

Reflections on a Century of College Admissions Tests

Richard C. Atkinson and Saul Geiser

The College Boards started as achievement tests designed to measure students' mastery of college preparatory subjects. Admissions testing has significantly changed since then with the introduction of the Scholastic Aptitude Test, Lindquist's creation of the ACT, renewed interest in subject-specific assessments, and current efforts to adapt K–12 standards-based tests for use in college admissions. We have come full circle to a renewed appreciation for the value of achievement tests. Curriculum-based achievement tests are more valid indicators of college readiness than other tests and have important incentive or signaling effects for K–12 schools as well: They help reinforce a rigorous academic curriculum and create better alignment of teaching, learning, and assessment along the pathway from high school to college.

Keywords: achievement; admissions; assessment; colleges; educational policy; testing

Standardized testing for college admissions has seen extraordinary growth over the past century and appears to be on the cusp of still more far-reaching changes. Fewer than 1,000 examinees sat for the first College Boards in 1901. Today more than 1.5 million students take the SAT, 1.4 million sit for the ACT, and many students take both. This does not count many more who take preliminary versions of college entrance tests earlier in school, nor does it include those who take the SAT Subject Tests and Advanced Placement (AP) exams. Admissions testing continues to be a growth industry, and further innovations such as computer-based assessments with instant scoring, adaptive testing, and “noncognitive” assessment are poised to make their appearance.

Despite this growth and apparent success, the feeling persists that all is not well in the world of admissions testing. College entrance tests and related test preparation activities have contributed mightily to what has been called the “educational arms race”—the ferocious competition for admission at highly selective institutions (Atkinson, 2001). Many deserving low-income and minority students are squeezed out in this competition, and questions about fairness and equity are raised with increasing urgency. The role of the testing agencies themselves has also come into question, and some ask whether the testing industry holds

too much sway over the colleges and universities it purports to serve. Underlying all of these questions is a deeper concern that the current regime of admissions testing may impede rather than advance our educational purposes.

This article reflects on the first century of admissions testing with a view to drawing lessons that may be useful as we now contemplate the second. Our aim is not to extrapolate from the past or to predict the specific forms and directions that admissions tests may take in the future. Rather, our intent is to identify general principles that may help guide test development going forward.

Putting Tests in Perspective: Primacy of the High School Record

A first order of business is to put admissions tests in proper perspective: High school grades are the best indicator of student readiness for college, and standardized tests are useful primarily as a supplement to the high school record.

High school grades are sometimes viewed as a less reliable indicator than standardized tests because grading standards differ across schools. Yet although grading standards do vary by school, grades still outperform standardized tests in predicting college outcomes: Irrespective of the quality or type of school attended, cumulative grade point average (GPA) in academic subjects in high school has proved to be the best overall predictor of student performance in college. This finding has been confirmed in the great majority of “predictive-validity” studies conducted over the years, including studies conducted by the testing agencies themselves (see Burton & Ramist, 2001, and Morgan, 1989, for useful summaries of studies conducted since 1976).¹

In fact, traditional validity studies tend to understate the true value of the high school record, in part because of the methods employed and in part because of the outcomes studied. Such studies usually rely on simple correlation methods. For example, they examine the correlation between SAT scores and college grades, and the size of the correlation is taken to represent the predictive power of the SAT. At most, these studies report multiple correlations involving only two or three variables, as, for example, when they examine the joint effect of SAT scores and high school grades in predicting first-year college grades (see, e.g., Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008).

But correlations of this kind can be misleading because they mask the contribution of socioeconomic and other factors to the prediction. Family income and parents' education, for example, are correlated with SAT scores and also with college outcomes, so that much of the apparent predictive power of the SAT actually

reflects the proxy effects of socioeconomic status. Princeton economist Jesse Rothstein (2004) conservatively estimates that traditional validity studies that omit socioeconomic variables overstate the predictive power of the SAT by 150%.² High school grades, on the other hand, are less closely associated with students' socioeconomic background and so retain their predictive power even when controls for socioeconomic status are introduced, as shown in validity studies that employ more fully specified multivariate regression models. Such models generate standardized regression coefficients that allow one to compare the predictive weight of different admissions factors when all other factors are held constant. Using this analytical approach, the predictive advantage of high school grades over standardized tests is more evident (Geiser, 2002; Geiser & Santelices, 2007).³

The predictive superiority of high school grades has also been obscured by the outcome measures typically employed in validity studies. Most studies have looked only at freshman grades in college; relatively few have examined longer term outcomes such as 4-year graduation or cumulative GPA in college. A large-scale study at the University of California (UC) that did track long-term outcomes found that high school grades were decisively superior to standardized tests in predicting 4-year graduation and cumulative college GPA (Geiser & Santelices, 2007). The California findings have been confirmed in a recent national study of college completion by William Bowen and his colleagues, *Crossing the Finish Line*, based on a sample of students at a broad range of public colleges and universities: "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" (Bowen, Chingos, & McPherson, 2009, pp. 113–114).

Why high school grades have a predictive advantage over standardized tests is not fully understood, as it is undeniable that grading standards differ across high schools. Yet standardized 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—term papers, labs, final exams—so it should not be surprising that prior performance in such activities would be predictive of later performance.

Whatever the precise reasons, it is useful to begin any discussion of standardized admissions tests with acknowledgment that a student's record in college preparatory courses in high school remains the best indicator of how the student is likely to perform in college. Standardized tests do add value. In our studies at the University of California, for example, we have found that admissions tests add an increment of about 6 percentage points to the explained variance in cumulative college GPA, over and above about 20% of the variance that is accounted for by high school GPA and other academic and socioeconomic factors known at point of admission (Geiser & Santelices, 2007). And tests can add value in other important ways, beyond prediction, that we shall consider later in this article.

Testing for Ability: The Saga of the SAT

The SAT, or Scholastic Aptitude Test, first made its appearance in 1926 as an alternative to the earlier College Boards. Whereas

the older tests were written, curriculum-based examinations designed to assess student learning in college preparatory subjects, the SAT promised something entirely new: an easily scored, multiple-choice instrument for measuring students' general ability or aptitude for learning (Lemann, 1999).

The similarity between the early SAT and IQ testing was not coincidental, and the two shared a number of assumptions that most now regard as problematic. The SAT grew out of the experience with IQ tests during the First World War, when 2 million men in military service were tested and assigned an IQ based on the results. The framers of those tests assumed that intelligence was a unitary, inherited attribute; it was not subject to change over a lifetime and could be measured in a single number. Although the SAT was more sophisticated from a psychometric standpoint, it evolved from the same questionable assumptions about human talent and potential.

Yet especially in the years after World War II, the idea of the SAT resonated strongly with the meritocratic ethos of American college admissions. The SAT was standardized in a way that high school grades were not, and it could be administered relatively inexpensively to large numbers of students. If aptitude for learning could be reliably measured, the SAT could help identify students from disadvantaged circumstances who were deserving of admission—thus improving access and equity in college admissions. Above all, the SAT offered a tool for prediction, providing admissions officers a means to distinguish between applicants who were likely to perform well or poorly in college. It is easy to understand why the test gained widespread acceptance in the postwar years.

