

The New SAT: A Test at War with Itself

Richard C. Atkinson*

My remarks today are based on a paper co-authored with Saul Geiser who has been my colleague and collaborator on the work I have done on college admissions testing. The full text of the paper, with citations and acknowledgements is posted on the website of the Center for Studies in Higher Education at the University of California, Berkeley (<http://cshe.berkeley.edu/publications/publications.php?id=335>). My talk will hit the highlights of the paper—with the hope that it will cause you to visit the website and study the full text more carefully. The title of the paper is “Reflections on a Century of College Admissions Tests”. In your AERA program, my talk is listed as “The New SAT: A Test at War with Itself”. That title is a little catchier than the title of the paper, but it does justice to the topics I’ll cover today.

I’ll begin with a brief history of my involvement in these matters. In the early 1990s I served as chair of the Board on Testing and Assessment (BOTA)—a board of the National Research Council charged with advising the federal government and other groups on issues of testing and assessment. The defining moment for me occurred at a meeting of BOTA in Washington, D.C. where representatives of the College Board (the non-profit organization that owns the SAT) presented their views on college admissions tests. I left that meeting less than satisfied. The College Board has a superb record on the technical aspects of test development and also on administering tests and ensuring their security. But at that meeting, the notion that the SAT was a “true measure of intelligence” dominated their perspective. They seemed oblivious to research suggesting that achievement tests were a better predictor of college success than aptitude tests.

On my way back to California I stopped in Florida to visit my grandchildren. I found my granddaughter, then in 6th grade, already diligently preparing for the SAT by testing herself on several books of verbal analogies. She also had a book with lists of quite obscure words to memorize and then construct analogies using the words. I was amazed at the amount of time and effort involved, all in anticipation of the SAT. Was this how I wanted my granddaughter spending her study time?

On the plane trip back to California I drafted an op-ed piece about college admissions tests. It made a series of points. One was that admissions tests should not try to measure “innate intelligence” but should focus on achievement—what the student actually learned during the high school years. Such tests should have an essay component requiring the student to produce an actual writing sample. And the tests should cover more mathematics than an eighth grade introduction to algebra.

Finally, I said that an important aspect of admissions tests was to convey to students, as well as their teachers and parents, the importance of learning to write and the necessity of mastering a firm grounding in mathematics.

* An invited presidential address at the annual meeting of the American Educational Research Association held in San Diego, California on April 15, 2009.

When I was asked to give the keynote address at the annual meeting of the American Council of Education (ACE) in February 2001, I decided to use the op-ed draft as the basis for the speech. I won't go into the details of the ACE speech. In a nutshell, I said that I intended to recommend to the University of California (UC) faculty that we cease using the SAT and instead rely on appropriate achievement tests. The speech attracted a great deal of public attention.

The time bomb in the ACE speech was a reference to a UC study that was released a few months later. For some years, UC had required the SAT and three SAT Subject Tests (writing, mathematics and a third of the student's choice) in the admissions process. The SAT Subject Tests are achievement tests offered in about 20 areas such as Physics, Chemistry, U.S. History and Literature. By the time I gave my ACE speech, we had four years of data on all freshmen admitted and subsequently enrolled at a UC campus. We had approximately 80,000 student protocols. A protocol included the student's high school grades, SAT scores (verbal and quantitative), three SAT Subject Test scores, family income, family educational background, the quality of the student's high school, race/ethnicity, and several other variables. And, of course, the protocol included the grade record of the student in her or his freshman year at a UC campus.

In brief, the study shows that the SAT Subject Tests were a better predictor of college grades than the SAT. The combination of high school grades and the three SAT Subject Tests accounted for 22.2% of the variance in first-year college grades. When the SAT was added, the explained variance increases from 22.2% to 22.3%, a trivial increment.

The data indicate that the predictive validity of the SAT Subject Tests are less affected by differences in socioeconomic background than is the SAT. After controlling for family income and parents' education, the predictive power of the SAT Subject Tests are undiminished, whereas the relationship between SAT scores and UC grades virtually disappears.

