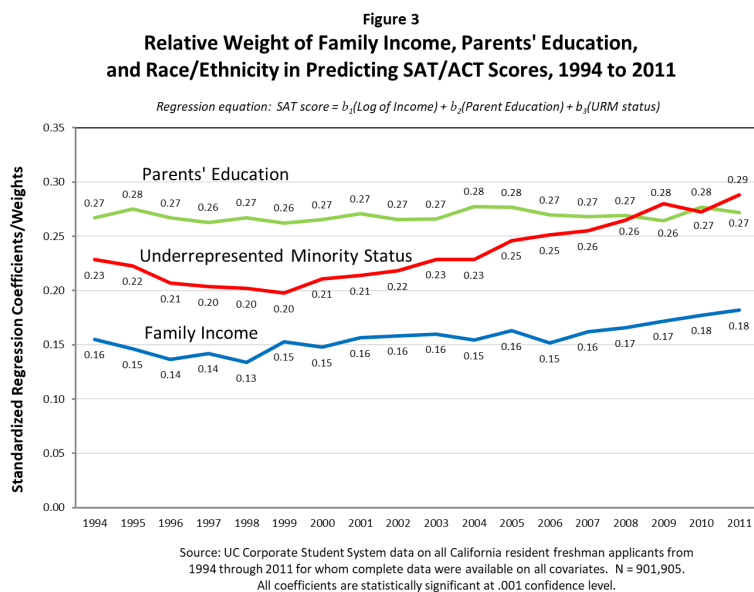
Race vs. Class as Correlates of Test Scores

Closer analysis of UC test-score data challenges an underlying assumption of class-based affirmative action: That racial and ethnic differences in test scores are mainly a byproduct of the fact that Latino and black applicants come disproportionately from poorer and less educated families. UC data show, to the contrary, that race has a large, independent, and growing statistical effect on students’ SAT/ACT scores after controlling for other factors. Race matters as much as, if not more than, family income and parents’ education in accounting for test-score differences.

¹ California is predominantly an “SAT state.” Fewer students take the ACT than the SAT, although that number has grown in recent years, and the ACT has overtaken the SAT nationally. Like UC, most American colleges and universities accept both tests and treat SAT and ACT scores interchangeably. Both are norm-referenced assessments designed to tell admissions officers where an applicant ranks in the national test-score distribution.

² A recent feature article in the New York Times found that underrepresentation of Hispanics in the UC system is more pronounced than in any other state, although underrepresentation of African Americans is slightly improved relative to their share of college-age population: [“Even With Affirmative Action, Blacks and Hispanics Are More Underrepresented at Top Colleges Than 35 Years Ago.”](#) (Ashsenas, Park, & Pearce, 2017).

This is not to suggest that family income, parents' education, or race/ethnicity "cause" test-score differences in any simple or direct fashion. Socioeconomic factors are mediated by other, more proximate experiences that affect test performance, such as access to test prep services or the quality of schools that students attend. Nevertheless, without being able to observe those intermediating experiences directly, regression analysis enables one to measure the relative importance of different socioeconomic factors in accounting for test-score differences. Figure 3 shows standardized regression coefficients for family income, parental education, and race/ethnicity in predicting test-score variation among UC applicants.³ The coefficients represent the predictive weight of each factor after controlling for the effects of the other two, thus allowing comparison of the relative importance of each.



Parents' education—the education level of the student's highest-educated parent—has long been the strongest predictor of test scores among UC applicants. In 1994, at the beginning of the period covered by this analysis, the predictive weight for parental education was 0.27. This means that for each one standard-deviation increase in parents' education, SAT/ACT scores rose by .27 of a standard deviation, or about 50 points, controlling for other factors. The predictive weight for parental education has remained about the same since then.

Family income, perhaps surprisingly, has the least effect on test scores after controlling for other factors, although it has shown a small but steady increase. The coefficient for income grew from 0.13 in 1994 to 0.18 in 2011. The latter represents a test-score differential of about 30 points.

But the most important trend has been the growing statistical effect of race and ethnicity. The standardized coefficient for race increased from 0.23 in 1994 to 0.29 in 2011 (about a 55-point test-score differential), the last year for which the author has obtained data. Underrepresented minority status—self-identification as Latino or African American—became the strongest predictor of test scores for the first time in 2009.⁴

A key implication of these findings is that racial and ethnic group differences in SAT/ACT scores are not simply reducible to differences in family income and parental education. Statistically, being born to black or Latino parents is as much of a disadvantage on the SAT or ACT as being born to poor or less-educated parents. Race matters as much as income or education in accounting for test-score differences.

³ The findings shown here and throughout the paper are based on a total sample of 1,144,047 California high school graduates who applied for freshman admission at UC between 1994 and 2011. The number of observations shown for particular figures and tables is less than total sample size due to missing data on one or more covariates.

⁴ Due to very small sample sizes and attendant concerns about confidentiality of student records, the data provided to the author by the UC president's office do not allow separate identification of American Indian applicants. For this reason, all data on underrepresented minority applicants shown here exclude American Indian or Native American students, who account for less than one percent of all UC applicants.

Racial Segregation and the “Test-Score Gap”

What accounts for the growing statistical correlation between race and test scores among UC applicants? It has occurred within a relatively short time span, less than a generation, so that any suggestion of a biogenic explanation is moot.

Nearly 20 years ago, the publication of Jencks and Phillips' (1998) *The Black-White Test Score Gap* sparked an extraordinary burst of research on the relationship between race and test scores. Most of that research focused on black-white differences and is not as directly relevant to the California experience as one might wish. The explosion of California's Chicano and Latino population together with continued growth among different Asian ethnic groups have overtaken the traditional, black-white dichotomy. Yet the earlier research has much to tell us.

Studies of the black-white test score gap can be divided into two strands: Those that seek to explain the gap primarily by reference to general socioeconomic factors, such as differences in family income and wealth; and those that emphasize factors specifically associated with race, such as discrimination and segregation.

A great deal of research on the black-white test score gap has favored the first type of explanation. Those studies have emphasized, among other influences, differences in family income,⁵ parental education,⁶ and quality of schools⁷ as factors underlying black-white test score differences. As a group, black students are disproportionately affected by all these factors. For example, black students have, on average, fewer resources in and out of the home, poorer health care, and less effective teachers, all of which can have an impact on test scores.⁸

The main problem with this body of research is that it leaves much of the black-white gap unexplained. Though class differences account for a portion of the test-score gap, they by no means eliminate it entirely. Conventional measures of socioeconomic status leave a large portion, even a majority of the gap, unexplained.⁹

Others who have studied the black-white test score gap have long noted the coincidence between trends in racial segregation and changes in the size of the gap. Following the US Supreme Court's decision in *Brown v. Board of Education*, racial segregation in US schools decreased dramatically during the 1960s and 1970s. The black-white test score gap narrowed significantly during the same period. School desegregation stalled in the 1990s as the result of court decisions restricting busing and other integration measures. Progress in narrowing the test score gap stalled at the same time.

The coincidence of these trends has prompted a great deal of research to determine whether they are causally linked. Berkeley economists David Card and Jesse Rothstein's (2007) study of SAT-score gaps in metropolitan areas across the US is generally regarded as dispositive: “We find robust evidence that the black-white test score gap is higher in more segregated cities” (p. 1).

Resurgence of Racial Segregation in California Schools

Racial segregation has increased sharply in California over the past quarter century, as documented by Gary Orfield and his colleagues at the UCLA Civil Rights Project. In 1993, about half of California's schools were still majority white schools, and only one-seventh were “intensely segregated,” meaning 90 percent or more Latino and black. By 2012, 71 percent of the state's schools had a majority of Latino and black students and 31 percent were intensely segregated.

California's schools now rank among the most segregated of all the states, including those in the deep South, on some measures. Measured by black students' exposure to white students, California ranks next to last. For Latino students, it ranks last. Today the typical Latino student attends schools where less than six percent of the other students are white.

A clear pattern of co-segregation of Latino and African American students has emerged. Slightly more than half of all Latino students now attend intensely segregated schools, and the comparable figure for black students in California is 39 percent. At the same time, “double segregation” by race and poverty has become increasingly common as the result of rising poverty levels within segregated school districts. Black students in California on average attend schools whose populations are two-thirds poor children, and Latinos attend schools that are more than 70 percent poor. Double segregation by race and poverty is overlaid with “triple

⁵ Magnuson & Duncan, 2006; Magnuson, Rosenbaum, & Waldfogel, 2008.