The SAT has evolved considerably since that time, and both the name of the test and the terminology describing what it is intended to measure have changed. In an effort to alter the perception of the test's link to the older IQ tradition, in 1990 the College Board changed the name to the Scholastic Assessment Test and then in 1996 dropped the name altogether, so that the initials no longer stand for anything. Official descriptions of what the test is supposed to measure have also changed over the years from "aptitude" to "generalized reasoning ability" and now "critical thinking," and the test items and format have been more or less continuously revised (Lawrence, Rigol, Van Essen, & Jackson, 2003). Throughout these changes, the one constant has been the SAT's claim to gauge students' general analytic ability, as distinct from their mastery of specific subject matter, and thereby to predict performance in college.

By the end of the 20th century, however, the SAT had become the object of increasing scrutiny, partly as a result of developments at our own institution, the University of California. After Californians voted to end affirmative action in 1996, the UC system undertook a sweeping review of its admissions policies in an effort to reverse plummeting Latino and African American enrollments. What we found challenged many established beliefs about the SAT.

Far from promoting equity and access in college admissions, we found that—compared with traditional indicators of academic achievement—the SAT had a more adverse impact on low-income and minority applicants.⁴ The SAT was more closely correlated than other indicators with socioeconomic status and so tended to diminish the chances of admission for underrepresented minority applicants, who come disproportionately from

lower socioeconomic backgrounds. For example, when UC applicants were rank ordered by SAT scores, roughly half as many Latino, African American, and American Indian students appeared in the top of the applicant pool as when the same students were ranked by high school grades (Geiser & Santelices, 2007).

Another surprise was the relatively poor predictive power of the SAT (then also known as the SAT I) as compared not only with high school grades but also with curriculum-based achievement tests, such as the SAT II subject tests and AP exams, which measure students' mastery of specific subjects. The SAT I's claim to assess general analytic ability, independent of curriculum content, was long thought to give it an advantage over achievement tests in predicting how students will perform in college.

The University of California had required applicants to take both the SAT I and a battery of achievement tests since 1968 and so had an extensive database to evaluate that claim. Our data showed that the SAT I reasoning test was consistently *inferior* to the SAT II subject tests in predicting student performance, although the difference was small and there was substantial overlap between the tests. It was not the size of the difference but the consistency of the pattern that was most striking. The subject tests—particularly the writing exam—held a predictive advantage over the SAT I reasoning test at all UC campuses and within every academic discipline (Geiser, 2002).^{5,6} And in later studies we found that the AP exams, which require the greatest depth of subject knowledge, exhibited an even greater predictive advantage (Geiser & Santelices, 2006). Mastery of curriculum content, it turns out, is important after all.

Another concern with the SAT I was its lack of fit with the needs of K–12 schools. After affirmative action was dismantled, UC massively expanded its outreach to low-performing schools throughout California in an effort to restore minority admissions over the long term. At their height, before later state budget cuts, UC outreach programs were serving 300,000 students and 70,000 teachers, and UC campuses had formed school–university partnerships with 300 of the lowest performing schools in the state. College admissions criteria can have a profound influence, for good or ill, on such schools—what Michael Kirst has called a “signaling effect” (Kirst & Venezia, 2004)—and it was evident that the SAT was sending the wrong signals.

The SAT I sent a confusing message to students, teachers, and schools. It featured esoteric items, like verbal analogies and quantitative comparisons, rarely encountered in the classroom. Its implicit message was that students would be tested on materials that they had not studied in school and that the grades they achieved could be devalued by a test that was unrelated to their course work. Especially troubling, the perception of the SAT I as a test of basic intellectual ability had a perverse effect on many students from low-performing schools, tending to diminish academic aspiration and self-esteem. Low scores on the SAT I were too often interpreted as meaning that a student lacked the ability to attend the University of California, notwithstanding his or her record in high school.⁷

These concerns prompted the first author of this article to propose dropping the SAT I in favor of curriculum-based achievement tests in UC admissions (Atkinson, 2001).⁸ The University of California accounts for a substantial share of the national

market for admissions tests, and the College Board responded to our concerns with a revised SAT in 2005.

The New SAT (now also known as the SAT-R, for “reasoning”) is clearly an improvement over the previous version of the test. The SAT II writing exam has been incorporated into the test, and verbal analogies have been dropped. Instead of deconstructing esoteric analogies, students must now perform a task they will actually face in college—writing an essay under a deadline. The old SAT featured math items, such as quantitative comparisons, that were known for their trickery but required only an introductory knowledge of algebra; the New SAT math section is more straightforward and covers some higher level topics in algebra. Reports indicate that the changes have galvanized a renewed focus on math and especially writing in many of the nation's schools (Noeth & Kobrin, 2007).

Nevertheless, as an admissions test the New SAT still falls short in important respects. The New SAT has three sections: writing, mathematics, and a third called critical reading. Not surprisingly, given the University of California's earlier findings, research by the College Board shows that writing is the most predictive of the three sections. Yet College Board researchers also find that, overall, the New SAT is not statistically superior to the old test in predicting success in college: “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). This result was unexpected, given the strong contribution of the writing test and the fact that the New SAT is almost an hour longer than the old test.⁹

A possible explanation is provided by another study by three economists at the University of Georgia (Cornwell, Mustard, & Van Parys, 2008). That study found that adding the writing section to the New SAT has rendered the critical-reading section almost entirely redundant so that it does not add significantly to the prediction. The critical-reading section is essentially the same as the verbal-reasoning section of the old SAT I. It appears that the College Board was trying to have the best of both worlds. The College Board could and did tell admissions officers that the critical-reading and math sections of the New SAT were comparable to the verbal- and mathematical-reasoning sections of the old SAT I. If admissions officers disliked the New SAT, they could ignore the writing exam and then for all practical purposes the old and new SAT would be equivalent.¹⁰

A more fundamental question is what, exactly, the new test is intended to measure. The SAT's underlying test construct has long been ambiguous, and the recent changes have only added to the confusion. Although the inclusion of the writing test and some higher level math items are evidently intended to position the New SAT as more of an achievement test, its provenance as a test of general analytic ability remains evident as well. The verbal and math sections continue to feature items that are remote from what students encounter in the classroom, and the College Board has emphasized the psychometric continuity between the old and new versions of the test (Camara & Schmidt, 2006). In a phrase, the New SAT appears to be “a test at war with itself” (Geiser, 2009), and it will be interesting to see which impulse prevails in future iterations of the test.

Although a significant improvement over the old test, the New SAT remains fundamentally at odds with educational priorities along the pathway from high school to college. The New

SAT's lack of alignment with high school curricula has become especially conspicuous now that more and more states have moved toward standards-based assessments at the K–12 level. Standards-based tests seek to align teaching, learning, and assessment. They give feedback to students and schools about specific areas of the curriculum where they are strongest and weakest, providing a basis for educational improvement and reform (Darling-Hammond, 2003). Aligning admissions tests with the needs of our schools—especially schools serving populations that have been traditionally underserved by higher education—must be a priority as we look to the next generation of standardized admissions tests.

Testing for Achievement: Enter the ACT

The ACT was introduced in 1959 as a competitor to the SAT. From its inception, the ACT has reflected an alternative philosophy of college admissions testing espoused by its founder, E. F. Lindquist (1958):

If the examination is to have the maximum motivating value for the high school student, it must impress upon him the fact that his chances of being admitted to college . . . depend not only on his “brightness” or “intelligence” or other innate qualities or factors for which he is not personally responsible, but even more upon how hard he has worked at the task of getting ready for college. . . . The examination must make him feel that he has *earned* the right to go to college by his own efforts, not that he is entitled to college because of his innate abilities or aptitudes, regardless of what he has done in high school. In other words, the examination must be regarded by him as an *achievement* test. (pp. 108–109)

From our vantage half a century later, Lindquist's vision of admissions testing seems remarkably fresh and prescient. His understanding of the signaling effect of college admissions criteria for K–12 students and schools reflects a modern sensibility, as does his admonition that educators must not allow their standards to be set, by default, by the tests they use. Assessment should flow from standards, not the other way round. Lindquist's concept of achievement testing was also quite sophisticated; as against those who would caricature such tests as measuring only rote recall of facts, he insisted that achievement tests can and should measure students' reasoning skills, albeit those developed within the context of the curriculum.

