

Research & Occasional Paper Series: CSHE.4.09

CSHE | Center for Studies in Higher Education

UNIVERSITY OF CALIFORNIA, BERKELEY

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

REFLECTIONS ON A CENTURY OF COLLEGE ADMISSIONS TESTS*

April 2009

Richard C. Atkinson

University of California — President Emeritus

Saul Geiser

University of California, Berkeley — Center for Studies in Higher Education

Copyright 2009 Richard C. Atkinson and Saul Geiser, all rights reserved.

ABSTRACT

Standardized testing for college admissions has grown exponentially since the first administration of the old “College Boards” in 1901. This paper surveys major developments since then: the introduction of the “Scholastic Aptitude Test” in 1926, designed to tap students’ general analytic ability; E.F. Lindquist’s creation of the ACT in 1959 as a competitor to the SAT, intended as a measure of achievement rather than ability; the renewed interest on the part of some leading colleges and universities in subject-specific assessments such as the SAT Subject Tests and Advanced Placement exams; and current efforts to adapt K-12 standards-based tests for use in college admissions. Looking back at the evolution of admissions tests, it is evident that we have come full circle to a renewed appreciation for the value of achievement tests. The original College Boards started out as achievement tests, designed to assess students’ mastery of college-preparatory subjects. A century of admissions testing has taught us that this initial premise may have been sounder than anyone realized at the time. But the journey has been useful, since we now have much better understanding of why assessment of achievement and curriculum mastery remains vital as a paradigm for admissions testing. Curriculum-based achievement tests are the fairest and most effective assessments for college admissions and have important incentive or “signaling effects” for our K-12 schools as well: They help reinforce a rigorous academic curriculum and create better alignment of teaching, learning, and assessment all along the pathway from high school to college.

Standardized testing for college admissions has seen extraordinary growth over the past century and appears on the cusp of still more far-reaching changes. Fewer than 1,000 examinees sat for the first “College Boards” in 1901. Today nearly three million high-school seniors take the SAT or ACT each year. And this does not count many more who take preliminary versions of college-entrance tests earlier in school or sit for the exams

* This paper served as the basis for an invited address by the first author at the annual meeting of the American Educational Research Association in San Diego, California on April 15, 2009. Richard C. Atkinson is president emeritus of the University of California. Saul Geiser is former director of admissions research at UC’s Office of the President and currently a research associate at the Center for Studies in Higher Education at the UC Berkeley campus.

multiple times, nor does it include those who take the SAT Subject Tests and Advanced Placement exams. Admissions testing continues to be a growth industry, and further innovations such as computer-based assessments with instant scoring, adaptive testing, and “non-cognitive” assessment are poised to make their appearance.

Despite this growth and apparent success, however, 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 one of us has 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 paper surveys 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.

Our thesis, in brief, is this: The original College Boards started out as achievement tests, designed to assess students’ mastery of college-preparatory subjects. A century of admissions testing has taught us that this initial premise may have been sounder than anyone realized at the time. After a prolonged detour with alternative approaches to admissions testing, today we have come full circle to a renewed appreciation for the value of achievement tests. But the journey has been useful, since we now have a much better understanding of *why* assessment of achievement and curriculum mastery remains vital as a paradigm for admissions testing. Curriculum-based achievement tests are the fairest and most effective assessments for college admissions and have important incentive or “signaling” effects for our K-12 schools as well: They help reinforce a rigorous academic curriculum and create better alignment of teaching, learning, and assessment all along the pathway from high school to college.

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. The reality is different. Though it is true that grading standards 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 in academic subjects in high school has proven consistently the best overall predictor of student performance in college. This finding has been confirmed in hundreds of “predictive-validity” studies conducted over the years, including studies conducted by the testing agencies themselves (see Morgan, 1989, and Burton and Ramist, 2001, for useful summaries of studies conducted since 1976).

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. Validity studies conducted by the testing agencies 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, et al., 2008).

But correlations of this kind can be misleading, since they mask the contribution of socioeconomic and other factors to the prediction. Family income and parents' education, for example, are correlated both 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 conservatively estimates that traditional validity studies that omit socioeconomic variables overstate the predictive power of the SAT by 150 percent (Rothstein, 2004).