⁶ Haveman & Wolfe, 1995; Cook & Evans, 2000.

⁷ Corcoran & Evans, 2008.

⁸ Phillips, Crouse, & Ralph 1998; Magnuson, Rosenbaum, & Waldfogel, 2008.

⁹ Haveman, Sandefur, Wolfe, & Voyer, 2004; Meyers, Rosenbaum, Ruhm, & Waldfogel, 2004; Phillips & Chin, 2004. Some researchers hold out hope that better measures of wealth (as opposed to income) might account for the remainder of the test-score gap, but this remains unproven. See, e.g., Wilson, 1998; Kahlenberg, 2014b.

segregation” by language, as these same schools enroll high proportions of English language learners. On top of all, intensely segregated schools are typically under-resourced and staffed by teachers with weaker educational credentials.

Racial segregation is associated, in short, with multiple forms of disadvantage that combine to magnify test-score disparities among Latino and black students.

There are two aspects to this conclusion: First, social and economic disadvantage—not poverty itself, but a host of associated conditions—depresses student performance, and second, concentrating students with these disadvantages in racially and economically homogenous schools depresses it further (Rothstein, 2014, p. 1).

If segregation does have an independent effect on test scores, as much research on the black-white gap suggests, that explanation is consistent with observed trends in the California data: Why the correlation between race and test scores has grown for Latino as well as black students, and why the correlation remains after controlling for conventional measures of socioeconomic status.

Holistic Review and Proposition 209

In 1998, the Educational Testing Service proposed a new test-score measure, known as the “Strivers” proposal (Carnevale & Haggihat, 1998). ETS researchers compared student’s actual SAT scores with their predicted scores based on socioeconomic and other factors, including race. Students whose actual score significantly exceeded their predicted score were deemed “strivers.” As an ETS official explained, “A combined score of 1000 on the SATs is not always a 1000. When you look at a Striver who gets a 1000, you’re looking at someone who really performs at a 1200. This is a way of measuring not just where students are, but how far they’ve come” (Marcus, 1999). The proposal was later withdrawn after it sparked controversy and was rejected by the College Board.

Yet the underlying idea of the Strivers proposal remains very much alive at selective universities, like UC, that practice holistic admissions. “Holistic” or comprehensive review considers the totality of information in applicants’ files. Admissions staff who read files are “normed” and trained to evaluate indicators of academic achievement, such as test scores, in light of applicants’ educational and socioeconomic circumstances. Though far less algorithmic than the Strivers proposal, holistic admissions shares the same impulse to assess “achievement in context,” as emphasized in UC admissions policy:

Standardized tests and academic indices as part of the review process must be considered in the context of other factors that impact performance, including personal and academic circumstances (e.g. low-income status, access to honors courses, and college-going culture of the school).

UC’s holistic review process differs, however, from that at most other selective universities in one key respect: UC admissions readers cannot consider race as a contextual factor when evaluating applicants’ SAT or ACT scores. Proposition 209 amended the California state constitution “... to prohibit public institutions from discriminating on the basis of race, sex, or ethnicity.” While it made no mention of affirmative action, Proposition 209 was widely viewed as a referendum on that policy. Until then, UC had considered underrepresented minority status a “plus factor” in selection decisions. Supporters of the ballot measure saw it as a vehicle to end “race preferences” and “reverse discrimination” in university admissions.

Race was removed from applicant files after Proposition 209 took effect in 1998. UC continues to collect data on applicants’ race and ethnicity, but that information is not given to admissions readers. Proposition 209 has thus had the effect of eliminating any attention to race, whether as a “plus factor” or a socioeconomic disadvantage, in UC’s admissions process. In barring consideration of race as an admissions criterion, it has also effectively barred consideration of how other admissions criteria—like SAT and ACT scores—are themselves conditioned by race.

A Brief History of Validity Studies at UC

Of course, tests may have adverse statistical effects without necessarily being biased. If tests are valid measures of student readiness for college, test-score gaps between different groups may simply reflect real, if unfortunate, differences in opportunity to learn. The validity of admissions tests like the SAT and ACT is measured by their ability to predict student performance in college. How well do test scores predict student success at UC?

Early validity studies by BOARS: The relatively weak predictive validity of the SAT and ACT is a longstanding issue that prevented their adoption at UC until later than most other selective US universities. UC experimented with the SAT as early as 1960, when it required the test on a trial basis to evaluate its effectiveness. “Extensive analysis of the data,” BOARS¹⁰ chair Charles

¹⁰ Board of Admissions and Relations with Schools, the university-wide faculty committee responsible for formulating UC admissions policy.

Jones concluded in 1962, “leave the Board wholly convinced that the Scholastic Aptitude Test scores add little or nothing to the precision with which the existing admissions requirements are predictive of success in the University.” BOARS voted unanimously to reject the test (Douglass, 2007:90).

Lobbied by the Educational Testing Service, in 1964 UC conducted another major study of the predictive value of achievement exams such as the SAT II Subject Tests, which assess student knowledge of specific college-preparatory subjects, like chemistry or US history. The study again showed that test scores were of limited value in predicting academic success at UC, although achievement exams were slightly superior to the SAT. High school GPA remained the best indicator, explaining 22% of the variance in university grades, while subject tests explained 8%. Combining high school GPA and test scores did yield a marginal improvement, but the study concluded that the gain was too small to warrant adoption of the tests at UC. Again, BOARS’ decision was to reject the tests (Douglass, 2007:91-92).¹¹

Growing use of test scores in UC eligibility and admission: If UC was slow to embrace the national exams, the “tidal wave” of California high school graduates that Clark Kerr and the Master Plan architects had foreseen would soon tip the scales in favor of the SAT, if for reasons other than predictive validity. The growing volume of UC applications made test scores more useful as an administrative tool to cull UC’s eligibility pool and limit it to the top 12.5% of California high school graduates prescribed by the Master Plan. In 1968, UC for the first time required all applicants to take the SAT or ACT but restricted use of test scores to specific purposes such as evaluating out-of-state applicants and assessing the eligibility of in-state students with very low GPAs (between 3.00 and 3.09), who represented less than 2% of all admits. Test scores were still not widely used at this time.

UC began using the SAT and ACT more extensively as part of its systemwide eligibility requirements in 1979, following a 1976 study by the California Postsecondary Education Commission. The CPEC study showed that UC eligibility criteria then in place were drawing almost 15% of the state’s high school graduates, well in excess of the Master Plan’s 12.5% target. Rather than tightening high-school GPA requirements, BOARS chair Allan Parducci proposed an “eligibility index,” combining grades and test scores, to address the problem. The index was an offsetting scale that required students with lower GPAs to earn higher test scores, and the converse. The effect was to extend a minimum test-score requirement to most UC applicants. The proposal was controversial because of its anticipated adverse effect on low-income and underrepresented minority applicants, and it was narrowly approved by the regents in a close vote (Douglass, 2007:116-117).

The SAT would soon feature more prominently in UC admissions as well, and for the same reason: Its utility as an administrative tool to sort the ballooning volume of applications after the UC system introduced multiple filing in 1986. Multiple filing enabled students to submit one application to all UC campuses via a central application processing system. The volume of applications increased sharply the first year, and Berkeley and UCLA for the first time received vastly more applications from UC-eligible students than they were able to admit.

The UC system responded in 1988 with a new policy on “admissions selection” at heavily impacted campuses. The policy established a two-tier structure for campus admissions. The top 40% to 60% of the freshman class were admitted based solely on high-school grades and SAT/ACT scores. The remainder admitted based on combination of academic and “supplemental” criteria. Race and ethnicity were explicitly included among the supplemental criteria. The guidelines were, in effect, a compromise between the competing goals of selectivity and diversity. The compromise would hold for only a few years, however, before being shattered in 1995 by UC regents’ resolution SP-1, eliminating race as a supplemental admissions criterion.

2001 UC validity study: It might be assumed that UC’s increasing reliance on the SAT and ACT during this period was accompanied by concurrent improvements in the validity of the tests, but that assumption would be mistaken. In the aftermath of SP-1 (and a year later, passage of Proposition 209), UC undertook a sweeping review of all admissions criteria, with a special focus on standardized tests, in an effort to reverse plummeting minority enrollments. The findings of that review were strikingly similar to what BOARS had found 40 years earlier.

