

Norm-Referenced Tests and Race-Blind Admissions

The Case for Eliminating the SAT and ACT at the University of California

Saul Geiser

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 Table 1.1. The findings shown here are based on an initial sample of 1,144,047 California high school graduates who applied for freshman admission at UC between 1994 and 2011. The sample was provided to the author by UC Office of the President (UCOP). Subsequently, UCOP institutional research staff replicated and extended the author's findings for the years 2012 through 2016, expanding the total sample to over 1.6 million applicants. Findings for 2012-2016 reported in later figures and tables below were kindly provided by UCOP research staff.

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

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

Table 1.1. 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

Note. UC Corporate Student System is source of data on all California residents who applied for freshman admission between 1994 and 2011 for whom complete data were available on all covariates.

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.

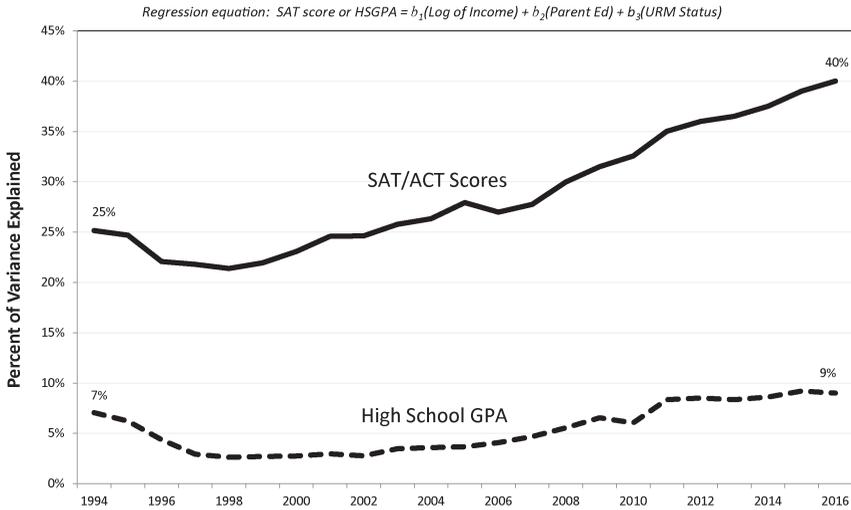
Yet the balance of costs and benefits has changed over time, and the uneasy détente over testing has become more difficult to sustain. This is evident at the University of California. Over the past 2 decades, the correlation between test scores and socioeconomic factors has increased sharply among UC applicants, as shown in Figure 1.1.

In 1995, family income, parental education, and race/ethnicity together accounted for 23% of the variance in test scores among UC applicants, compared to 5% of the variance in high school grades. By 2016, those same factors accounted for 40% of the variance in SAT/ACT scores, compared to 9% for high school GPA. What this means is that 40% of the variation in students’ SAT/ACT scores is attributable to differences in socioeconomic circumstance. It also means that test scores have become a deterrent to admission of poor, first-generation college, and Latinx 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

Figure 1.1. Variance in SAT/ACT Scores and High School GPA Explained by Family Income, Parents' Education, and Race/Ethnicity, 1994 to 2016



Note. UC Corporate Student System is source of data on all California residents who applied for freshman admission from 1994 through 2016 for whom complete data were available on all covariates.

backgrounds. In addition to holistic review, UC introduced the Eligibility in the Local Context (ELC) program in 2001, extending eligibility for admission to the top 4% 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% 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 deemphasis 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 U.S. Supreme Court guidelines reiterated most recently in *Fisher v. University of Texas* (2016).

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 (e.g., see Kahlenberg, 2014b).

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% of UC undergraduates were eligible for Pell grants (federal financial aid for low-income students) and 44% 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.