Reflecting Lindquist's philosophy, the ACT from the beginning has been tied more closely than the SAT to high school curricula. The earliest forms of the test grew out of the Iowa Tests of Educational Development and included four sections—English, mathematics, social studies reading, and natural sciences reading—reflecting Iowa's high school curriculum. As the ACT grew into a national test, its content came to be based on national curriculum surveys as well as analysis of state standards for K–12 instruction. In 1989 the test underwent a major revision and the current four subject areas were introduced (English, mathematics, reading, and science), and in 2005 the ACT added an optional writing exam in response, in part, to a request from the University of California.

The ACT exhibits many of the characteristics that one would expect of an achievement test. It is developed from curriculum surveys. It appears less coachable than the SAT, and the consensus

among the test prep services is that the ACT places less of a premium on test-taking skills and more on content mastery. The ACT also has a useful diagnostic component to assist students as early as the eighth grade to get on and stay on track for college—another function that Lindquist believed an admissions test should perform (ACT, 2009b).

Yet the ACT still falls short of a true achievement test in several ways. Like the SAT, the ACT remains a norm-referenced test and is used by colleges and universities primarily to compare students against one another rather than to assess curriculum mastery. The ACT is scored in a manner that produces almost the same bell curve distribution as the SAT. It is true that the ACT also provides standards-based interpretations indicating the knowledge and skills that students at different score levels generally can be expected to have learned (ACT, 2009a). But those interpretations are only approximations and do not necessarily identify what an examinee actually knows. It is difficult to reconcile the ACT's norm-referenced scoring with the idea of a criterion-referenced assessment or to understand how one test could serve both functions equally.

The ACT lacks the depth of subject matter coverage that one finds in other achievement tests such as the SAT Subject Tests or AP exams. The ACT science section, for example, is intended to cover high school biology, chemistry, physics, and earth/space science. But the actual test requires little knowledge in any of these disciplines, and a student who is adept at reading charts and tables quickly to identify patterns and trends can do well on this section—unlike the SAT Subject Tests or AP exams in the sciences, which require intensive subject matter knowledge.

In a curious twist, the ACT and SAT appear to have converged over time. Whereas the SAT has shed many of its trickier and more esoteric item types, like verbal analogies and quantitative comparisons, the ACT has become more SAT-like in some ways, such as the premium it places on students' time management skills. It is not surprising that almost all U.S. colleges and universities now accept both tests and treat ACT and SAT scores interchangeably.

Finally, another fundamental problem for the ACT—or for any test that aspires to serve as the nation's achievement test—is the absence of national curriculum standards in the United States. The ACT has tried to overcome this problem through its curriculum surveys, but the “average” curriculum does not necessarily reflect what students are expected to learn in any given state, district, or school. The lack of direct alignment between curriculum and assessment has led the National Association for College Admissions Counseling (NACAC; 2008) to criticize the practice followed by some states, such as Colorado, Illinois, and Michigan, of requiring all K–12 students to take the ACT, whether or not they plan on attending college, and using the results as a measure of student achievement in the schools. This practice runs counter to the American Educational Research Association's guidelines on testing: “Admission tests, whether they are intended to measure achievement or ability, are not directly linked to a particular instructional curriculum and, therefore, are not appropriate for detecting changes in middle school or high school performance” (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999, p. 143).

Of course, using the ACT to assess achievement in high school is not the same as using it to assess readiness for college. But the same underlying problem—the loose alignment between curriculum and assessment—is evident in both contexts. It may be that no one test, however well designed, can ever be entirely satisfactory in a country with a strong tradition of federalism and local control over the schools. A single national achievement test may be impossible in the absence of a national curriculum.

Assessing Achievement in Specific Subjects: SAT Subject Tests and AP Exams

In place of a single test, another approach taken at some colleges and universities is to require several achievement tests in different subjects. The assessments most often used are the SAT II subject tests and AP exams.

During the 1930s, the College Board developed a series of multiple-choice tests in various subject areas to replace its older, written exams. These later became known as the SAT IIs and are now officially called the SAT Subject Tests. In 1955 the College Board introduced the Advanced Placement program and with it, the AP exams. As their name indicates, the AP exams were originally intended for use in college placement: Colleges and universities used AP exam scores mainly to award course credits, allowing high-achieving students to place out of introductory courses and move directly into more advanced college work. Over time, however, AP has come to play an increasingly important role in admissions at selective institutions, and its role in admissions is now arguably more important than its placement function.¹¹

Of all nationally administered tests used in college admissions, the SAT Subject Tests and AP exams are the best examples of achievement tests currently available. The SAT Subject Tests are offered in about 20 subject areas and the AP exams in more than 30. The SAT Subject Tests are hour-long, multiple-choice assessments, whereas the AP exams take 2 to 3 hours and include a combination of multiple-choice, free-answer, and essay questions. Students frequently sit for the tests after completing high school course work in a given subject, so that tests often serve, in effect, as end-of-course exams. Test prep services such as the Princeton Review advise students that the most effective way to prepare for subject exams is through course work, and in a telling departure from its usual services, the Review offers content-intensive coursework in mathematics, biology, chemistry, physics, and U.S. history to help students prepare for these tests (Princeton Review, 2009).

Until the SAT II Writing exam was discontinued and became part of the New SAT in 2005, the University of California had for many years required three subject tests for admission to the UC system: SAT Writing, SAT II Mathematics, and a third SAT II subject test of the student's choosing.¹² The elective test requirement was established to give students an opportunity to demonstrate particular subjects in which they excel and to assist them in gaining admission to particular majors. Students can also elect to submit AP exam scores, which, though not required, are considered in admission to individual UC campuses.¹³

The idea that students should be able to choose the tests they take for admission may seem anomalous to those accustomed to viewing the SAT or ACT as national “yardsticks” for measuring readiness for college. But the real anomaly may be the idea that

all students should take one test or that one test is suitable for all students. Our research showed that a selection of three SAT II subject tests—including one selected by students—predicted college performance better than either of the generic national assessments, although scores on all of the tests tended to be correlated and the predictive differences were relatively small. Of the individual SAT II exams, the elective SAT II subject test proved a relatively strong predictor, ranking just behind the SAT II Writing test (Geiser, 2002; Geiser & Santelices, 2007). The AP exams proved even better predictors. Although mere participation in AP classes bore no relation to performance in college, students who took and scored well on the AP exams tended to be very successful: AP exam scores were second only to high school grades in predicting student performance at the University of California (Geiser & Santelices, 2006).