A 2001 study commissioned by BOARS showed once again that test scores yielded only a small incremental improvement over high school GPA in predicting freshman grades at UC. High school GPA accounted for 15.4% of the variance in first-year grades.

¹¹ To those unfamiliar with UC admissions, the superiority of high school GPA over standardized test scores in predicting UC student performance may seem odd, given that grading standards vary widely among schools. Despite this, high school GPA in college-preparatory coursework has long been shown to be the best overall predictor of student outcomes at UC, irrespective of the quality or type of high school attended. One reason may be that test scores are based on a single sitting of 3 or 4 hours, whereas high school GPA is based on repeated sampling of student performance over a period of years. And college preparatory classes present many of the same academic challenges that students will face in college—quizzes, term papers, labs, final exams—so it not should not be surprising that prior performance in such activities would be predictive of later performance.

Adding students' scores from the SAT into the prediction model improved the explained variance to 20.4%. SAT scores provided an incremental gain, in other words, of about five percentage points, close to what BOARS had found in 1964.

Figure 4
Percent of Variance in UC First-Year Grades
Predicted by High School GPA and Test Scores

Predictor variables/equations:	Explained variance:
(1) HSGPA	15.4%
(2) HSGPA + SAT	20.8%
(3) HSGPA + Subject Tests	22.2%
(4) HSGPA + Subject Tests + SAT	22.3%
SAT increment (4 - 3)	0.1%

Source: Geiser & Studley, 2002, Table 2. Based on a sample of 77,893 first-time freshmen entering UC from 1996 through 1999.

The 2001 study replicated another result from BOARS' earlier studies: Achievement exams such as the SAT II Subject Tests were slightly superior to the SAT I "reasoning test" in predicting first-year performance at UC. The subject tests provided an incremental gain of about seven percentage points over high school GPA alone. After taking account of high school GPA and SAT II Subject Test scores, SAT I scores yielded little further incremental improvement and became almost entirely redundant. Based in part on this finding, then UC president Richard C. Atkinson proposed eliminating generic, norm-referenced tests like the SAT or ACT in favor of curriculum-based achievement exams like the SAT II Subject tests.¹²

The College Board responded with a revised SAT in 2005. The new test eliminated two of its more controversial item-types, verbal analogies and quantitative comparisons, and added a writing test, thus addressing many of Atkinson's and BOARS' criticisms and moving the exam more in the direction of a curriculum-based assessment. But the changes did not alter the test's basic design and, like the ACT, it has remained a norm-referenced assessment. Nor have the changes improved the test's predictive validity. According to the College Board: "The results show that the changes made to the SAT did not substantially change how well the test predicts first-year college performance" (Kobrin, et al, 2008:1).¹³

Later validity studies: The SAT and ACT are designed primarily to predict freshman grades, and most validity studies published by the national testing agencies have focused on that outcome criterion. Later studies by independent researchers have examined whether the tests are useful in predicting other measures of success in college, like graduation, that many would regard as more important. The most definitive study of US college graduation rates to date is Bowen, Chingos, and McPherson's (2009) *Crossing the Finish Line*. Based on a massive sample of freshmen at 21 state flagship universities and four state higher education systems, the late William G. Bowen and his colleagues found:

High school grades are a far better predictor of both four-year and six-year graduation rates than are SAT/ACT test scores—a central finding that holds within each of the six sets of public universities that we study . . . Test scores, on the other hand, routinely fail to pass standard tests of statistical significance when included with high school GPA in regressions predicting graduation rates . . . (pp:113-115).¹⁴

With respect to first-year outcomes, later research has shown that traditional validity studies overstate the predictive value of the tests. Validity studies conducted by the College Board and ACT typically consider only two predictors: high-school grades and SAT/ACT scores. This simple, two-variable prediction model can be misleading, however, because it masks the effects of socioeconomic factors on prediction estimates. SES is correlated both with test scores and with outcomes like first-year grades, so that much of the apparent predictive value of test scores reflects the "proxy" effect of SES. When SES is included in the model, the predictive value of test scores falls. An independent analysis of UC test-score data by Berkeley economist Jesse Rothstein has found that the proxy effect is substantial:

¹² Atkinson, 2001. See Atkinson & Geiser, 2009 for an extended discussion of the preferability of criterion-referenced over norm-referenced assessments in university admissions.

¹³ The College Board introduced yet another revision of the SAT in 2016, but no change is expected in test validity.

¹⁴ Validity studies conducted at UC similarly show that SAT/ACT scores are very weak predictors of graduation rates (Geiser & Santelices, 2007).

The results here indicate that the exclusion of student background characteristics from prediction models inflates the SAT's apparent validity, as the SAT score appears to be a more effective measure of the demographic characteristics that predict UC FGPA [freshman grade-point average] than it is of preparedness conditional on student background . . . [A] conservative estimate is that traditional methods and sparse models [i.e., those that do not take socioeconomic factors into account] overstate the SAT's importance to predictive accuracy by 150 percent (Rothstein, 2004:297).

How well do test scores predict student success at UC?: As this brief history teaches, the answer depends not only on the outcome measure chosen but on what other academic and socioeconomic information is included in the prediction model. The advent of holistic review in UC admissions has substantially added to the body of information considered in admissions decisions. After taking that information into account, how much do SAT/ACT scores *uniquely* add to the prediction of student success at UC?

An answer is provided in a 2008 study by Sam Agronow, former director of policy, planning, and analysis at Berkeley. In a regression model predicting first-year grades, Agronow entered all available data from the UC application. In addition to high school GPA and SAT scores, these included: students' course totals in the UC-required "a-g" sequence, whether the student ranked in the top 4% of their class, scores on two SAT II Subject Tests, family income, parental education, language spoken in the home, participation in academic preparation programs, and the rank of the student's high school on the state's Academic Performance Index.

Entering all these factors into the prediction model, Agronow found that they explained 21.7% of the variance in students' first-year grades at Berkeley. When he eliminated SAT scores from the model, thus isolating their effect, the explained variance fell to 19.8%. SAT scores accounted for less than 2 percent of the variance in students' first-year grades at Berkeley. Across all UC undergraduate campuses, SAT scores contributed an increment of 1.6 percentage points.¹⁵

Test scores do add a statistically significant increment to the prediction of freshman grades at UC. But in the context of all the other applicant information now available, they are largely redundant, and their unique contribution is small.

The Problem of Prediction Error

It may be replied that any improvement in prediction, however small, is useful if it adds information. More information is always better than less, on this view. Using SAT and ACT scores in tandem with other applicant data may improve the overall quality of admissions decision-making and yield a stronger admit pool.

One problem with this argument is prediction error. In highly competitive applicant pools like UC's, where most applicants have relatively high scores, relying on small test-score differences to compare and rank students yields almost as many wrong predictions as correct ones.

Consider two applicants who are evenly matched in all other respects--same high school grades, socioeconomic characteristics, and so forth--except that the first student scored 1300 on the SAT and the second 1400. The choice would seem clear. Test scores are sometimes used as a tie-breaker to choose between applicants who are otherwise equally qualified, and in this case the tie would go to the second student.

What this ignores, however, are two measurement issues. First, because SAT/ACT scores account for such a small fraction of the variance in college grades, their effect size is quite small. Controlling for other academic and supplemental factors in students' files, a 100-point increment in SAT or ACT-equivalent scores translated into a predicted improvement of just 0.13 of a grade point, or an increase in college GPA from 3.00 to 3.13 for the UC sample.

Second, the confidence intervals or error bands around predicted college grades are large and, indeed, substantially larger than the effect size itself. The confidence interval around predicted GPA for the UC sample was plus or minus 0.81 grade points at the 95 percent confidence level, the minimum level of statistical significance normally accepted in social-science research. The error band around the predictions, in other words, was over six times larger than the predicted difference in the two applicants' test scores. For both students, the most that can be said is that actual performance at UC is likely to fall somewhere in a broad range between an A- and a C+ average.