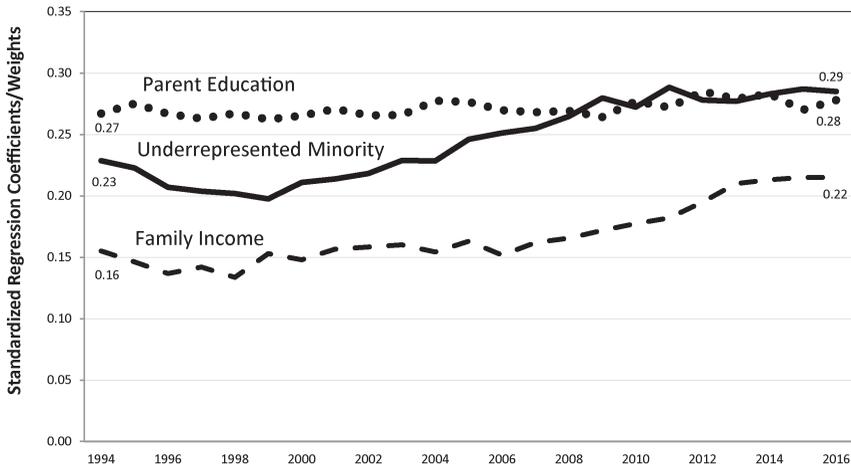
But the racial impact has been limited. Underrepresented minority enrollments—Black, Latinx, and American Indian students—plummeted after Proposition 209 took effect at UC in 1998 and did not return to pre-209 levels until 2005. Today, underrepresented groups account for about 28% of all new UC freshmen, up from about 20% 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 Latinx. 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 Chicanx and Latinx students increased by 68%, 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 VERSUS 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 Latinx 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

Figure 1.2. Relative Weight of Family Income, Parents' Education, and Race/Ethnicity in Predicting SAT/ACT Scores, 1994 to 2016



Note. UC Corporate Student System is source of data on all California resident freshman applicants from 1994 through 2016 for whom complete data were available on all covariates. All coefficients are statistically significant at .001 confidence level.

factors. Race matters as much as, if not more than, family income and parents' education in accounting for test score differences.

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 1.2 shows standardized regression coefficients for family income, parental education, and race/ethnicity in predicting test score variation among UC applicants. 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 1995, at the beginning of the period covered by this analysis, the predictive weight for parental education was 0.28. This means that for each one standard-deviation increase in parents' education, SAT/ACT scores rose by .28 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 0.16 in 1995 to 0.22 in 2016. The latter represents a test score differential of about 40 points.

But the most important trend has been the growing statistical effect of race and ethnicity. The standardized coefficient for race increased from 0.22 in 1995 to 0.29 in 2016 (about a 55-point test score differential), the last year for which the author has obtained data. Underrepresented minority status—self-identification as Latinx or African American—became the strongest predictor of test scores for the first time in 2009.³

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 Latinx parents is as much of a disadvantage on the SAT or ACT as being born to poor or less-educated parents.

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.

Over 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 Chicax and Latinx 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 (Magnuson & Duncan, 2006; Magnuson, Rosenbaum, & Waldfogel, 2008), parental education (Cook & Evans, 2000; Haveman & Wolfe, 1995), and quality of schools (Corcoran

& Evans, 2008) 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 (Magnuson et al., 2008; Phillips, Crouse, & Ralph, 1998).

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 (Haveman, Sandefur, Wolfe, & Voyer, 2004; Meyers, Rosenbaum, Ruhm, & Waldfogel, 2004; Phillips & Chin, 2004).⁴

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 U.S. Supreme Court’s decision in *Brown v. Board of Education* (1954), racial segregation in U.S. 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 (2006) study of SAT score gaps in metropolitan areas across the United States is generally regarded as dispositive: “We find robust evidence that the Black–White test score gap is higher in more segregated cities” (p. ii).

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 (Orfield & Ee, 2014). In 1993 about half of California’s public schools were still majority White schools, and only one-seventh were “intensely segregated,” meaning 90% or more Latinx and Black. By 2012, 71% of the state’s schools had a majority of Latinx and Black students and 31% 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 Latinx students, it ranks last. Today the typical Latinx student attends schools where less than 6% of the other students are White.