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 regression models (Geiser with Studley, 2002; Geiser and Santelices, 2007).¹

The predictive superiority of high-school grades has also been obscured by the outcome measures typically employed in conventional validity studies. Most studies have looked only at freshman grades in college – the outcome measure that standardized admissions tests are optimized to predict — while relatively few have examined longer-term outcomes such as four-year graduation or cumulative GPA in college. A recent, large-scale study at the University of California that did track long-term outcomes found that high-school grades were decisively superior to standardized tests in predicting four-year graduation and cumulative college GPA (Geiser and Santelices, 2007).

Why high-school grades have a predictive advantage over standardized tests is not fully understood, as it is undeniable that grading standards do vary across high schools. Yet standardized test scores are based on a single sitting of three or four hours, whereas

¹ In a recent study sponsored by the College Board, Paul Sackett and his colleagues have defended the SAT, asserting that its predictive power is not substantially diminished when controls for socioeconomic status (SES) are introduced (Sackett, et al., 2009). Sackett's study, however, examined only the overall, bivariate correlation between SAT scores and college outcomes (first-year college grades) and failed to consider the independent contribution of high-school grades (HSGPA) to the prediction. 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. Sackett's study is uninformative on that question. Looking at the unique portion of the variance in SAT scores – the portion *not* shared with HSGPA – studies using more fully specified regression models have found that the predictive power of the SAT is decisively diminished when controls for SES are introduced. SES has much less of an effect on HSGPA and the variance that SAT scores share with HSGPA (Geiser with Studley, 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 has acknowledged in a personal communication with the authors.

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 they are 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 six percentage points to the explained variance in cumulative college GPA, over and above about 20 percent of the variance that is accounted for by high-school GPA and other academic and socioeconomic factors known at point of admission (Geiser and Santelices, 2007). And tests can add value in other important ways, beyond prediction, that we shall consider later in this paper. Yet after more than a century of standardized admissions testing, it bears remembering that high-school grades are still the most meaningful and effective measure of student readiness for college.

Testing for ability: The saga of the SAT

The “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 two million men were tested and assigned an IQ based on the results. The framers of those tests assumed that intelligence was a unitary, inherited attribute, that it was not subject to change over a lifetime, that it could be measured in a single number, and that individuals could be ranked and assigned their place in society accordingly. 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 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, despite inferior schooling, were nevertheless 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 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 and its 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 evolved over the years from “aptitude” to “generalized reasoning ability” and now “critical thinking,” and test items and format have been more or less continuously revised (Lawrence et al., 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 and criticism, partly as a result of developments at the authors’ 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 greater equity and access in college admissions, we found that, compared to traditional measures of academic achievement, the SAT (then also known as the “SAT I”) had a more adverse impact on low-income and minority applicants. For example, when UC applicants were rank-ordered by SAT I scores, roughly half as many Latino, African-American and American Indian students appeared in the top of the applicant pool than when the same students were ranked by high-school grades (Geiser and Santelices, 2007).

Another surprise was the relatively poor predictive power of 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 Advanced Placement exams, which measure students’ mastery of specific subjects like biology or U.S. history. The SAT’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.

UC has 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 at UC, 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 with Studley, 2002). And in later studies we found that the AP exams, which require the greatest depth of subject knowledge, had an even greater predictive advantage (Geiser and Santelices, 2006). Mastery of curriculum content, it turns out, is important after all.

Yet another concern with the SAT I was its lack of fit with the needs of our 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 and 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 coursework. 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 UC, notwithstanding his or her record of accomplishment in high school.

These concerns prompted the first author of this paper to propose dropping the SAT I in favor of curriculum-based achievement tests in UC admissions (Atkinson, 2001).² UC 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 new mathematics section is more demanding, but fairer; while the old SAT featured item-types that were known for their trickery but required only a basic knowledge of algebra, the new math section is more straightforward and covers some higher-level math. Reports indicate that the changes have galvanized a renewed focus on writing and math in many of the nation’s schools.

Yet as an admissions test, the New SAT still falls short in important respects. The New SAT is almost an hour longer, which presumably should have improved its predictive validity. Yet research by the College Board indicates that the new test is no better in predicting student success in college than the test it replaced. In a nationwide study of 110 college and universities, College Board researchers found that while the writing exam, as expected, was the most predictive of the New SAT’s three component tests (critical reading, mathematics, and writing), their overall verdict was that “... the changes made to the SAT did not substantially change how well the test predicts first-year college performance” (Kobrin, et al., 2008).