Using SAT/ACT scores to rank applicants introduces a substantial amount of error in admissions decision-making. Two types of error are involved. First are "false positives," that is, students admitted based on higher test scores who later underperform in college. Second are "false negatives," that is, individuals denied admission because of lower test scores who would have performed

¹⁵ Compare regression models 21 and 22 in Agronow, 2008, pp. 4 and 107.

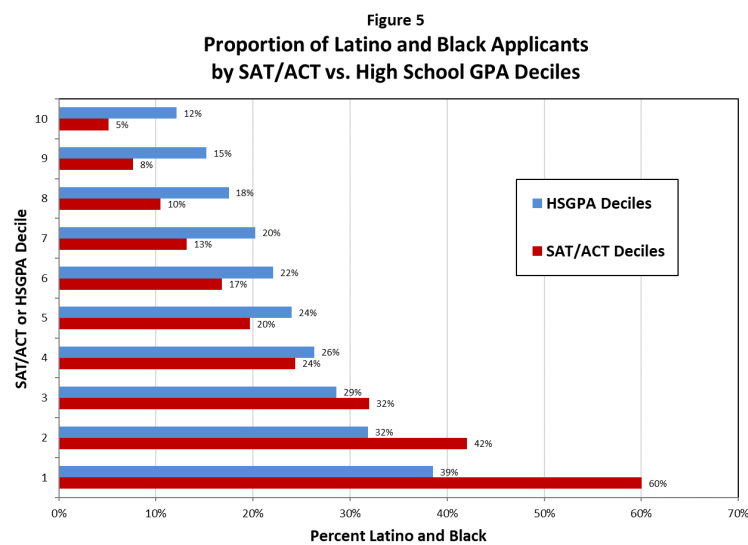
better than many of those admitted. Both types of error are inevitable when the predictive power of tests is low and score differences are small.

It might be countered that some predictive error is the price to pay for the utility of the SAT and ACT at large, public institutions that receive huge numbers of applications. Even if their predictive power is small, test scores do have statistically significant effect “on average,” that is, over large numbers of students. Whatever random errors they produce in individual cases may be outweighed by their utility as an administrative tool.

The problem is that the effects of the SAT and ACT are non-random. When used as a tool to rank applicants, SAT/ACT scores create greater racial stratification than other admissions criteria. More information is not always better than less if the new information introduces systematic bias in admissions decisions.

Norm-Referenced Tests and Racial Stratification

Norm-referenced tests like the SAT and ACT magnify racial stratification among UC applicants. Figure 5 will illustrate. All UC applicants were first ranked into ten equal groups, or deciles, based on their high school grades (blue bars). The same students were then ranked into deciles based on their SAT/ACT scores (red bars), and the proportion of Latino and black students in both distributions was compared.



Source: UC Corporate Student System data on all CA resident freshman applicants from 1994 through 2011 for whom complete data were available on SAT/ACT scores, high school GPA, and race/ethnicity. N = 1,127,267.

The difference is dramatic. Although SAT/ACT scores and high school GPA both create racial stratification, the demographic footprint of the tests is far more pronounced. At the bottom of the pool, underrepresented minorities comprise 60 percent of the lowest SAT/ACT decile, compared to 39 percent of the lowest HSGPA decile. Conversely, at the top of the pool—those most likely to be admitted—Latinos and blacks make up 12 percent of the top HSGPA decile but just 5 percent of the top SAT/ACT decile. Underrepresented minority applicants are less than half as likely to rise to the top of the pool when ranked by test scores in place of high-school grades.

The tendency of test scores to magnify racial and ethnic differences stems in part from the way norm-referenced assessments like the SAT and ACT are designed. Before going further, it should be emphasized that test publishers such as the College Board and ACT take strenuous measures to eliminate test bias, so that any such effect is unintentional. Gone are the days when test items such as “runner is to marathon as oarsman is to regatta” would survive the rigorous, multi-step review process that test publishers now follow.

Despite this, there are grounds for believing that exams like the SAT or ACT may unintentionally disadvantage Latino and black applicants because of the way those tests are developed. Before any item is included in the SAT, it is reviewed for reliability. Reliability is measured by the internal correlation between performance on that item and overall performance on the test in a reference population. If the correlation drops below 0.30, the item is typically flagged and dropped from the test. Some test experts argue that the test-development process tends systematically to exclude items on which minority students perform well, and *vice versa*, even though the items appear unbiased on their face:

Such a bias tends to be obscured because Whites have historically scored higher on the SAT than African Americans and Chicanos. The entire score gap is usually attributed to differences in academic preparation, although a significant and unrecognized portion of the gap is an inevitable result of ... the development process (Kidder & Rosner, 2002:159).

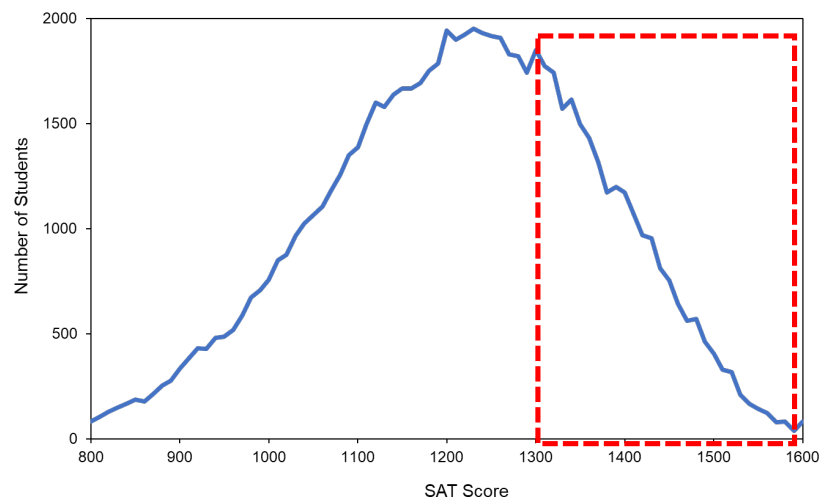
One measure of bias is known as “differential item functioning,” or DIF. DIF occurs “when equally able test takers differ in their probabilities of answering a test item correctly as a function of group membership” (AERA/APA/NCME, 2014:51).¹⁶ Although test developers go to great lengths to eliminate items that exhibit DIF, there is evidence that some residual DIF remains on the finished exams. One of the most rigorous studies to date is by psychometricians Veronica Santelices and Mark Wilson of Berkeley’s graduate school of education. Based on analysis of several SAT test forms offered in different years, the researchers found that about 10% of all items exhibited large DIF for black examinees and 3% to 10% for Latino examinees, depending on the year and form. Moderate-to-low levels of DIF were found on a substantially larger percentage of items for both subgroups (Santelices & Wilson, 2012:23).¹⁷

Yet item bias may be less important than the general design of norm-referenced tests in understanding why test scores tend to magnify racial differences. Before the SAT, the original “College Boards” were written exams designed to measure students’ academic preparation. Multiple-choice exams like the SAT and ACT did not come into prominence until later, as admissions became more competitive and colleges needed to make finer distinctions within applicant pools where most students were high performers. The SAT and later the ACT were designed to meet that need.

Although they differ in some other ways, both the SAT and ACT are “norm-referenced” tests, designed to compare an examinee’s performance with that of other examinees. They differ from “criterion-referenced” assessments that measure student performance against a fixed standard. A student’s SAT or ACT score measures how he or she performs relative to other test takers. How others perform matters as much as one’s own performance in determining one’s score. Scores on norm-referenced exams tell admissions officers where an applicant falls within the national test-score distribution.

Norm-referenced tests are designed to produce the familiar bell curve so beloved of statisticians, with most examinees scoring in the middle and sharply descending numbers at the top and bottom. To create this distribution, test designers must make the test neither too easy nor too hard. Designers adjust test difficulty by using plausible-sounding “distractors” to make multiple-choice items harder and by the “speededness” of the test, how many items the student must complete in a short space of time. Most important is the choice of items to be included. The ideal test item is one that splits the test-taking population evenly. Items that too many students can answer correctly are dropped from the test.

Figure 6
Frequency Distribution of Test Scores
Among California SAT Takers



¹⁶ Some have criticized DIF as a measure of item bias. DIF assesses the fairness of individual items by their correlation with test performance among students with similar abilities, which is measured by overall test scores. But if the overall test is itself biased, DIF may not necessarily detect item bias even when present (Kidder & Rosner, 2002).