A clear pattern of co-segregation of Latinx and African American students has emerged. Slightly more than half of all Latinx students now attend intensely segregated schools, and the comparable figure for Black students in California is 39%. 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 Latinx attend schools that are more than 70% poor. Double segregation by race and poverty is overlaid with “triple segregation” by language, as these same schools enroll high proportions of English language learners. On top of all, intensely segregated schools are typically underresourced 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 Latinx 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 Latinx 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 (ETS) proposed a new test score measure, known as the “Strivers” proposal (Carnevale & Haghghat, 1998). ETS researchers compared students’ 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” (quoted in 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 Academic Senate, 2014, p. 3)

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 “preference” in public education 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 “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 U.S. universities. UC experimented with the SAT as early as 1960, when it required the test on a trial basis to evaluate its effectiveness. In 1962 Charles Jones, chair of the Board of Admissions and Relations with Schools (BOARS),⁵ came to the following conclusion: “Extensive analysis of the data, 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 (quoted in Douglass, 2007, p. 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 U.S. 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’s decision was to reject the tests (Douglass, 2007, pp. 91–92).⁶

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 (CPEC). 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, pp. 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 were admitted based on a 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 (see Table 1.2 for study results).

Adding students’ scores from the SAT into the prediction model improved the explained variance to 20.8%. SAT scores provided an

Table 1.2. 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%

Note. Data were taken from “UC and the SAT: Predictive Validity and Differential Impact of the SAT I and SAT II at the University of California,” by S. Geiser(with R. Studley, 2002, *Educational Assessment*, 8(1), p. 6 (Table 2). Results were based on a sample of 77,893 first-time freshmen entering UC from 1996 through 1999.

incremental gain, in other words, of about five percentage points, close to what BOARS had found in 1964.

The 2001 study replicated another result from BOARS’s 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 (2001) proposed eliminating generic, norm-referenced tests like the SAT or ACT in favor of curriculum-based achievement exams like the SAT II Subject Tests.⁷

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’s 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, p. 1).⁸

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 U.S. 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 4 state higher education systems, the late William G. Bowen and his colleagues found that

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)⁹

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. Socioeconomic status (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 (2004) has found that the proxy effect is substantial:

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. (p. 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 test scores *uniquely* add to the prediction of student success at UC?

An answer is provided in a 2007 study by institutional researchers Sam Agronow and Roger Studley of UC Office of the President. In a regression model predicting first-year grades, the researchers entered all available data from the UC application. In addition to high school GPA and test scores (including SAT II Subject Test scores as well as scores on the SAT), those variables included: number of UC-required “a-g” courses taken, number of AP or honors courses taken, class rank in high school, family income, parental education, race/ethnicity, 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, the researchers found that they explained 28.6% of the variance in UC freshman grades. When they eliminated test scores from the model, thus isolating their effect, the explained variance fell to 25.4%. Test scores, including SAT II Subject Tests as well as the SAT I, accounted for about 3% of the variance in students’ first-year grades at UC.¹⁰

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

Some may argue 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 tiebreaker to choose between applicants who are otherwise equally qualified, and in this case the decision 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% 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 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 a 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.

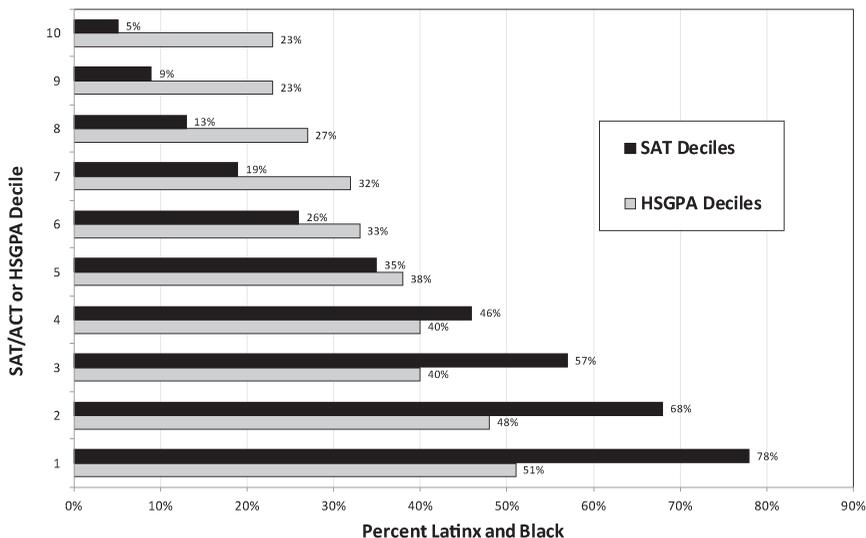
The problem is that the effects of the SAT and ACT are nonrandom. When used as a tool to rank applicants, SAT/ACT scores create greater racial stratification than other admissions criteria. More information is not always better than less if the new information introduces systematic bias in admissions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