Our findings in California on the superiority of achievement tests, and especially the AP exams, have been confirmed by Bowen et al.'s (2009) recent national study of college completion. Based on a large sample of students at public colleges and universities, Bowen and his colleagues found that AP exam scores were

a far better incremental predictor of graduation rates than were scores on the regular SAT/ACT and, as in the case of the SAT IIs, including this achievement-test variable in the regression equation entirely removed any positive relationship between the SAT/ACT scores and graduation rates. . . . It is also important to emphasize that achievement tests are better predictors than SAT scores for all students, including minority students and students from low-SES backgrounds. (pp. 130–131)

In the national admissions community there is growing awareness of the value of subject tests. NACAC has recently called on colleges and universities to reexamine their emphasis on the SAT and ACT and to expand use of subject tests in admissions. NACAC's commission on testing, which wrote the report, included many high-profile admissions officials and was chaired by William Fitzsimmons, dean of admissions at Harvard. The report is unusually thoughtful and worth quoting at some length:

There are tests that, at many institutions, are both predictive of first-year and overall grades in college and more closely linked to the high school curriculum, including the College Board's AP exams and Subject Tests as well as the International Baccalaureate examinations. What these tests have in common is that they are—to a much greater extent than the SAT and ACT—achievement tests, which measure content covered in high school courses; that there is currently very little expensive private test preparation associated with them, partly because high school class curricula are meant to prepare students for them; and that they are much less widely required by colleges than are the SAT and ACT. . . .

By using the SAT and ACT as one of the most important admission tools, many institutions are gaining what may be a marginal ability to identify academic talent beyond that indicated by transcripts, recommendations, and achievement test scores. In contrast, the use of . . . College Board Subject Tests and AP tests, or International Baccalaureate exams, would create a powerful incentive for American high schools to improve their curricula and their teaching. Colleges would lose little or none of the information they need to make good choices about entering classes, while benefiting millions of American students who do not enroll in highly selective colleges and positively affecting teaching and learning in America's schools (NACAC, 2008, p. 44).

The main counterargument to expanding use of such tests in college admissions is the fear that they might harm minority, low-income, or other students from schools with less rigorous curricula. Currently the SAT Subject Tests and AP exams are considered in admissions only at a few, highly selective colleges and universities so that the population of test takers is smaller, higher achieving, and less diverse than the general population that takes the SAT or ACT. The fear is that if subject tests were used more widely, students from disadvantaged schools might perform more poorly than on tests less closely tied to the curriculum.

Experience at the University of California suggests that this fear is unfounded. After introducing its Top 4 Percent Plan in 2001, which extended eligibility for admission to top students in low-performing high schools, the university saw a significant jump in the number of students in these schools who took the three SAT II subject tests that the university required. Yet low-income and minority students performed at least as well on these tests, and in some cases better, than they did on the SAT I reasoning test or ACT. Scores on the SAT II subject tests were in most cases *less* closely correlated than SAT I or ACT scores with students' socioeconomic status.¹⁴ Interestingly, the elective SAT II subject test had the lowest correlation of any exam with students' socioeconomic status, while remaining a relatively strong indicator of their performance at the University of California (Geiser, 2002).

Nevertheless, as achievement tests, the SAT Subject Tests and AP exams do have limitations. Scoring on both tests is norm referenced, despite the fact that colleges often treat them as proficiency tests (especially the AP exams, which are used for college placement as well as admissions). Oddly, for tests designed to assess curricular achievement, scores are not criterion referenced even though they are often interpreted as such.

Another issue is how well the tests actually align with high school curricula. The SAT Subject Tests and AP exams differ in this regard. The latter exams are intended primarily for students who have completed Advanced Placement courses in high school. This arrangement has both advantages and disadvantages. The advantage is that the exams are tied to the AP curriculum, but it also means that the tests are not necessarily appropriate for students who have not taken AP, thus limiting the usefulness of the exams in college admissions. Also, the AP program has come under fire from some educators who charge that, by "teaching to the test," AP classes too often restrict the high school curriculum and prevent students from exploring the material in depth; a number of leading college preparatory academies have dropped AP for that reason (Hammond, 2008).

The SAT Subject Tests, on the other hand, are not tied as directly to particular instructional approaches or curricula but are designed to assess a core of knowledge common to all curricula in a given subject area: "Each Subject Test is broad enough in scope to be accessible to students from a variety of academic backgrounds, but specific enough to be useful to colleges as a measure of a student's expertise in that subject" (College Board, 2009b). This enhances their accessibility for use in admissions, but at a cost: The SAT Subject Tests are less curriculum intensive than the AP exams, and perhaps for that reason, they are also somewhat less effective in predicting student success in college (Geiser & Santelices, 2006).

Without question, the SAT Subject Tests and AP exams have the strongest curricular foundations of any college entrance tests now available, and more colleges and universities should find them attractive for that reason. But both fall short of being fully realized achievement tests.

Adapting K–12 Standards-Based Tests for Use in College Admissions

The best examples of pure achievement tests now available are employed not in U.S. higher education but in our K–12 schools: standards-based assessments developed by the various states as part of the movement to articulate clearer standards for what students are expected to learn, teach to the standards, and assess student achievement against those standards.¹⁵ The schools are well ahead of colleges and universities in this regard. In its recent report, NACAC's commission on testing raised the possibility of adapting K–12 standards-based assessments for use in college admissions:

As one aspect of the standards movement that has swept across American elementary and secondary public education over the past quarter-century, many states now require all public high school students to take achievement-based exams at the end of high school. These tests vary in quality; the better ones, such as those in New York, include end-of-course tests that students take upon completion of specific courses. Not all state high school exams are sufficient to measure the prospect of success in postsecondary education. However, if such tests can be developed so they successfully predict college grades as well as or better than the SAT, ACT, AP, International Baccalaureate exams, and Subject Tests do, and align with content necessary for college coursework, the Commission would urge colleges to consider them in the admission evaluation process. (NACAC, 2008, p. 44)

The idea of adapting K–12 standards-based assessments for use in college admissions has obvious attractions. In the ideal case, students' performance on end-of-course tests or exit exams could serve the dual function of certifying both their achievement in high school and their readiness for college. The burden on students and the amount of testing they must endure could be greatly reduced. College entrance criteria would be aligned directly with high school curricula, and the message to students would be clear and unequivocal: Working hard and performing well in one's high school course work is the surest route to college.

This is surely a compelling and worthwhile vision. At the same time, however, there are significant obstacles to its realization. Our experience in California is not necessarily representative of other states but may help illustrate some of the difficulties involved.

In 2000 the University of California began to explore possible alternative assessments to the SAT and ACT that were more closely aligned with California's K–12 curriculum yet suitable for use in UC admissions. Some UC faculty were skeptical of this effort in view of the volatile political environment surrounding the state's K–12 assessment system, where new testing regimes came and went with alarming frequency. In 1997, however, the State Board of Education launched a major effort to articulate clear curriculum standards for the schools and to align all state tests with those standards, which seemed to promise greater stability and continuity going forward.

It soon became evident, however, that most statewide tests were inadequate for use in UC admissions. Designed to measure achievement across the entire range of the K–12 student population, the California Standards Test lacked sufficient differentiation and reliability at the high end of the achievement distribution, from which the University of California draws its students. A similar problem existed with the California High School Exit Exam, then in its planning stages: An exam designed to determine whether students meet the minimum standards required for high school graduation is unlikely to be useful in a highly selective admissions environment.