¹⁷ Santelices and Wilson’s main finding was that DIF was inversely related to item difficulty. Surprisingly, score differences between Latino and black students and others was greater on *easier* rather than on more difficult SAT items.

The resulting bell curve is especially useful for admissions officers at highly competitive institutions like UC, where most applicants are concentrated in the higher-scoring, right-hand tail of the distribution. At those institutions, the problem is how to select from an applicant pool where most students are high performing. Norm-referenced tests assist by accentuating the importance of small score differences at the high end of the test-score range. There the distribution is steepest, and a small difference in test scores makes a bigger difference in a student's percentile rank. This has led to the creation of an entire test-prep industry, as students and their parents seek any small advantage in an increasingly competitive admissions environment. It is not surprising that tests designed to magnify the importance of small score differences should magnify underlying racial, economic, and educational differences at the same time.

National Standards for Fairness in Testing

In 2014, the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education jointly issued the latest revision of their *National Standards for Educational and Psychological Testing*. First issued in 1955, and previously revised in 1974, 1985, and 1999, the *Standards* address all aspects of test design, development, and use for educational purposes. While they do not carry the force of law, the *Standards* provide guidance on the responsibilities of test users to ensure "fairness in testing." What guidance can they offer UC regarding the difficult issues of race and testing?

The *Standards* obligate test users to be vigilant about the adverse effect of test scores on different social groups, such as racial and ethnic minorities. Adverse statistical effect is not enough by itself, however, to judge whether a test is unfair or biased. Where there are group differences in students' "opportunity to learn," a test may have adverse statistical impact but still be a valid indicator of what it is designed to measure. Adverse impact must be weighed against test validity.

When score gaps between groups are observed, the *Standards* advise test users to look for deeper psychometric warning signs, or "threats to fairness," that may indicate a problem with the validity of the test. "[G]roup differences in testing outcomes should trigger heightened scrutiny for possible sources of test bias." One warning sign is test content: Is there evidence that test items unfairly favor one group over another? Another is test context, that is, aspects of the testing environment that may favor or disfavor certain groups of examinees. Still another is differential prediction, as when a test predicts better for one group than another.

When tests exhibit such symptoms, the *Standards* offer this guidance to test users:

Standard 3.16: When credible research indicates that test scores for some relevant subgroups are differentially affected by construct-irrelevant characteristics of the test or of the examinees, when legally permissible, test users should use the test only for those subgroups for which there is sufficient evidence of validity to support score interpretations for the intended uses (AERA/APA/NCME, 2014:70).

All of the warning signs the *Standards* describe are evident in the admissions tests that UC employs. Test content is a first issue. As noted above, UC psychometricians have found differential item functioning on the SAT. Despite efforts of test developers to remove it, some statistically significant DIF remains for both black and Latino examinees (Santelices & Wilson, 2012).

Test context is another issue for underrepresented minority students:

[M]ultiple aspects of the test and testing environment ... may affect the performance of an examinee and consequently give rise to construct-irrelevant variance in the test scores. As research on contextual factors (e.g., stereotype threat) is ongoing, test developers and test users should pay attention to the emerging empirical literature on these topics so that they can use this information if and when the preponderance of evidence dictates that it is appropriate to do so (AERA/APA/NCME, 2014:54).

The *Standards'* reference to stereotype threat as a contextual factor in testing owes to the work of Stanford social psychologist Claude Steele. Steele and Aronson's (1995) study was the first to demonstrate that awareness of racial stereotypes has a measurable effect on SAT performance among black examinees.

Finally, differential prediction, or prediction bias, is also evident in the UC test-score data. This problem arises "... when differences exist in the pattern of associations between test scores and other variables for different groups, bringing with it concerns about bias in the inferences drawn from the use of test scores" (AERA/APA/NCME, 2014:51). For example, SAT/ACT scores are relatively

weak predictors of graduation rates for all UC students, but they are even weaker predictors for black and Latino undergrads.¹⁸ Tests predict least well for groups most disadvantaged by test scores.

When such warning signs are found, the *Standards* hold test users responsible for mitigating the adverse impact of their tests to ensure fairness for students from affected groups. In most states outside California, race and ethnicity can be taken into account in evaluating applicants' test scores, in the same way that family income and parental education are considered in identifying "strivers" who have overcome socioeconomic disadvantages. But how does Standard 3.16 apply in California and the seven other states where it is *not* "legally permissible" to consider race as a factor in admission to public colleges and universities?

In the end, the *Standards* are silent and punt the ball back to test users. As applied to UC admissions, a literal reading of Standard 3.16 would mean exempting Latino, African American, and American Indian applicants from the test requirement if test validity was deemed insufficient for those groups. SAT/ACT scores would be used only in admitting other applicants. Such an outcome is highly unlikely in California. Exempting any racial or ethnic group from the test requirement would be construed as a violation of Proposition 209. If UC were to exempt underrepresented minority applicants from submitting SAT/ACT scores on grounds of insufficient test validity, it would need to exempt all applicants.

Opportunity to Learn and Admissions Testing

Yet the problem goes beyond race. The larger problem is that SAT/ACT scores systematically disadvantage applicants who have had less opportunity to learn in favor of those who have had more, even when individual ability is equal.

The current *détente* over use of SAT/ACT scores in US college admissions rests crucially on the concept of opportunity to learn. The *Standards* define this as "... the extent to which individuals have had exposure to instruction or knowledge that affords them the opportunity to learn the content and skills targeted by the test ..." (p. 56). Differences in opportunity to learn are a threat to test validity and fairness. For example, if a school system establishes a test requirement for graduation, it is unfair to hold all students to the requirement if some have received inadequate instruction.

But the circumstances are different in college admissions, according to the *Standards*:

It should be noted that concerns about opportunity to learn do not necessarily apply to situations where the same authority is not responsible for both the delivery of instruction and the testing and/or interpretation of results. For example, in college admissions decisions, opportunity to learn may be beyond the control of the test users and it may not influence the validity of tests interpretations for their intended use (e.g., selection and/or admissions decisions). (p. 57).

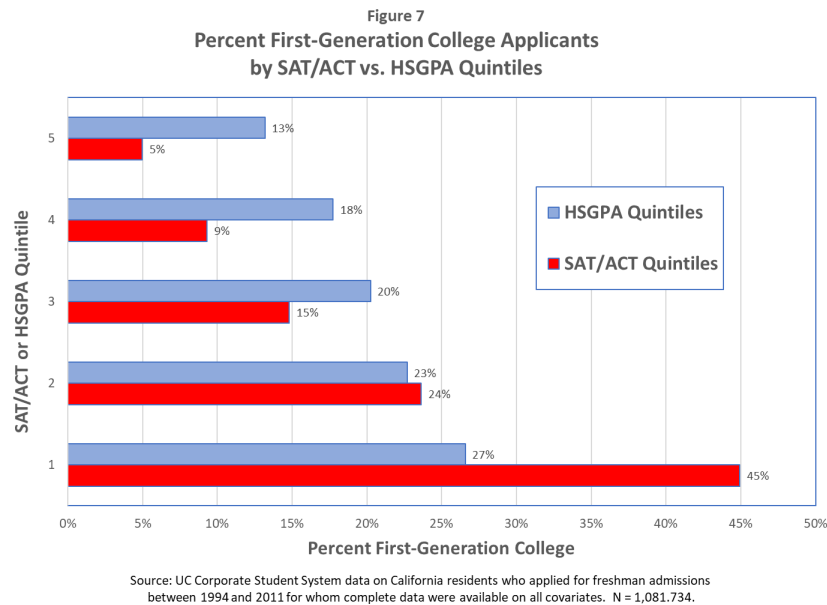
Put differently: Despite disparities in students' opportunity to learn and thus to perform well on the SAT and ACT, universities are justified in using them for admissions purposes if they are valid predictors of how students will perform in college.

The national *détente* over opportunity to learn in testing does not sit well in California. To start with, SAT/ACT scores don't predict student outcomes very well at UC. Test scores add little incremental validity in predicting UC student outcomes such as first-year grades or four-year graduation rates. Another issue is the size of the effect of opportunity to learn on test scores. Student socioeconomic characteristics now account for over a third of the variance in SAT/ACT scores among California high school graduates who apply to UC. Even granting that SES differences are beyond UC's control, the sheer size of the effect makes it difficult to justify using test scores to rank applicants.