NORM-REFERENCED TESTS AND RACIAL STRATIFICATION

Norm-referenced tests like the SAT and ACT magnify racial stratification among UC applicants, as Figure 1.3 will illustrate. All UC applicants were first ranked into 10 equal groups, or deciles, based on their high school grades. The same students were then ranked into deciles based on their SAT/ACT scores, and the proportion of Latinx and Black students in both distributions was compared.

The difference is dramatic. Although SAT/ACT scores and high school GPA both create racial stratification, the demographic footprint of the tests

Figure 1.3. Proportion of Black and Latinx Applicants by SAT/ACT vs. High School GPA Deciles



Note. UC Corporate Student System is source of data on all CA resident freshman applicants from 2016.

is far more pronounced. At the bottom of the pool, underrepresented minorities comprise 78% of the lowest SAT/ACT decile, compared to 51% of the lowest HSGPA decile. Conversely, at the top of the pool—those most likely to be admitted—Latinx and Blacks make up 23% of the top HSGPA decile but just 5% of the top SAT/ACT decile. Underrepresented minority applicants are far less 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, multistep review process that test publishers now follow.

Despite this, there are grounds for believing that exams like the SAT or ACT may unintentionally disadvantage Latinx 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, p. 159)

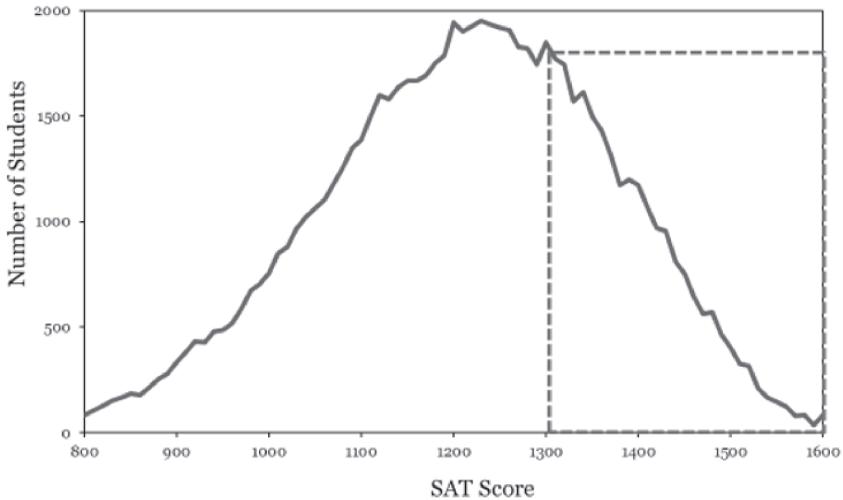
One measure of bias is known as *differential item functioning*, or DIF. DIF occurs “when equally able test takers differ in their probabilities of answering a test item correctly as a function of group membership” (AERA/APA/NCME, 2014, p. 51).¹¹ Although test developers go to great lengths to eliminate 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 Latinx 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, p. 23).¹²

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

Figure 1.4. Frequency Distribution of Test Scores Among California SAT Takers



Source: College Board College-Bound Seniors Report for California (n.d.).

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 incorrect answers, or *distractors*, to make multiple-choice items harder, and by the *speededness* of the test, or 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.