Some have been surprised by the lack of improvement in the predictive power of the New SAT, given the addition of the writing test. A possible explanation is provided by another recent study by three economists at the University of Georgia. Their study found that adding the writing exam to the SAT-R has rendered the old verbal-reasoning test

² For 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 UC, see “College admissions and the SAT: A personal perspective” (Atkinson, 2004).

(now called critical reading) almost entirely redundant, with the result that there has been no overall gain in prediction (Cornwell, Mustard and Van Parys, 2008).³

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 been at pains to demonstrate the psychometric continuity between the old and new versions of the test (Camara and 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 will prevail in future iterations of the test.

Though 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 SAT-R's lack of alignment with high-school curricula has become especially conspicuous now that most states, like California, have moved decisively 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. 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:

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 ... (Lindquist, 1958; emphasis in original).

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

³ In a recent article reviewing the New SAT, the authors have suggested significantly reducing or even eliminating the critical reading section, which would not only shorten the test but also possibly improve its predictive validity (Atkinson and Geiser, 2008).

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-sciences reading and natural-sciences reading – reflecting that state'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 important function that Lindquist believed an admissions test should perform.

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. 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 also 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 content 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 do require intensive subject-matter knowledge.

In a curious twist, the ACT and SAT appear to have converged over time. While the SAT has shed many of its trickier and more esoteric item-types, like verbal analogies, 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 U.S. 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 American Educational Research Association 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:

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. (AERA, 1999)

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 that has been 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 Advanced Placement 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.

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 over 30. The SAT subject tests are hour-long, multiple-choice assessments, while the AP exams take two to three hours and include a combination of multiple-choice, free-answer, and essay questions. Students frequently sit for the tests after completing high-school coursework 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 coursework, 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.

The University of California has for many years required three subject tests for admission to the UC system: the SAT II Mathematics exam, the SAT II Writing exam (until it was discontinued and became part of the New SAT in 2005), and a third SAT II subject test of the student's choosing. The elective test requirement was established to give students an opportunity to demonstrate particular subjects in which they excel and can 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.

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. In fact, 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. In studies of almost 125,000 students entering UC between 1996 and 2001, we found that students' combined scores on the three SAT II subject tests were consistently superior to SAT I or ACT-equivalent scores in predicting student success at all campuses and in all academic disciplines. 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 with Studley, 2002). And later research showed that the AP exams were even better predictors. Though mere enrollment in AP classes bore no relation to performance in college, students who 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 UC (Geiser and Santelices, 2006).

There is growing awareness of the value of subject tests within the national admissions community. The National Association for College Admissions Counseling has recently called on American colleges and universities to re-examine 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 and financial aid 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).

The main counter-argument 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 elite colleges and universities, so that the self-selected population of test-takers is smaller, higher achieving, and less diverse than the general population who take 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 that are tied less closely to the curriculum.

Our experience at the University of California, however, suggests that this fear is unfounded. After UC introduced its Top 4 Percent Plan in 2001, extending eligibility for admission to top students in low-performing high schools, we saw a significant jump in the number of students in these schools who took the three SAT II subject tests that UC 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. 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 UC (Geiser with Studley, 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 directly 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 general 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 schools have dropped AP for that reason.

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.

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 not employed 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. The schools are well ahead of colleges and universities in this regard. In its recent report, NACAC's commission on testing has raised the intriguing 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, IB 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).

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 coursework 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 at least help to illustrate some of the difficulties involved.

In 2000, UC began to explore possible alternative assessments to either the SAT or 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 seemed to come and go with alarming frequency. Beginning in 1997, however, the State Board of Education had launched a major effort to articulate clear curriculum standards for the schools and to align all state tests against those standards, which seemed to promise greater stability and continuity going forward.

Up until that time the main assessment in California's schools was the Standardized Testing and Reporting (STAR) Program, which utilized a nationally norm-referenced, standardized examination to test all pupils in grades two through eleven. But with the advent of state curriculum standards, the STAR program was augmented by the California Standards Test, a standards-based achievement test that assessed student learning in English-language arts, mathematics, history/social science, science, and written composition.

The idea that the state's universities might also make use of the new standards-based test was intriguing, and both UC and the California State University system held extended discussions with K-12 representatives to explore its feasibility. It soon became evident, however, that at least for UC, the California Standards Test was psychometrically inadequate for use in admissions. Designed to measure achievement across the entire range of the K-12 student population, it lacked sufficient differentiation and reliability at the high end of the achievement distribution, from which UC draws its students. California's Master Plan for Higher Education mandates that UC admit students from the top eighth of the state's high-school graduates.