SAT/ACT scores create built-in "headwinds"¹⁹ that reduce the chances of admission for socioeconomically disadvantaged students throughout the applicant pool. But the effect is most pronounced at the top of the pool, among the ablest students. Figure 7 shows the effect on first-generation college applicants at UC.

¹⁸ Geiser, 2014: 8-9.

¹⁹ Kidder, W., & Rosner, J. (2002). "How the SAT creates built-in headwinds: An educational and legal analysis of disparate impact."



When applicants are ranked by test scores instead of grades, first-generation college applicants are much less likely to place within the upper strata of the pool, where the chances of admission are greatest. SAT/ACT scores systematically disfavor the ablest applicants—as measured by their achievement in school—who have had less opportunity to learn.

For this reason, norm-referenced admissions tests are ill-suited to the mission of public universities like UC. Public institutions have a special obligation to expand opportunity for able students from disadvantaged backgrounds. SAT/ACT scores create built-in deterrents to that mission. Only by making special accommodations to compensate for disparities in opportunity to learn, such as admissions preferences for low-income or first-generation college students, can their use be justified.

Conclusion: Eliminating the SAT and ACT in UC Admissions

National standards for fairness in testing oblige test users to be vigilant about the differential impact of test scores on racial and ethnic minorities, beyond what may be warranted by test validity. Until 1998, UC met this obligation by means of a two-tiered admissions process. The top half of the pool was admitted based on grades and test scores only. The bottom half was admitted using a combination of academic and supplemental criteria, including race. Though far from ideal, two-tiered admissions did allow sensitivity to the differential validity of SAT/ACT scores for underrepresented minority applicants.²⁰

All of this changed with SP-1 and Proposition 209, barring use of race as a supplemental admissions criterion. But that change has also effectively barred consideration of how other admissions criteria, like SAT/ACT scores, are themselves affected by race. The correlation between race and test scores has grown substantially among UC applicants over the past 25 years, mirroring the growing concentration of Latino and black students in California's poorest, most intensely segregated schools. Statistically, underrepresented minority status is now a stronger predictor of SAT/ACT scores than either family income or parents' education.

UC is thus faced with a choice of some consequence. One is to continue to employ a selection criterion with known collateral effects on underrepresented minorities, even while admissions officials are prevented by law from acting on that knowledge. Continuing to use the SAT and ACT under the constraints of Proposition 209 means accepting adverse impacts on black and Latino applicants beyond what can be justified by test validity.

The alternative is to eliminate use of SAT/ACT scores in UC admissions. If the university cannot legally consider race as a socioeconomic disadvantage in admissions, neither should it consider scores on nationally normed tests. Race-blind implies test-blind admissions.

The question remains: How would UC replace standardized test scores in its admissions process? Following are thoughts on four possible paths.

²⁰ The two-tiered admissions selection process ended in 2001, when the UC regents, on the recommendation of the faculty and the president, mandated comprehensive review for all UC applicants.

1. Test-optional: One possibility is the “test optional” approach taken by a number of US colleges. In fact, the term is associated with a number of related, but distinct alternatives including “test-flexible” and “test-blind” admissions policies. Test-optional allows applicants to choose whether to submit SAT/ACT scores and have them considered in the admit decision. Test-flexible allows students to pick from a larger smorgasbord of tests. Test-blind eliminates any consideration of test scores in admit decisions.

Test-optional admissions is the most widely employed approach at US colleges and has been most studied (Soares, 2012). As of this writing, about half of national liberal arts colleges ranked in the “Top 100” by *U.S. News and World Report* do not require ACT or SAT scores from all or most applicants. The list includes Bates, Bowdoin, Bryn Mawr, Holy Cross, Pitzer, Smith, and Wesleyan. Wake Forest is the first national university ranked in the top 30 to go test-optional.

Although disputed by the College Board and ACT, most independent evaluations of test-optional admissions are generally positive (e.g., Bates College, 2004; McDermott, 2008; Bryn Mawr College, 2009). Test-optional colleges report increases in application and enrollment rates for both low income and underrepresented minority students, though the effect appears to be greater under a test-blind approach (Espenshade & Chung, 2012). At the same time, test-optional institutions report little if any change in college outcomes such as grades and graduation rates. An added attraction is that test-optional boosts SAT/ACT averages among applicants who do submit scores, a fact that has not gone unnoticed at colleges concerned to improve their institutional ranking.

Whether test-optional might work at UC is an open question. Test-optional does nothing to lessen the built-in advantage for those who do submit scores (and who have usually enjoyed the greatest opportunity to learn). For those who do not submit, it is unclear how UC would replace test scores as a tool for sorting the large volume of applications it receives. Also of concern is the potential for “gaming” the system that test-optional creates, both for institutions as well as students.

2. Strivers approach: Another possible path is the Strivers approach, that is, controlling statistically for the effects of socioeconomic circumstance on SAT/ACT scores. The most thorough-going effort to assess how such an approach might work at UC is a 2001 UCOP study by Roger Studley. Based on a sample of California SAT takers who applied to UC, the study adjusted students’ test scores to account for socioeconomic circumstance: neighborhood of residence (as defined by zip code), high school attended, and family characteristics including income, parents’ education, and whether English was the students’ first language. These factors were used to generate a predicted SAT score for each student. Students whose actual scores exceeded their predicted score were deemed high-ability applicants, as in the Strivers proposal.

The results showed that adjusting SAT/ACT scores for SES substantially reduced score deficits for Latino and black students relative to other applicants. To a lesser extent this was also true for Asian students, who tended to score higher than white students in similar circumstances (Studley, 2001). Of course, class-based adjustments cannot eliminate racial gaps entirely, as noted earlier. But a Strivers-type approach might reduce those gaps in significant measure.

3. Curriculum-based achievement tests: Until 2012, UC was one of the few universities outside the Ivy League to require both curriculum-based achievement tests as well as nationally normed tests like the SAT and ACT. During the 1930s, the College Board developed a series of multiple-choice tests to replace its older, written exams. These later became known as the SAT IIs and are now officially called the SAT Subject Tests. They are hour-long assessments that measure mastery of college-preparatory subjects and are offered in 20 subject areas, like chemistry or literature.

A student may sit for up to three subject tests at a time at one flat rate. Though used primarily to assess students’ present level of achievement rather than predict future performance, UC validity studies have long shown that performance on a battery of three achievement tests predicts student outcomes at least as well as SAT/ACT scores.

In his 2001 speech to the American Council on Education, UC president Richard Atkinson argued that, as well as offering the same or better prediction, curriculum-based achievement tests were preferable to the SAT on educational grounds. Unbeknownst to Atkinson, BOARS had been moving independently to the same conclusion.²¹ Atkinson’s speech triggered an extraordinary period of discussion, research, and policy development under BOARS chair Dorothy Perry. Among the products of this work were the holistic-review policy for admissions selection and the Top 4%/ELC policy for eligibility.

In 2002, BOARS issued what was believed to be the first statement of principles by any US university on admissions testing. Among the more important principles were: Admissions tests should measure students’ mastery of college-preparatory subjects.

²¹ Douglass, 2007:214-234; Pelfrey, 2012:115-138.

Tests should align with and support teaching of those subjects in California high schools. Tests should have a reasonable level of predictive validity. They should be fair across demographic groups. They should have diagnostic value in helping students identify specific areas of strength and areas where they need to devote more study. Above all, admissions tests should provide an incentive for students to take rigorous coursework in high school (BOARS, 2002).

Applying these principles, BOARS concluded that curriculum-based achievement tests like the SAT II Subject Tests were the clear choice over tests of general analytic ability or “aptitude” like the SAT and ACT. As for Atkinson, the decisive consideration for BOARS was not prediction but the educational value of achievement tests and their better alignment with the needs of students and schools. For students, low SAT/ACT scores signaled lack of individual ability, as against broader factors such as unequal access to good schools and well-trained teachers. The message could be damaging to self-esteem and academic aspiration. For schools, norm-referenced admissions tests incentivized test-prep activities outside the classroom. Their implicit message was that test-taking skills learned outside the classroom could override achievement in school.