The resulting bell curve is especially useful for admissions officers at highly competitive institutions like UC (see Figure 1.4), 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 (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) jointly issued the latest revision of their *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. “Group differences in testing outcomes should trigger heightened scrutiny for possible sources of test bias” (AERA/APA/NCME, 2014, p. 54). 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, p. 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 Latinx examinees (Santelices & Wilson, 2012).

Test context is another issue for underrepresented minority students:

Multiple 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, p. 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, p. 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 Latinx undergrads (Geiser, 2015, pp. 8–9). 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 Latinx, 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 U.S. 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” (AERA/APA/NCME, 2014, 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*:

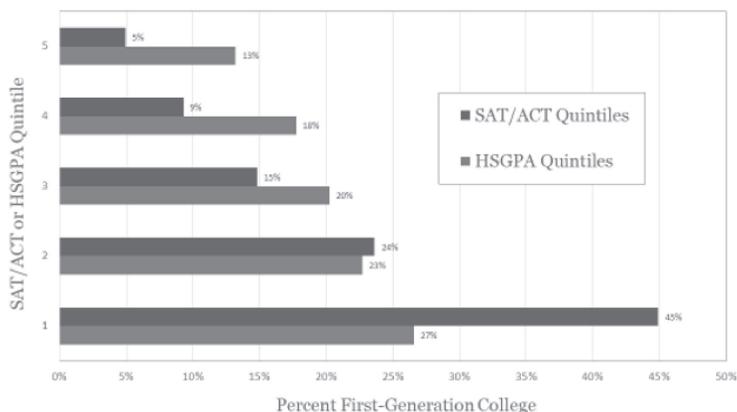
It should be noted that concerns about opportunity to learn do not necessarily apply to situations where the same authority is not responsible for both the delivery of instruction and the testing and/or interpretation of results. For example, in college admissions decisions, opportunity to learn may 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). (AERA/APA/NCME, 2014, 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 4-year graduation rates. Another issue is the size of the effect of opportunity to learn on test scores. Student socioeconomic characteristics now account for about 40% 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 head winds” (Kidder & Rosner, 2002). that reduce the chances of admission for socioeconomically disadvantaged students throughout the applicant pool. But the effect is most pronounced

Figure 1.5. Percent First-Generation College Applicants by SAT/ACT vs. HSGPA Quintiles



Note. UC Corporate Student System is source of data on all California residents who applied for freshman admissions between 1994 and 2011.

at the top of the pool, among the ablest students. Figure 1.5 shows the effect on first-generation college applicants at UC.

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.¹³

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 Latinx 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 Latinx applicants beyond what is warranted 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.

1. Test-Optional

One possibility is the "test-optional" approach taken by a number of U.S. 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 U.S. colleges and has been most studied (Soares, 2012). About half of national liberal arts colleges ranked in the "Top 100" by *U.S. News and World Report* no longer require ACT or SAT scores from all or most applicants. The list includes Bates, Bowdoin, Bryn Mawr, Holy Cross, Pitzer, Smith, and Wesleyan. As of this writing, the University of Chicago, ranked number 6 among national universities, is the first elite U.S. university to go test-optional.

Although disputed by the College Board and ACT Inc., most independent evaluations of test-optional admissions are generally positive (e.g., Bryn Mawr College, 2009; McDermott, 2008; see also Chapters 6 and 7 of this volume). 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 UC Office of the President (UCOP) study by Roger Studley (2004). 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 Latinx 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, 2004). 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 hourlong assessments that measure mastery of college-preparatory subjects and are offered in 20 subject areas, like chemistry or literature. 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 (Douglass, 2007, pp. 214–234; Pelfrey, 2012, pp. 115–138). 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 U.S. university on admissions testing. Among the more important principles was: Admissions tests should measure students' mastery of college-preparatory subjects. 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. 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, in BOARS' view, created better alignment between instruction and assessment. 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-G") 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, p. iii, italics in original).

The BOARS 2002 statement suggests a possible path for UC today: replace the nationally normed tests with a battery of curriculum-based achievement exams. The SAT Subject Tests remain one possibility, but another possibility has emerged in the years since 2002: employing K–12 standards-based exams in university admissions.