But one test did hold promise: the Golden State Examinations (GSEs), which had been established in 1983 to assess achievement in specific academic subjects. The California Department of Education, the state's K–12 administrative arm, had long championed the GSEs as part of a broader program to improve student achievement, similar to the national AP program. The exams were voluntary and geared as honors-level assessments. Matching the state's test records to our own student database, we found that GSE scores predicted first-year performance at the University of California almost as well as the SAT I reasoning test, although not nearly as well as the SAT II subject tests. Although the GSEs lacked some of the technical sophistication of the national tests, we were hopeful that those issues could be resolved; the state had contracted with ACT, Inc., to help improve the tests' psychometric quality.¹⁶

Those hopes were dashed when funding for the GSE program was eliminated from the state's 2003 budget. The test had fallen victim to political infighting between the California Department of Education, which was promoting the test, and the State Board of Education, which viewed the GSEs as a departure from its new curriculum standards. Some state education officials also viewed the University of California's efforts to adapt the GSEs for use in admissions as an incursion on the Board of Education's authority over K–12 curriculum standards.

California's experience illustrates a more general problem likely to confront efforts to develop standards-based assessments that bridge the institutional divide between state university and K–12 school systems: Standards for what is expected of entering freshmen at selective colleges and universities are different and usually much more rigorous than K–12 curriculum standards. They overlap, to be sure, but they are not the same, and institutional conflicts over standards and testing are probably inevitable for this reason. College and university faculty are right to be skeptical about using K–12 tests in admissions if it means relinquishing control over entrance standards. And it is understandable that secondary school educators are concerned that, in seeking to adapt and modify K–12 tests for use in admissions, colleges and universities may exert undue influence over curriculum standards for the schools.

A first step toward getting past this problem is for colleges and universities to band together in articulating their own standards for what is expected of entering freshmen, as distinct from high school graduates. This has occurred in California. The academic senates of the three main segments of the state's higher education system—the University of California, the California State University, and the California Community Colleges—have collaborated on a joint statement of specific “competencies” in both

English and mathematics expected of all students entering California higher education (Intersegmental Committee of the Academic Senates, 1997, 1998). The statements are intended to inform students about the preparation they will need for college beyond the minimum requirements for high school graduation, so that students do not graduate only to find themselves unready for college-level work. Although it is a useful first step, the standards have yet to result in any changes in admissions tests.

Nationally, the most ambitious effort to develop standards of college readiness is Standards for Success, a project sponsored by the American Association of Universities (AAU) and the Pew Charitable Trusts. Led by David Conley at the Center for Education Policy Research at the University of Oregon, the project convened representatives from AAU institutions to identify content standards for what students need to know to succeed in entry-level courses at those institutions. The standards covered English, mathematics, natural sciences, social sciences, second languages, and the arts. Then, in the most interesting phase of the project, researchers used the standards as a reference point to evaluate alignment of K–12 standards-based tests. The project evaluated 66 exams from 20 states, finding that although a few were closely aligned with the standards, most bore only an inconsistent relationship to the knowledge and skills needed for college (Brown & Conley, 2007).

Whether K–12 standards-based assessments can be successfully adapted for use in college admissions may depend in part on the response of the testing agencies. The Standards for Success project ended in 2003, and the standards were subsequently licensed to the College Board. The College Board has announced that the standards are now being used in reviewing test specifications for the SAT, the Preliminary SAT/National Merit Scholarship Qualifying Test, and AP exams. Like ACT, the College Board has sought to have its tests adopted by the states for assessing K–12 student achievement (Hupp & Morgan, 2008), but there is as yet no indication that the standards will be used to adapt state-level exams for admissions purposes (College Board, 2009a).

In its call for American colleges and universities to “take back the conversation” on standardized admissions testing, NACAC's (2008) blue-ribbon commission on testing had this to say about the role of the testing agencies:

Institutions must exercise independence in evaluating and articulating their use of standardized test scores. There is also a need for an independent forum for inter-institutional evaluation and discussion of standardized test use in admission that can provide support for colleges with limited resources to devote to institutional research and evaluation.

While support for validity research is available from the testing agencies, the Commission does not believe that colleges and universities should rely solely on the testing agencies for it. . . . Rather, this Commission suggests that colleges and universities create a new forum for validity research under the auspices of NACAC. Such an independent discussion might begin to address questions the Commission and other stakeholders have posed about the tests. (pp. 21, 23)

NACAC's call for independent research on admissions tests is a useful reminder that until now most research on the SAT and ACT has been conducted by the testing agencies themselves.

Much of this work is published outside the academic journals, without benefit of normal peer review, and the findings are invariably supportive of the agencies' test products. Whether or not there is an actual conflict of interest, the appearance of a conflict is inevitable, and the parallel with some recent issues in medical research is troubling.

These considerations underscore the need for colleges and universities collectively to reclaim their authority over admissions testing—and, most vitally, over the standards on which admissions tests are built. Only college and university faculty are in a position to set academic standards for what is expected of matriculants, and this critical task can be neither delegated to the schools nor outsourced to the testing agencies.

Shifting the Paradigm: From Prediction to Achievement

Looking back at the arc of admissions testing over the 20th century, the signs of a paradigm shift are increasingly apparent. Ever since the 1930s, when Henry Chauncey suggested that Carl Brigham's new Scholastic Aptitude Test could predict student success at Harvard, the idea of prediction has captivated American college admissions. The preoccupation continues to this day and still drives much research on admissions testing. Yet the preoccupation with prediction has gradually given way to another idea. Lindquist's philosophical opposition to the SAT and his introduction of the ACT, the renewed interest in subject tests at some colleges and universities, the explosion of standards-based tests in K–12 schools, and the as-yet unsuccessful efforts to adapt them for use in college admissions—all point the way to assessment of achievement and curriculum mastery as an alternative paradigm for admissions testing.

Our ability to predict student performance in college on the basis of factors known at point of admission remains relatively limited. After decades of predictive-validity studies, our best prediction models (using not only test scores but high school grades and other academic and socioeconomic factors) still account for only about 25% to 30% of the variance in outcome measures such as college GPA. This means that some 70% to 75% of the variance is unexplained. That should not be surprising in view of the many other factors that affect student performance after admission, such as social support, financial aid, and academic engagement in college. But it also means that the error bands around our predictions are quite broad. Using test scores as a tiebreaker to choose between applicants who are otherwise equally qualified, as is sometimes done, is not necessarily a reliable guide, especially where score differences are small.

Moreover, there is little difference among the major national tests in their ability to predict student performance in college. Although the New SAT, ACT, SAT Subject Tests, and AP exams differ in design, content, and other respects, they tend to be highly correlated and thus largely interchangeable with respect to prediction. It is true that subject-specific tests (in particular the AP exams) do have a statistically significant predictive advantage (Bowen et al., 2009; Geiser & Santelices, 2006), but the statistical difference by itself is too small to be of practical significance or to dictate adoption of one test over another. The argument for achievement tests is not so much that they are better predictors

than other kinds of tests but that they are no worse: "The benefits of achievement tests for college admissions—greater clarity in admissions standards, closer linkage to the high-school curriculum—can be realized without any sacrifice in the capacity to predict success in college" (Geiser, 2002, p. 25).

For these reasons, we believe that prediction will recede in importance, and other test characteristics will become more critical in designing standardized admissions tests in the future. We will still need to "validate" our tests by demonstrating that they are reasonably correlated with student performance in college; validation remains especially important where tests have adverse impacts on low-income and minority applicants. But beyond some acceptable threshold of predictive validity, decisions about what kinds of assessments to use in college admissions will be driven less by small statistical differences and more by educational policy considerations.