Achievement tests created better alignment between instruction and assessment, in BOARS’ view. On achievement tests, the best test-prep was regular classroom instruction. A low score on an achievement test meant that the student had not mastered the required content. This might be due to inadequate instructional resources and inferior teaching—or lack of hard work on the part of the student. Achievement tests focused attention on determinants of student performance that were alterable, at least in principle, and so were better suited to the needs of students and schools:

Aptitude-type tests send the message that academic success is based in some part on immutable characteristics that cannot be changed and are, therefore, independent of good study skills and hard work. In contrast, a policy requiring achievement tests reinforces the primary message that the University strives to send to students and schools (and that is embedded in the recent decision to adopt comprehensive admissions review for all applicants): *the best way to prepare for post-secondary education is to take a rigorous and comprehensive college-preparatory curriculum and to excel in this work*. This is what the University’s coursework (“A-C”) and scholarship (GPA) requirements articulate. An appropriate battery of achievement tests would reinforce this message—independently and in a way that is consistent across all schools and test-takers (BOARS, 2002:iii, italics in original).

BOARS’ 2002 statement suggests a possible path for UC today: replace the nationally normed tests with a battery of curriculum-based achievement exams. The main argument against achievement exams like the SAT Subject Tests is that the pool of students who take them is much smaller than the pool of SAT/ACT takers.²² Except for a handful of selective colleges, few US universities any longer require subject tests, and the pool of test takers is less demographically diverse. It was for this reason that BOARS dropped the SAT Subject Tests from UC eligibility requirements in 2012 as part of a larger package of policy changes aimed at improving eligibility rates among socioeconomically disadvantaged students.²³

Yet the smaller, less diverse pool of achievement-test takers is not a fatal flaw, and there are steps universities can take to remedy the problem. When UC introduced the ELC program in 2001, there was concern that top 4% students in low-performing schools might not take the three Subject Tests then required to complete UC eligibility. UC mounted an aggressive outreach effort to maximize test participation, writing to each ELC student informing them of UC’s test requirements and encouraging them to take the tests. The result was a substantial increase in test-taking rates in schools that previously had sent few students to the university. ELC stimulated approximately 2,000 new UC applicants in the first year, over half of whom were underrepresented minorities. Virtually all the new applicants had taken the SAT II Subject Tests (UCOP, 2002).

UC is one of the biggest consumers of test scores and so has an outsize influence on the testing market. From that standpoint, the relatively small size of the pool who take the Subject Tests could prove an advantage. Were UC again to require a battery of Subject Tests to replace the SAT and ACT, in a stroke it would become the largest single user of Subject Tests in the US. Its market position could provide leverage to move those tests more fully in the direction BOARS envisioned in its statement of principles in 2002: Aligning UC admissions tests as closely as possible with classroom instruction to reinforce teaching and learning of a rigorous college-preparatory curriculum in all California schools.

²² In 2017, about 219,000 students took at least one SAT Subject Test, and a total of 542,000 subject tests were taken, compared to about 1.8 million who took the SAT (Jaschik, 2017). Other curriculum-based achievement tests that might be considered are the Advanced Placement and International Baccalaureate exams, which are associated with educational programs of the same names. However, program participation is sharply skewed along socioeconomic lines, limiting the usefulness of AP or IB exams for general admissions purposes as opposed to college placement (Geiser & Santelices, 2006).

²³ The major change in UC eligibility in 2012 was to increase the percentage of students Eligible in the Local Context from 4% to 9% of each California high school. A new eligibility category of Entitled to Review was also created.

4. Eliminate SAT/ACT scores in admissions selection: A final path is to eliminate use of test scores in admissions decisions. This is not as draconian a step as might first appear. Eliminating tests in admissions need not mean eliminating them entirely. Test scores are employed at two main decision points at UC: eligibility and admissions selection. SAT/ACT scores could continue to be used in determining students' eligibility for the UC system, if not their campus of choice. Unlike the test-optional approach, all UC applicants would still be required to take the tests. But test scores would be removed from the applicant files reviewed by admissions readers, in the same way that race was removed from admissions files after Proposition 209 took effect in 1998. Admit decisions would be based on the thirteen other selection criteria that UC employs.

Discontinuing SAT/ACT scores in admissions selection but continuing their use in eligibility might permit a natural experiment at one or more UC campuses. Though not used for admissions purposes, test-score data would be available for all applicants and admits at those campuses. This would facilitate in-depth analysis of the impact of the change on both the quality of admissions decisions and the demographic composition of the admitted pool. Should the experiment prove unsuccessful, it would be a relatively straightforward matter to reintroduce SAT/ACT scores into the campus admissions process.

While it is true that the SAT and ACT raise many of the same concerns for eligibility as for admissions selection, the two contexts differ in important ways. In eligibility, test scores affect a relatively small number of students at the margin of the pool, those with very low GPAs on the UC Eligibility Index. In admissions selection, test scores are used to rank applicants throughout the entire pool, so that many more students are affected. Another difference is that eligibility is based on just two other factors, high school grades and class rank, the latter of which is itself based on grades. Incorporating test scores in the UC Eligibility Index arguably adds more predictive value there than in holistic review, where so much more information is available on each applicant.

Using the SAT and ACT in admissions selection is another matter. More than any other selection criterion, norm-referenced admissions tests magnify economic, educational, and especially racial/ethnic differences among UC applicants. The effect is evident across the entire applicant pool but is most pronounced at the top, among the ablest applicants as demonstrated by achievement in high school. SAT/ACT scores create built-in head winds to admission of top black and Latino high school graduates, especially at UC's most selective campuses. The adverse racial impact of the tests has grown far out of proportion to their limited utility to predict how students will perform at UC.

Since Proposition 209 took effect two decades ago, the University of California has taken extraordinary steps to mitigate the adverse effects of the national admissions tests on underrepresented minority applicants. In addition to class-based admissions preferences, those steps have included holistic review and Eligibility in the Local Context. They have proven only partially successful. The next step is to replace or eliminate SAT/ACT scores in admissions selection.

REFERENCES

- Agronow, S. (2008). Examining the predictive value of the SAT II subject exams in the prediction of first-year UC GPA – A report to BOARS. Appendix IV (pp. 88-132) in BOARS' revised "Proposal to reform UC's freshman eligibility policy." Oakland, CA: University of California Academic Senate. Accessed on March 22, 2016 from <http://senate.universityofcalifornia.edu/underreview/sw.rev.eligibility.02.08.pdf>.
- Ashsenas, J., Park, H., & Pearce, A. (2017). Even with affirmative action, blacks and Hispanics are more underrepresented at top colleges than 35 years ago. *New York Times*, August 24, 2017, p. 1.
- Atkinson, R.C. (2001). Standardized tests and access to American universities. The 2001 Robert H. Atwell Distinguished Lecture, delivered at the 83rd Annual Meeting of the American Council on Education, Washington, DC. Accessed on March 22, 2016 from <http://rca.ucsd.edu/speeches/satspch.pdf>.
- Atkinson, R., & Geiser, S. (2009). Reflections on a century of college admissions tests. *Educational Researcher*, 38(9), 665-676.
- Bates College. (2004). SAT study: 20 years of optional testing. Institutional research paper. Lewiston, ME: Bates College.
- Board of Admissions and Relations with Schools. (2002). The use of admissions tests by the University of California. Accessed on November 24, 2017 from http://senate.universityofcalifornia.edu/_files/committees/boars/admissionstests.pdf.
- Bowen, W., Chingos, M., & McPherson, M. (2009). *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton, NJ: Princeton University Press.
- Bryn Mawr College. (2009). Bryn Mawr adopts new testing policy promoting greater flexibility and emphasis on subject mastery. Accessed June 23, 2009 from <http://news.brynmawr.edu/?p=2973>.
- Card, D., & Rothstein, J. (2006). Racial segregation and the black-white score gap. NBER Working Paper 12078. Cambridge, MA: National Bureau of Economic Research.
- Carnevale, A., & Haggihat, E. (1998). Selecting the strivers: A report on the preliminary results of the ETS 'Educational Strivers' study. Pp. 122-128 in *Hopwood, Bakke, and Beyond: Diversity on Our Nation's Campuses*, edited by D. Bakst. Washington, DC: American Association of Collegiate Registrars and Admissions Officers.
- Cook, M. & Evans, W. (2000). Families or schools? Explaining the convergence in white and black academic performance. *Journal of Labor Economics* 18(4), 729-54.