The best examples of pure achievement tests now available are employed not in U.S. higher education but in our K–12 public schools: standards-based assessments developed by the various states as part of the movement to articulate clearer curriculum standards, teach to the standards, and assess student achievement against those standards. The schools are well ahead of colleges and universities in this regard.

In 2014, California adopted one such exam, the Smarter Balanced Assessment, as part of its effort to align K–12 assessment and accountability policies with newly adopted Common Core curriculum standards. Smarter Balanced includes an 11th-grade summative exam, and sufficient time has now passed for an initial look at whether that exam might also prove useful for purposes of university admissions. Preliminary results are quite positive. Smarter Balanced performs as well as the SAT in predicting both first-year grades and second-year persistence at UC, yet produces a more socioeconomically and racially diverse applicant pool (Kurlaender & Cohen, 2019).

The idea of adapting K–12 achievement tests for use in college admissions has obvious attractions. In the ideal case, such exams could serve the dual purpose of certifying both achievement in high school and UC eligibility. The burden on students and the testing time they must endure could be greatly reduced. College entrance criteria would be directly aligned with high school curricula, and the signal to students would be clear and unequivocal: Working hard and performing well in one's high school coursework is the most direct route to college.

This is surely an attractive vision. Yet a key question remains unanswered. Standards for what is expected of entering freshmen at highly selective institutions like UC are different and usually much more rigorous than K–12 curriculum standards. They overlap, to be sure, but they are not the same, and it is unclear whether any one exam can serve both institutional purposes equally.

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 key decision points at UC. *Eligibility* involves assessing whether students are qualified for the UC system, if not their campus of choice. *Admissions selection* involves selecting from among qualified students at oversubscribed UC campuses that have more eligible applicants than places available.

Were the SAT/ACT eliminated in admissions selection, standardized tests could continue to be used in determining students' eligibility for the UC system. Unlike the test-optional approach, all UC applicants would still be required to submit test scores. But 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 13 other selection criteria that UC employs.

Discontinuing standardized tests 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.

BOARS's 2002 testing policy implies a change not only in the *kinds* of tests that UC employs but *how* tests are used and for what purpose. Norm-referenced exams like the SAT or ACT are designed to compare and rank applicants. Criterion- or content-referenced exams like Smarter Balanced, on the other hand, are designed to assess whether students have mastered prerequisite subjects and skills. As James Crouse and Dale Trusheim (1988) put it in their classic brief, *The Case Against the SAT*, standardized tests are much more effective as a tool for "certifying competence"—assessing whether students are academically qualified—than for rank-ordering them on their predicted probability of earning a higher freshman GPA. Tests are much better at certification than prediction.

The clear implication of BOARS's 2002 statement in favor of curriculum-based "achievement" exams is that standardized tests are more appropriate for UC eligibility than for admissions selection. In fact, neither norm-referenced exams like the SAT/ACT nor content-referenced exams like Smarter Balanced possess sufficient predictive validity to justify their adverse socioeconomic and racial impact in admissions selection. Though

content-referenced exams have less adverse impact than the SAT/ACT, that impact is still considerable, and predictive validity is too weak to draw fine distinctions among applicants without unwarranted adverse impacts on low-income, first-generation, and underrepresented minority applicants. Simply replacing the SAT/ACT with content-referenced exams in admissions selection is not necessarily the answer. Moving to content-referenced assessments, and using them primarily for purposes of certification, is the logical conclusion of BOARS's testing principles.

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.

NOTES

1. California is predominantly an "SAT state." Fewer students take the ACT than the SAT, although that number has grown in recent years. 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" (Ashseenas, Park, & Pearce, 2017).

3. 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.

4. 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., Kahlenberg, 2014a; Wilson, 1998.

5. The Board of Admissions and Relations with Schools, known as BOARS, is the university-wide faculty committee responsible for formulating UC admissions policy.

6. 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 should not be surprising that prior performance in such activities would be predictive of later performance.