In contrast to prediction, the idea of achievement offers a richer paradigm for admissions testing and calls attention to a broader array of characteristics that we should demand of our tests:

1. Admissions tests should be *criterion referenced* rather than norm referenced: Our primary consideration should not be how an applicant compares with others but whether he or she demonstrates sufficient mastery of college preparatory subjects to benefit from and succeed in college.
2. Admissions tests should have *diagnostic utility*: Rather than a number or a percentile rank, tests should provide students with curriculum-related information about areas of strength and areas where they need to devote more study.
3. Admissions tests should exhibit not only predictive validity but *face validity*: The relationship between the knowledge and skills being tested and those needed for college should be transparent.
4. Admissions tests should be *aligned with college preparatory coursework*: Assessments should be linked as closely as possible to materials that students encounter in the classroom and should reinforce teaching and learning of a rigorous academic curriculum in our high schools.
5. Admissions tests should *minimize the need for test preparation*: Although test prep services will probably never disappear entirely, admissions tests should be designed to reward mastery of curriculum content over test-taking skills so that the best test prep is regular classroom instruction.
6. Finally, admissions tests should *send a signal to students*: Our tests should send the message that working hard and mastering academic subjects in high school is the most direct route to college.

The core feature of achievement testing is criterion-referenced or standards-based assessment. This approach to assessment is now widely established in the nation's K–12 schools but has yet to take hold in college admissions, where norm-referenced assessments still prevail. Norm-referenced tests like the SAT or ACT are often justified as necessary to help admissions officers sort large numbers of applicants and evaluate their relative potential for success in college.

Once started, however, norm-referenced assessment knows no stopping point. The competition for scarce places at top institutions drives test scores ever higher, and average scores for this

year's entering class are almost always higher than last year's. Tests are used to make increasingly fine distinctions within applicant pools where almost all students have relatively high scores. Small differences in test scores often tip the scales against admission of lower scoring applicants, when in fact such differences have marginal validity in predicting college performance. The ever-upward spiral of test scores is especially harmful to low-income and minority applicants. Even where these students achieve real gains in academic preparation, as measured on criterion-referenced assessments, they lag further behind other applicants on norm-referenced tests.¹⁷ The emphasis on "picking winners" makes it difficult for colleges and universities to extend opportunities to those who would benefit most from higher education. And the preoccupation with test scores at elite institutions spreads outward, sending mixed messages to other colleges and universities and to the schools.

Criterion-referenced tests, on the other hand, presuppose a very different philosophy and approach to college admissions. Their purpose is to certify students' knowledge of college preparatory subjects, and they help to establish a baseline or floor for judging applicants' readiness for college. Along with high school grades, achievement test scores tell us whether applicants have mastered the foundational knowledge and skills required for college-level work.

When we judge students against this standard, two truths become evident. First is that the pool of qualified candidates who could benefit from and succeed in college is larger than can be accommodated at selective institutions. Second is that admissions criteria other than test scores—special talents and skills, leadership and community service, opportunity to learn, and social and cultural diversity—are more important in selecting whom to admit from among this larger pool. Admissions officers often describe their work as "crafting a class," a phrase that nicely captures this meaning.

Achievement testing reflects a philosophy of admissions that is at once more modest and more expansive than predicting success in college. It is more modest in that it asks less of admissions tests and is more realistic about what they can do: Our ability to predict success in college is relatively limited, and the most we should ask of admissions tests is to certify students' mastery of foundational knowledge and skills. It is more expansive in holding that beyond some reasonable standard of college readiness, other admissions criteria must take precedence over test scores if we are to craft an entering class that reflects our broader institutional values. And beyond the relatively narrow world of selective college admissions, testing for achievement and curriculum mastery can have a broader and more beneficial "signaling effect" throughout all of education.

It is not our intention to try to anticipate the specific forms or directions that admissions testing may take in the 21st century. Yet we believe that the general principles just outlined—and the paradigmatic idea of achievement testing that unites them—will be useful and relevant as a guide for evaluating new kinds of assessments that may emerge in the future. For example, these principles lead us to be initially skeptical about efforts to develop "noncognitive" assessments for use in college admissions insofar as those efforts sometimes blur the crucial distinction between achievement and personality traits over which the student has

little control. On the other hand, notwithstanding the many difficulties involved in adapting K–12 standards-based tests for use in admissions, we conclude that this is unquestionably a worthwhile goal if it can be realized.

It should be evident that no existing admissions tests satisfy all of the principles we have outlined. Our purpose is not to endorse any particular test or set of tests but to contribute to the national dialogue about admissions testing and what we expect it to accomplish. Two decades ago in their classic brief *The Case Against the SAT*, James Crouse and Dale Trusheim (1988) argued persuasively for a new generation of achievement tests that would certify students' mastery of college preparatory subjects, provide incentives for educational improvement, and encourage greater diversity in admissions tests. What is new is that today, more than at any time in recent memory, American colleges and universities seem open to the possibility of a fresh start in standardized admissions testing.

NOTES

¹The superiority of high school grade point average (GPA) over standardized test scores in predicting college outcomes is sometimes obscured in descriptions of validity studies. For example, in a recent survey of predictive-validity studies conducted over the past several decades, College Board researchers described their findings this way:

The SAT has proven to be an important predictor of success in college. Its validity as a predictor of success has been demonstrated through hundreds of validity studies. These validity studies consistently find that high school grades and SAT scores together are substantial and significant predictors of achievement in college. In these studies, *although high school grades typically are slightly better predictors of achievement* [italics added], SAT scores add significantly to the prediction. (Camara & Echternacht, 2000)

²In a recent study sponsored by the College Board, Paul Sackett and his colleagues defend the SAT, asserting that its predictive power is not substantially diminished when controls for socioeconomic status (SES) are introduced (Sackett, Kuncel, Arneson, Cooper, & Waters, 2009). Sackett's study, however, examined the extent to which SES affected the overall, bivariate correlation between SAT scores and college outcomes (first-year college grades) but failed to consider the independent contribution of high school grades (HSGPA) and other indicators in predicting college outcomes. In real-world admissions, the key question is what SAT scores uniquely add to the prediction of college outcomes, beyond what is already provided by a student's HSGPA and other indicators. Looking at the unique portion of the variance in SAT scores—the portion *not* shared with HSGPA or other indicators—studies using more fully specified regression models have found that the predictive power of the SAT is significantly reduced when controls for SES are introduced (Geiser, 2002; Rothstein, 2004). Thus there is no actual conflict between Sackett's study and others that show that the value added by the SAT is heavily conditioned by SES, as Sackett acknowledges (personal communication, January 14, 2009).

³An example of how simple correlations can be misleading is a study cited on the College Board's website in introducing the New SAT: "In the California study, SAT scores were slightly more predictive than high school grade point average (HSGPA)" (College Board, 2009c). The study referred to was conducted at the University of California (UC). The claim that the New SAT is more predictive than HSGPA was based on the UC study's initial finding that the univariate correlation between New SAT scores and first-year college GPA (FYGPA) was slightly greater

than that between HSGPA and FYGPA (Agronow & Studley, 2007, Figure 1, Models 1 and 4). The same study, however, also presented more fully specified, multivariate regression models that allowed direct comparison of the predictive weights of HSGPA and SAT scores when both were included side-by-side in the same model along with other academic and socioeconomic factors. In the more fully specified models, HSGPA had by far the greatest predictive weight (Agronow & Studley, 2007, Table 1, Model 22).

⁴It is important to be clear about what is meant by the term *adverse impact*. Both the College Board and ACT go to great lengths to eliminate test bias, and we do not question those efforts. Notwithstanding those efforts, however, it remains the case that, compared with other admissions indicators such as high school grades and the SAT II subject tests, SAT scores are more closely correlated with measures of socioeconomic status such as family income and parental education. As a result, the latter test has a greater adverse statistical impact on underrepresented minority applicants, who come disproportionately from lower socioeconomic backgrounds.