- Corcoran, S., & Evans, W. (2008). The role of inequality in teacher quality. Chapter 6 in *Steady Gains and Stalled Progress: Inequality and the Black-White Test Score Gap*, edited by K. Magnuson and J. Waldfogel. New York: Russell Sage Foundation.
- Douglass, J. A. (2007). *The Conditions of Admission: Access, Equity, and the Social Contract of Public Universities*. Stanford, CA: Stanford University Press.
- Espenshade, T., & Chung, C. (2012). Diversity outcomes of test-optional policies. Chapter 12 in Soares, J. (editor), *SAT Wars: The Case for Test-Optional College Admissions*. New York and London: Teachers College Press.
- Geiser, S. (2015). The growing correlation between race and SAT scores: New findings from California. Research and Occasional Paper Series No. 10-54 (October, 2015). Berkeley, CA: Center for Studies in Higher Education. Accessed April 1, 2016 from http://www.cshe.berkeley.edu/sites/default/files/shared/publications/docs/ROPS.CSHE_10.15.Geiser.RaceSAT.10.26.2015.pdf.
- Geiser, S. (2016). A proposal to eliminate the SAT in Berkeley admissions. Research and Occasional Papers Series No. 4-16 (May 2016). Berkeley, CA: Center for Studies in Higher Education.
- Geiser, S., & Studley, R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, 8, 1-16.
- Geiser, S., & Santelices, V. (2006). The role of Advanced Placement and honors courses in college admissions. Chapter 4 in Gandara, P., Orfield, G., and Horn, C. (editors), *Expanding Opportunity in Higher Education: Leveraging Promise*. Albany, NY: SUNY Press.
- Geiser, S., & Santelices, V. (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. Research and Occasional Paper Series No. 9-07 (June 2017). Berkeley, CA: Center for Studies in Higher Education.
- Haveman, R., Sandefur, G., Wolfe, B., and Voyer, A. (2004). Trends in children's attainments and their determinants as family income inequality has increased. In *Social Inequality*, edited by K. Neckerman. New York: Russell Sage Foundation.
- Haveman, R., & Wolfe, B. (1995). The determinants of children's attainments: A review of methods and findings. *Journal of Economic Literature* 23(4):1829-78.
- Jaschik, S. (2017). The slow, steady erosion of SAT Subject Tests. *Inside Higher Education* (October 23, 2017). Accessed on November 24, 2017 from <https://www.insidehighered.com/admissions/article/2017/10/23/admissions-officials-consider-impact-erosion-sat-subject-tests>.
- Jencks, C., & Phillips, M. (1998). Introduction. In *The Black-White Test Score Gap*, edited by C. Jencks and M. Phillips. Washington, D.C.: Brookings Institution Press.
- Kahlenberg, R. (2014a). No longer black and white: Why liberals should let California's affirmative action ban stand. *Slate* (online magazine) (March, 2014). Accessed on October 25, 2017 from http://www.slate.com/articles/news_and_politics/jurisprudence/2014/03/california_affirmative_action_ban_why_liberals_should_let_it_stand.html.
- Kahlenberg, R. (2014b). Introduction. Chapter 1 in *The Future of Affirmative Action: New Paths to Higher Education after Fisher v. University of Texas*, edited by R. Kahlenberg. New York: Century Foundation Press.
- Kidder, W., & Rosner, J. (2002). How the SAT creates built-in-headwinds: An educational and legal analysis of disparate impact, 43, *Santa Clara L. Rev.* 131. Accessed on January 6, 2015 from <http://digitalcommons.law.scu.edu/lawreview/vol43/iss1/3>.
- Magnuson, K., and Duncan, G. (2006). The role of family socioeconomic resources on racial test score gaps. *Developmental Review* 26(4), 365-399.
- Magnuson, K., & Waldfogel, J. (2008). Introduction. In *Steady Gains and Stalled Progress: Inequality and the Black-White Test Score Gap*, edited by K. Magnuson and J. Waldfogel. New York: Russell Sage Foundation.
- Magnuson, K., Rosenbaum, D., & Waldfogel, J. (2008). Inequality and black-white achievement trends in the NAEP. Chapter 1 in *Steady Gains and Stalled Progress: Inequality and the Black-White Test Score Gap*, edited by K. Magnuson and J. Waldfogel. New York: Russell Sage Foundation.
- Marcus, A. (1999). New weights can alter SAT scores: Family is factor in determining who's a 'striver'. *Wall Street Journal*, August 31, 1999, pp. B1, B8. Accessed on June 27, 2015 from <http://info.wsj.com/classroom/archive/wsje.99nov.casestudy.html>.
- McDermott, A. (2008). Surviving without the SAT. *The Chronicle of Higher Education*. (September 25, 2008). Accessed from <http://chronicle.com/article/Surviving-Without-the-SAT/18874>.
- Meyers, M., Rosenbaum, D., Ruhm, C., & Waldfogel, J. (2004). Inequality in early childhood education and care: What do we know? In *Social Inequality*, edited by K. Neckerman. New York: Russell Sage Foundation.
- Orfield, G., & Ee, J. (2014). Segregating California's Future: Inequality and its Alternative 60 Years after Brown v. Board of Education. Los Angeles, CA: UCLA Civil Rights Project. Accessed on August 19, 2015 from <http://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/segregating-california2019s-future-inequality-and-its-alternative-60-years-after-brown-v.-board-of-education>.
- Patterson, B.F., Mattern, K.D., & Kobrin, J. (2009). Validity of the SAT for predicting FYGPA: 2007 SAT validity sample (College Board Statistical Report No. 2009-1). New York: The College Board.
- Pelfrey, P. (2012). *Entrepreneurial President: Richard Atkinson and the University of California, 1995-2003*. Berkeley and Los Angeles: University of California Press.
- Phillips, M., Crouse, J., & Ralph, J. (1998). Does the black-white test score gap widen after children enter school? Chapter 7 in *The Black-White Test Score Gap*, edited by C. Jencks and M. Phillips. Washington, DC: Brookings Institution Press.
- Phillips, M., and Chin, T. (2000). School inequality: What do we know? In *Social Inequality*, edited by K. Neckerman. New York: Russell Sage Foundation.
- Rothstein, J. (2004). College performance predictions and the SAT. *Journal of Econometrics*, 121, 297-317.
- Rothstein, R. (2014). The racial achievement gap, segregated schools, and segregated neighborhoods – A constitutional insult. *Race and Social Problems* 6 (4). Accessed on August 19, 2015 from <http://www.epi.org/publication/the-racial-achievement-gap-segregated-schools-and-segregated-neighborhoods-a-constitutional-insult/>.

- Santelices, M.V., & Wilson, M. On the relationship between differential item functioning and item difficulty: An issue of methods? Item response theory approach to differential item functioning. *Educational and Psychological Measurement*, 72(1), 5–36.
- Soares, J. (2012). *SAT Wars: The Case for Test-Optional Admissions*. New York and London: Teachers College Press.
- Steele, C., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology* 69(5):797-811.
- Studley, R. (2001). Inequality, student achievement, and college admissions: A remedy for underrepresentation. Pp. 321-343 in R. Zwick (editor), *Rethinking the SAT: The Future of Standardized Testing in University Admissions*. New York and London: RoutledgeFarmer.
- University of California Academic Senate. (2014). Guidelines for Implementation of University Policy on Undergraduate Admissions. UC Academic Senate: Oakland, CA. Accessed on March 9, 2016 from http://senate.universityofcalifornia.edu/committees/boars/GUIDELINES_FOR_IMPLEMENTATION_OF_UNIVERSITY_POLICY_on_UG_ADM_Revised_January2014.pdf.
- University of California Office of the President. (1977). Final Report on the Task Force for Undergraduate Admissions. Oakland, CA: Author.
- University of California Office of the President. (2002). Eligibility in the Local Context program evaluation report. Report prepared for May 2002 Regents meeting. Oakland, CA: Author. <http://www.universityofcalifornia.edu/regents/regmeet/may02/304attach.pdf>.
- Wilson, W. (1998). The role of the environment in the black-white test score gap. Chapter 15 in *The Black-White Test Score Gap*, edited by C. Jencks and M. Phillips. Washington, D.C.: Brookings Institution Press.