7. See Atkinson & Geiser (2009) for an extended discussion of the preferability of criterion-referenced over norm-referenced assessments in university admissions.

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

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

10. Compare regression models 17 and 22 in Agronow & Studley, 2007, Table 1.

11. 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).

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

13. 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.

REFERENCES

- Agronow, S. & R. Studley. (2007, November 4). *Prediction of college GPA from new SAT scores – A first look. A paper presented at the annual meeting of the California Association for Institutional Research*. Retrieved from cair.org/wp-content/uploads/sites/474/2015/07/Agronow.pdf.
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Ashsenas, J., Park, H., & Pearce, A. (2017, August 24). Even with affirmative action, Blacks and Hispanics are more underrepresented at top colleges than 35 years ago. *New York Times*, p. 1.
- Atkinson, R. C. (2001, February 18). *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. Retrieved from rca.ucsd.edu/speeches/satpoch.pdf

- Atkinson, R. C., & Geiser, S. (2009). Reflections on a century of college admissions tests. *Educational Researcher*, 38(9), 665–676. doi.org/10.3102/0013189X09351981
- Board of Admissions and Relations with Schools (BOARS). (2002). *The use of admissions tests by the University of California. Executive summary*. Retrieved from senate.universityofcalifornia.edu/_files/committees/boars/testsummary.pdf
- Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2009). *Crossing the finish line: Completing college at America's public universities*. Princeton, NJ: Princeton University Press.
- Brown v. Board of Education*, 347 US 483 (1954).
- Bryn Mawr College. (2009, June 3). Bryn Mawr adopts new testing policy promoting greater flexibility and emphasis on subject mastery. Retrieved from news.blogs.brynmawr.edu/2009/06/03/bryn-mawr-adopts-new-testing-policy-promoting-greater-flexibility-and-emphasis-on-subject-mastery/
- Card, D., & Rothstein, J. (2006). *Racial segregation and the Black–White test score gap* (NBER Working Paper No. 12078). Cambridge, MA: National Bureau of Economic Research. Retrieved from www.nber.org/papers/w12078
- Carnevale, A., & Haggihat, E. (1998). Selecting the strivers: A report on the preliminary results of the ETS “Educational strivers” study. In D. Bakst (Ed.), *Hopwood, Bakke, and beyond: Diversity on our nation's campuses* (pp. 122–128). 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. In K. Magnuson & J. Waldfogel (Eds.), *Steady gains and stalled progress: Inequality and the Black–White test score gap* (pp. 212–249). New York, NY: Russell Sage Foundation.
- Crouse, J. & D. Trusheim. (1988). *The case against the SAT*. Chicago, IL: University of Chicago Press.
- Douglass, J. A. (2007). *The conditions of admission: Access, equity, and the social contract of public universities*. Stanford, CA: Stanford University Press.
- Espenshade, T. J., & Chung, C. Y. (2012). Diversity outcomes of test-optional policies. In J. A. Soares. (Ed.), *SAT wars: The case for test-optional college admissions* (pp. 177–200). New York, NY: Teachers College Press.
- Fisher v. University of Texas*, 579 U.S. ____ (2016).
- Geiser, S. (2015, October). *The growing correlation between race and SAT scores: New findings from California* (Research & Occasional Paper