⁵Given the highly selective nature of UC admissions, some have questioned whether range restriction might account for the diminished predictive value of the SAT I as compared with high school GPA and SAT II subject tests in the UC sample. The UC data were examined carefully for range restriction effects, however, and there was no evidence that this was the case. Comparing the variances in HSGPA, SAT I, and SAT II scores in the UC applicant pool versus the pool of admitted students, we found that HSGPA—the primary selection criterion used in UC admissions—was the most range restricted of all admissions criteria even though it retained the greatest predictive weight. Restriction on both SAT I and SAT II scores was less pronounced and quite similar. Range restriction, in short, does not appear to account for the relative predictive weights of HSGPA, SAT I, and SAT II scores found in the UC sample (Geiser, 2002, note 4; Geiser & Santelices, 2007, note xix).

⁶In an independent reanalysis of the UC data, Zwick and her colleagues found the same small but consistent predictive advantage for the SAT II subject tests (Zwick, Brown, & Sklar, 2004). The same finding was also confirmed in a 2001 College Board study of a larger sample of institutions that required both the SAT I and SAT II, including Barnard, Bowdoin, Colby, Harvard, Northwestern, and Vanderbilt, as well as four UC campuses (Bridgeman, Burton, & Cline, 2001).

⁷These and other conclusions about the problematic effects of the SAT for California's K–12 schools were summarized in a policy paper, "The Use of Admissions Tests by the University of California," adopted by the UC faculty in 2001 after intensive debate and study. The paper was one of the first comprehensive policy statements on standardized admissions tests to be adopted by a major U.S. university and strongly endorsed "curriculum-based achievement tests" over "aptitude-type" tests (University of California, 2002).

⁸For an account of events immediately leading up to and following Atkinson's 2001 address to the American Council on Education, proposing elimination of the SAT at the University of California, see "College Admissions and the SAT: A Personal Perspective" (Atkinson, 2004).

⁹College Board researchers had expected inclusion of the writing exam in the New SAT "to add modestly to the prediction of college performance when critical reading and mathematics scores are considered" (Kobrin & Kimmel, 2006, p. 7).

¹⁰In a recent article reviewing the New SAT, the authors suggested significantly reducing or even eliminating the critical-reading section, which would not only shorten the test but also possibly improve its predictive validity. Along with this shortened SAT, students might be required to take two subject tests in areas of their choosing (Atkinson & Geiser, 2008).

¹¹About 70% of all U.S. high schools now award "bonus points" for Advanced Placement (AP) classes, according to a survey by the National

Association for College Admissions Counseling (2004). This boosts students' GPAs and improves admissions profiles, and a growing number of students now enroll in AP for this reason.

¹²The University of California currently requires two SAT Subject Tests, both of which are now elective: These must be in two different areas, chosen from the following: English, history and social studies, mathematics (Level 2 only), science, or language other than English.

¹³The UC regents have recently approved a policy change that would appear to reverse that institution's long-standing reliance on achievement tests in admissions. As part of a broader set of changes in UC admissions policies, in February 2009 the regents approved a proposal to eliminate the SAT Subject Tests and require only the New SAT (or ACT with writing) for admission to the UC system beginning in 2012. Understandably, some have viewed the regents' action as an endorsement of the New SAT and a rejection of previous UC policy favoring achievement tests. But according to UC President Mark Yudof, this is not the case:

It is important to note that although the subject examinations will no longer be *required*, students for whom these tests represent an opportunity to demonstrate achievement in a particular area are still *encouraged* to take the tests. . . . Eliminating the subject exam requirement in no way validates or confirms the use of other tests like the SAT reasoning exam. (Letter to Asian Pacific Islander Legislative Caucus, February 24, 2009)

¹⁴Regarding our contention that, compared with the SAT I, curriculum-based achievement measures such as the SAT II subject tests are less affected by students' socioeconomic status (SES), one reviewer of this article objected that achievement tests are also correlated with SES. Our point, however, is not that achievement test scores are *unrelated* to SES—virtually all academic indicators are correlated with SES to one degree or another—but that achievement indicators are *less* correlated with SES compared with the SAT. The UC studies showed that high school GPA had by far the lowest correlation with measures of SES such as family income, parental education, and high school quality; the SAT I had the strongest correlation; and the SAT II subject tests fell generally in between (Geiser, 2002; Geiser & Santelices, 2007). College Board researchers have also noted the stronger association between SAT I scores and SES than between SAT II scores and SES (see Kobrin, Camara, & Milewski, 2002, Figure 1A).

¹⁵There are substantial differences among the states in the quality of their assessments and the extent to which their curriculum standards are integrated with comprehensive school reform efforts. As Linda Darling-Hammond (2003) has noted,

In a number of states, the notions of standards and "accountability" have become synonymous with mandates for student testing that are detached from policies that might address the quality of teaching, the allocation of resources, or the nature of schooling. . . . States and districts that have relied primarily on test-based accountability emphasizing sanctions for students and teachers have often produced greater failure, rather than greater success, for their most educationally vulnerable students. More successful reforms have emphasized the use of standards for teaching and learning to guide investments in better prepared teachers, higher quality teaching, more performance-oriented curriculum and assessment, better designed schools, more equitable and effective resource allocations, and more diagnostic supports for student learning. (para. 3, 6)

¹⁶For an overview of the assessments used in California secondary and postsecondary education, and the alignment (or lack thereof) between them, see Venezia (2000).

¹⁷As Darling-Hammond (2003) notes,

Use of norm-referenced tests . . . makes it impossible to gauge progress accurately, as items are removed from the test as greater numbers of students can answer them, thus guaranteeing continuing high rates of failure, especially for certain subpopulations of students. (para. 9)

One of the main problems with No Child Left Behind, she argues, is that its testing requirements “push states back to the lowest common denominator, undoing progress that has been made to improve the quality of assessments and delaying the move from antiquated norm-referenced tests to criterion-referenced systems” (para. 11)