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) had been established in 1983 to assess achievement in key academic subjects for students in grades seven to twelve. 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 in the schools, similar to the national AP program. The GSEs were offered in thirteen subject areas and included written-response as well as multiple-choice questions; the GSE science exams also included laboratory tasks. The exams were voluntary and geared as honors-level assessments. Students who performed well were designated Golden State Scholars, which could be helpful in gaining admission to college as well as financial aid. Although the exams had been established well before the State Board of Education's new curriculum standards, they were clearly designed as achievement tests, and the California Department of Education had taken steps to ensure that all current forms of the GSE were fully aligned with those standards.

For UC, the Golden State Exams looked promising not only from the standpoint of alignment, but also from the standpoint of predictive validity. Matching the state's test records to our own student database, we found that GSE scores predicted first-year performance at UC almost as well as the SAT I reasoning test, though not nearly as well as the SAT II subject tests. While 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.

Alas, those hopes were dashed when funding for the GSE program was eliminated in 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. It is also likely that some viewed UC's efforts to adapt the GSEs for use in admissions as an incursion on the Board's authority over K-12 curriculum standards. The one silver lining to the story is that the California State University system was able to negotiate an agreement to use the California Standards Test to identify high-school juniors needing remediation in math or English, thus providing an extra year for them to rectify their academic deficiencies before entering CSU. But the broader and more significant potential for using California's standards-based assessments for admission to either of the state's public university systems has yet to be realized.

California's experience illustrates a more general problem likely to confront any effort 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 may 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 should be 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 – UC, 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 and 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. But while a useful first step, the standards have yet to result in any actual changes in admissions tests.

Nationally, the most ambitious effort to develop standards of college readiness is “Standards for Success,” a project supported jointly by the American Association of Universities 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 in order 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 most bore only an inconsistent relationship to the knowledge and skills needed for college (Brown and Conley, in press).

The project ended in 2003, and Standards for Success was subsequently licensed to the College Board. Given the College Board's vested institutional interest in its own stable of tests, however, it is an open question whether this project, though extremely promising, will stimulate further efforts to adapt state standards-based tests for use in college admissions. The College Board indicates that the standards are now being used in reviewing test specifications for the SAT, PSAT/NMSQT and AP exams (College Board, 2009a), but it seems unlikely that the Board would actively encourage development of alternative assessments that might compete with its existing line of test products.

For it is evident that the testing industry itself is another institutional barrier to the creation of the next generation of admissions tests. In its call for American colleges and universities to "take back the conversation" on standardized admissions testing, NACAC's blue-ribbon commission on testing had this to say:

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 (NACAC, 2008).

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 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. Unlike Kuhnian paradigm shifts in the sciences, however, this shift has been gradual and almost unnoticed. 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. E.F. Lindquist's philosophical opposition to the SAT and his introduction of the ACT, the renewed interest in subject tests at leading colleges and universities, the explosion of standards-based tests in our 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.

In fact, our ability to predict student performance in college based on 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 percent of the variance in outcome measures such as college GPA. This means that some 70 to 75 percent of the variance is unaccounted for and 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, but the statistical difference by itself is too small to be of practical significance or to dictate adoption of one test over another.

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 minimum threshold of predictive validity, our decisions about what kinds of assessments to use in college admissions will be driven less by small statistical differences and much 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:

- 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.
- 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 as well as areas where they need to devote more study.

- 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.
- Admissions tests should be *aligned with high-school curricula*: Assessments should be linked as closely as possible to materials that students encounter in the classroom and should reinforce teaching and learning of rigorous college-preparatory curriculum in our schools.
- Admissions tests should *minimize the need for test preparation*: Though 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.
- Finally and most important, 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 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 K-12 standards-based assessments, they lag further behind other applicants on norm-referenced tests. 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 a relatively few, 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 far 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, economic disadvantage, and social and cultural diversity – are far 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 at the same time 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. But it also suggests a more expansive vision: 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 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 current efforts to develop “non-cognitive” 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 could 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 longstanding 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 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 (Crouse and Trusheim, 1988). 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.