- Series: CSHE.10.15). . Berkeley, CA: Center for Studies in Higher Education. Retrieved from cshe.berkeley.edu/publications/growing-correlation-between-race-and-sat-scores-new-findings-california-saul-geiser
- Geiser, S. (with 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), 1–26.
- Geiser, S., & Santelices, M. V. (2007, June). *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: CHSE.6.07). Berkeley, CA: Center for Studies in Higher Education. Available at cshe.berkeley.edu/publications/validity-high-school-grades-predicting-student-success-beyond-freshman-year-high-school
- Haveman, R., Sandefur, G., Wolfe, B., & Voyer, A. (2004). Trends in children's attainments and their determinants as family income inequality has increased. In K. Neckerman (Ed.), *Social inequality* (pp. 149–188). New York, NY: Russell Sage Foundation.
- Haveman, R., & Wolfe, B. (1995). The determinants of children's attainments: A review of methods and findings. *Journal of Economic Literature*, 33(4), 1829–1878.
- Jencks, C., & Phillips, M. (1998). Introduction. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 1-51). Washington, DC: Brookings Institution Press.
- Kahlenberg, R. D. (2014a). Introduction. In R. D. Kahlenberg (Ed.), *The future of affirmative action: New paths to higher education diversity after Fisher v. University of Texas*. New York, NY: Century Foundation Press.
- Kahlenberg, R. D. (2014b, March 11). No longer Black and White: Why liberals should let California's affirmative-action ban stand. *Slate*. Retrieved from slate.com/news-and-politics/2014/03/california-affirmative-action-ban-why-liberals-should-let-it-stand.html
- Kidder, W. C., & Rosner, J. (2002). How the SAT creates built-in headwinds: An educational and legal analysis of disparate impact. *Santa Clara Law Review*, 43(1/3), 131–211. Retrieved from digitalcommons.law.scu.edu/lawreview/vol43/iss1/3
- Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *Validity of SAT for predicting first-year college grade point average* (College Board Research Report No. 2008-5). New York, NY: The College Board. Retrieved from eric.ed.gov/?id=ED563202
- Kurlaender, M. & Cohen, K. (2019, March). Predicting college success: How do different high school assessments measure up? Retrieved from edpolicyinca.org/publications/predicting-college-success-how-do-different-high-school-assessments-measure-2019
- Magnuson, K., & Duncan, G. (2006). The role of family socioeconomic resources on racial test score gaps. *Developmental Review*, 26(4), 365–399.

- Magnuson, K., Rosenbaum, D., & Waldfogel, J. (2008). Inequality and Black–White achievement trends in the NAEP. In K. Magnuson & J. Waldfogel (Eds.), *Steady gains and stalled progress: Inequality and the Black–White test score gap* (pp. 33–65). New York, NY: Russell Sage Foundation.
- Marcus, A. (1999, August 31). New weights can alter SAT scores: Family is factor in determining who’s a ‘striver.’ *Wall Street Journal*, pp. B1, B8. Retrieved from info.wsj.com/classroom/archive/wsje.99nov.casestudy.html
- McDermott, A. B. (2008, September 25). Surviving without the SAT. *The Chronicle of Higher Education*. Retrieved from www.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 K. Neckerman (Ed.), *Social inequality* (pp. 223–270). New York, NY: Russell Sage Foundation.
- Orfield, G., & Ee, J. (2014, May 14). Segregating California’s future: Inequality and its alternative 60 years after Brown v. Board of Education. In *The Civil Rights Project / Proyecto Derechos Civiles*. Retrieved from 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/
- Pelfrey, P. A. (2012). *Entrepreneurial president: Richard Atkinson and the University of California, 1995–2003*. Berkeley, CA: University of California Press.
- Phillips, M., & Chin, T. (2004). School inequality: What do we know? In K. Neckerman (Ed.), *Social inequality* (pp. 467–520). New York, NY: Russell Sage Foundation.
- Phillips, M., Crouse, J., & Ralph, J. (1998). Does the Black–White test score gap widen after children enter school? In C. Jencks & M. Phillips (Eds.), *The Black–White test score gap* (pp. 229–272). Washington, DC: Brookings Institution Press.
- Rothstein, J. M. (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). Retrieved from www.epi.org/publication/the-racial-achievement-gap-segregated-schools-and-segregated-neighborhoods-a-constitutional-insult/
- Santelices, M. V., & Wilson, M. (2012). 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. A. (Ed.). (2012). *SAT wars: The case for test-optional college admissions*. New York, NY: 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. (2004). Inequality, student achievement, and college admissions: A remedy for underrepresentation. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 321–343). New York, NY: RoutledgeFarmer.
- Wilson, W. (1998). The role of the environment in the Black–White test score gap. In C. Jencks & M. Phillips (Eds.), *The Black–White test score Gap* (pp. 501–510). Washington, DC: Brookings Institution Press.