REFERENCES

- ACT. (2009a). *College readiness standards for the ACT*. Iowa City, IA: Author. Retrieved July 26, 2009, from <http://www.act.org/standard/guides/act/index.html>
- ACT. (2009b). *Educational planning and assessment*. Iowa City, IA: Author. Retrieved July 26, 2009, from <http://www.act.org/epas/index.html>
- Agronow, S., & Studley, R. (2007, November). *Prediction of college GPA from New SAT test scores—A first look*. Paper presented at annual meeting of the California Association for Institutional Research, Monterey, CA.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Atkinson, R. (2001). *Standardized tests and access to American universities*. The 2001 Robert H. Atwell Distinguished Lecture, American Council on Education, Washington, DC. Retrieved November 18, 2009, from <http://www.rca.ucsd.edu/comments/satspch.html>
- Atkinson, R. (2004, April). *College admissions and the SAT: A personal perspective*. Invited address at the annual meeting of the American Educational Research Association, San Diego, CA. (Republished in *Journal of the Association for Psychological Science, Observer*, 18, 15–22, 2005).
- Atkinson, R., & Geiser, S. (2008). The new SAT: A work in progress. *Observer: A Journal of the Association for Psychological Science*, 21(10), 23–24.
- Bowen, W., Chingos, M., & McPherson, M. (2009). *Crossing the finish line: Completing college at America's public universities*. Princeton, NJ: Princeton University Press.
- Bridgeman, B., Burton, N., & Cline, F. (2001). *Substituting SAT II: Subject tests for SAT I: reasoning tests: Impact on admitted class composition and quality* (College Board Research Rep. No. 2001–3). New York: College Board.
- Brown, R., & Conley, D. (2007). Comparing state high school assessments and standards for success in entry-level university courses. *Educational Assessment*, 12, 137–160.
- Burton, N., & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980* (College Board Research Rep. No. 2001–2). New York: College Board.
- Camara, W., & Echternacht, G. (2000). *The SAT I and high school grades: Utility in predicting success in college* (College Board Rep. No. RN-10). New York: College Board.
- Camara, W., & Schmidt, A. (2006). *The New SAT facts [PowerPoint presentation]*. New York: College Board. Retrieved March 7, 2009, from http://www.collegeboard.com/prod_downloads/forum/forum06/the-new-sat_a-comprehensive-report-on-the-first-scores.PPT
- College Board. (2009a). *College Board standards for college success*. New York: College Board. Retrieved March 10, 2009, from <http://professionals.collegeboard.com/k-12/standards>
- College Board. (2009b). *Frequently asked questions about SAT Subject Tests*. New York: College Board. Retrieved March 6, 2009, from http://www.compassprep.com/subject_faq.shtml#faq2
- College Board. (2009c). *SAT validity studies*. New York: College Board. Retrieved July 22, 2009, from <http://professionals.collegeboard.com/data-reports-research/sat/validity-studies>
- Cornwell, C., Mustard, D., & Van Parys, J. (2008). *How does the New SAT predict academic performance in college?* (Working paper). Athens: University of Georgia. Retrieved November 18, 2009, from <http://www.terry.uga.edu/~mustard/New%20SAT.pdf>
- Crouse, J., & Trusheim, D. (1988). *The case against the SAT*. Chicago: University of Chicago Press.
- Darling-Hammond, L. (2003, February 16). Standards and assessments: Where we are now and what we need. *Teachers College Record*. Retrieved November 18, 2009, from <http://www.tcrecord.org> (ID No. 11109)
- 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, 1–26.
- Geiser, S. (2009). Back to the basics: In defense of achievement (and achievement tests) in college admissions. *Change*, 41(1), 16–23.
- Geiser, S., & Santelices, M. V. (2006). The role of Advanced Placement and honors courses in college admissions. In P. Gandara, G. Orfield, & C. Horn (Eds.), *Expanding opportunity in higher education: Leveraging promise* (pp. 75–114). Albany: State University of New York Press.
- Geiser, S., & Santelices, M.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*. Berkeley: Center for Studies in Higher Education, University of California, Berkeley. Retrieved November 18, 2009, <http://cshe.berkeley.edu/publications/publications.php?id=265>.
- Hammond, G. (2008). Advancing beyond AP courses. *Chronicle of Higher Education*, 54(34), B17.
- Hupp, D., & Morgan, D. (2008). *The SAT as a state's NCLB assessment: Rationale and issues confronted*. Paper presented at the National Conference on Student Assessment, Orlando, FL. Retrieved July 29, 2009, from the College Board website: <http://professionals.collegeboard.com/data-reports-research/cb/other-conf/nclb-state-asmnt>
- Intersegmental Committee of the Academic Senates. (1997). *Statement of competencies in mathematics expected of entering college students*. Sacramento: California Education Round Table. Available at <http://www.certicc.org>
- Intersegmental Committee of the Academic Senates. (1998). *Statement of competencies in English expected of entering college students*. Sacramento: California Education Round Table. Available at <http://www.certicc.org>
- Kirst, M., & Venezia, A. (Eds.). (2004). *From high school to college: Improving opportunities for success in postsecondary education*. San Francisco: Jossey-Bass.
- Kobrin, J., Camara, W., & Milewski, G. (2002). *The utility of the SAT I and SAT II for admissions decisions in California and the nation* (College Board Research Rep. No. 2002–6). New York: College Board.
- Kobrin, J., & Kimmel, E. (2006). *Test development and technical information on the writing section of the SAT reasoning test* (College Board Research Rep. No. RN-25). New York: College Board.
- Kobrin, J., Patterson, B., Shaw, E., Mattern, K., & Barbuti, S. (2008). *Validity of the SAT for predicting first-year college grade point average* (College Board Research Rep. No. 2008–5). New York: College Board.
- Lawrence, I., Rigol, G., Van Essen, T., & Jackson, C. (2003). *A historical perspective on the content of the SAT* (College Board Research Rep. No. 2003–03). New York: College Board.
- Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. New York: Farrar, Straus and Giroux.

- Lindquist, E. F. (1958, November 1). *The nature of the problem of improving scholarship and college entrance examinations* (Paper presented at Educational Testing Service invitational conference on testing problems). Princeton, NJ: Educational Testing Service.
- Morgan, R. (1989). *Analysis of the predictive validity of the SAT and high school grades from 1976 to 1983* (College Board Rep. No. 89-7). New York: College Board.
- National Association for College Admissions Counseling. (2004). *National school counselor survey*. Alexandria, VA: Author.
- National Association for College Admissions Counseling. (2008). *Report of the Commission on the Use of Standardized Tests in Undergraduate Admissions*. Arlington, VA: Author.
- Noeth, J., & Kobrin, J. (2007). *Writing changes in the nation's K-12 school system* (College Board Research Rep. No. RN-34). New York: College Board.
- Princeton Review. (2009). *Prep for SAT Subject Tests*. Framingham, MA: Author. Retrieved July 26, 2009, from <http://www.princetonreview.com/college/sat-subject-test-prep.aspx>
- Rothstein, J. (2004). College performance predictions and the SAT. *Journal of Econometrics*, *121*, 297-317.
- Sackett, P., Kuncel, N., Arneson, J., Cooper, S., & Waters, S. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological Bulletin*, *135*, 1-22.
- University of California. (2002). *The use of admissions tests by the University of California*. Oakland, CA: UC Board of Admissions and Relations With Schools. Retrieved November 18, 2009, from <http://www.universityofcalifornia.edu/senate/committees/boars/admissionstests.pdf>
- Venezia, A. (2000). Connecting California's K-12 and higher education systems: Challenges and opportunities. In E. Burr, G. C. Hayward, B. Fuller, & M. Kirst (Eds.), *Crucial issues in California education 2000: Are the pieces fitting together?* (pp. 153-176). Berkeley: Policy Analysis for California Education.
- Zwick, R., Brown, T., & Sklar, J. (2004). *California and the SAT: A reanalysis of University of California admissions data*. Berkeley: Center for Studies in Higher Education, University of California, Berkeley. Retrieved November 18, 2009, from <http://cshe.berkeley.edu/publications/publications.php?id=68>

AUTHORS

RICHARD C. ATKINSON is president emeritus of the University of California and professor emeritus of cognitive science and psychology at the University of California, San Diego, 5320 Atkinson Hall, 9500 Gilman Drive, La Jolla, CA 92093-0436; RCA@ucsd.edu. His research is on memory, perception, and cognition.

SAUL GEISER is a research associate at the Center for Studies in Higher Education at the University of California, Berkeley, 771 Evans Hall, No. 4650, Berkeley, CA 94720-4650; sgeiser@berkeley.edu. He is a former director of research for admissions and outreach for the University of California system.

Manuscript received May 26, 2009

Revision received August 4, 2009

Accepted August 10, 2009