REFERENCES

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. Standards for educational and psychological testing. Washington, D.C.: American Educational Research Association, 1999.
- Atkinson, R. (2001). Standardized tests and access to American universities. The 2001 Robert H. Atwell Distinguished Lecture. American Council on Education, Washington, D.C.
- Atkinson, R. (2004). College admissions and the SAT: A personal perspective. Invited address at the American Educational Research Association annual meeting, San Diego, CA, April 14, 2004. Subsequently published in *Journal of the Association for Psychological Science, Observer*, Vol. 18, pp.15-22 (2005).
- Atkinson, R. and S. Geiser. (2008). Beyond the SAT. *Forbes.com*, August 13, 2008. http://www.forbes.com/2008/08/13/college-admissions-sat-oped-college08-cx_ra_sg_0813atkinson.html.
- Brown, R. and D. Conley. (in press). Comparing state high school assessments and standards for success in entry-level university courses. *Educational Assessment*. Pre-publication version available at <http://cepr.uoregon.edu/Brown-Conley%20MS05-027.pdf>.
- Burton, N. and L. Ramist. (2001). Predicting success in college: SAT studies of classes graduating since 1980. College Board Research Report No. 2001-2. New York: College Board.
- Camara, W. and A. Schmidt. (2006). The New SAT facts. Powerpoint presentation. New York: College Board. Downloaded on March 7, 2009 from College Board website: 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. Downloaded March 10, 2009 from College Board website: <http://professionals.collegeboard.com/k-12/standards>. New York: College Board.
- College Board. (2009b). Frequently asked questions about SAT Subject Tests. Downloaded on March 6, 2009 from College Board website: http://www.compassprep.com/subject_faq.shtml#faq2.
- Cornwell, C., Mustard, D. and J. Van Parys. (2008). How does the New SAT predict academic performance in college? Working paper. Athens, Georgia: University of Georgia. <http://www.terry.uga.edu/~mustard/New%20SAT.pdf>.
- Crouse, J. and D. Trusheim. (1988). *The Case Against the SAT*. Chicago, IL: University of Chicago Press.

- Geiser, S. (2009). Back to the basics: In defense of achievement (and achievement tests) in college admissions. *Change*, vol. 41, no 1, pp.16-23.
- Geiser, S. and M. Santelices. (2006). The role of advanced placement and honors courses in college admissions. Pp. 75 -114 in P. Gandara, G. Orfield and C. Horn, (eds.), *Expanding Opportunity in Higher Education: Leveraging Promise*. Albany, NY: SUNY Press.
- Geiser, S. and M. Santelices. (2007). Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes. Center for Studies in Higher Education, UC Berkeley: <http://cshe.berkeley.edu/publications/publications.php?id=265>.
- Geiser, S. with R. Studley. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, vol. 8, no. 1, pp. 1-26.
- Intersegmental Committee of the Academic Senates. (1997). Statement of competencies in mathematics expected of entering college students. Sacramento, California: California Education Round Table. <http://www.certicc.org>.
- Intersegmental Committee of the Academic Senates. (1998). Statement of competencies in English expected of entering college students. Sacramento, California: California Education Round Table. <http://www.certicc.org>.
- Kirst, M. and A. Venezia, (eds.). (2004). *From High School to College: Improving Opportunities for Success in Postsecondary Education*. San Francisco: Jossey-Bass.
- Kobrin, J., Patterson, B., Shaw, E., Mattern, K., and S. Barbuti. (2008). Validity of the SAT for predicting first-year college grade point average. College Board Research Report No. 2008-5. New York: College Board.
- Lemann, N. (1999). *The Big Test: The Secret History of the American Meritocracy*. New York: Farrar, Straus and Giroux.
- Lindquist, E. F. (1958). The nature of the problem of improving scholarship and college entrance examinations. Paper presented at Educational Testing Service invitational conference on testing problems, November 1, 1958. Princeton, New Jersey: ETS.
- Lawrence, I., Rigol, G., Van Essen, T. and C. Jackson. (2003). A historical perspective on the content of the SAT. College Board Research Report No. 2003-03. New York: College Board.
- Morgan, R. (1989). Analysis of the predictive validity of the SAT and high school grades from 1976 to 1983. College Board Report No. 89-7. New York: College Board.

National Association of College Admissions Counseling. (2008). Report of the commission on the use of standardized tests in undergraduate admissions. Arlington, VA: NACAC

Rothstein, J. (2004). College performance predictions and the SAT. *Journal of Econometrics*, vol. 121, pp. 297-317.

Sackett, P., Kuncel, N., Anreson, J., Cooper, S., and S. Waters. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological Bulletin*, vol. 135, pp. 1-22.