The UC data yield another important result. Of the five admissions tests—the verbal and quantitative sections of the SAT and the three SAT Subject Tests—the best single predictor of student performance was the writing test. Given the importance of writing at the college level, it is not surprising that a test of actual writing skills correlates strongly with college grades.

Once the results of our study were fully understood, the College Board decided to replace the SAT with what I'll call the New SAT. As will be clear later, I believe the New SAT is a significant improvement over the old test, but it still falls short in important respects.

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 predictor of a student's readiness for college. Standardized admissions tests should be used, but primarily as a supplement to the high-school record.

High-school grades are sometimes viewed as less reliable than standardized tests because grading standards differ across schools. This is undeniably true, nevertheless, grades still outperform standardized tests in predicting college outcomes. This finding has been confirmed in hundreds of “predictive-validity” studies conducted over the years (see Morgan, 1989, and Burton and Ramist, 2001, for useful summaries of studies conducted since 1976).

In fact, traditional validity studies tend to underestimate the true value of the high-school record. Validity studies conducted by the testing agencies usually rely on simple correlations. At most they 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 grades in college (see, e.g., Kobrin, et al., 2008). But correlations of this kind can be misleading, since they mask the contribution of factors such as socioeconomic status (SES). Family income and parents’ education are correlated both with SAT scores and 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 (Geiser with Studley, 2002; Geiser and Santelices, 2007).¹

Why high-school grades have so great a predictive advantage over standardized admissions tests is not fully understood. Whatever the precise reasons, any discussion of standardized admissions tests should begin by acknowledging that a student’s high school record is the best indicator of how they are likely to perform in college.

Testing for ability: The saga of the SAT

The “Scholastic Aptitude Test” made its appearance in 1926 as an alternative to the original “College Boards” which were first used in 1901. Whereas the original College Boards were written curriculum-based examinations, the SAT promised something

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.

entirely new: an easily scored, multiple-choice instrument for measuring students' aptitude for learning, independent of any specific curricular materials (Lemann, 1999).

The similarity between the early SAT and IQ testing was not coincidental. The SAT grew out of the experience with IQ tests during the First World War, when over two million men in the armed forces were given IQ tests. The framers of those tests assumed that intelligence was a unitary, inherited attribute, and not subject to change over a lifetime. Although the SAT was more sophisticated from a psychometric standpoint, it was based on the same questionable assumptions about human talent and potential.

The SAT has evolved considerably since that time. 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 from the "Scholastic Aptitude Test" to the "Scholastic Assessment Test" and then in 1996 dropped the name altogether, so that the initials "SAT" 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," (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.

The "New SAT" introduced in 2005 (now also known as the "SAT-R," for "reasoning") is clearly an improvement over the previous version of the test. The writing subject test 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. The new mathematics section is more demanding, but fairer—the old SAT featured item-types that were known for their trickery but required a minimal knowledge of algebra whereas the new math section is more straightforward and covers some higher-level math. Reports from a variety of sources indicate that the changes have galvanized a renewed focus on writing and math in the nation's schools.

The New SAT has three sections: writing, mathematics and a third called critical reading. It is no surprise, given the University of California data, that recent research by the College Board shows that writing is the most predictive of the three sections. However, the College Board reports that overall the New SAT is not statistically superior to the old test in predicting success in college (Kobrin, et al., 2008). This is a remarkable result given the strong contribution of the writing test, and the fact that the New SAT is an hour longer than the old test. Increasing testing time by an hour surely should have improved the test's predictive validity.

A possible explanation is provided by another study by three economists at the University of Georgia (Cornwell, Mustard and Van Parys, 2008). This study found that the writing section of the New SAT made the critical reading section almost entirely redundant.² That is, when you knew the writing and math scores, adding the critical reading score did

² 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 possibly improve its predictive validity (Atkinson and Geiser, 2008).

not improve predictive validity. The critical reading section in the New SAT is basically the verbal-reasoning section of the old SAT. In a sense, the College Board was trying to have the best of both worlds. They could and did tell admissions officers that the critical reading and math sections of the New SAT matched the verbal-reasoning and mathematical-reasoning sections of the old SAT. If admissions officers didn't like the New SAT, they could ignore the writing test and then the old and new SAT tests would be equivalent for all practical purposes. One could say that the College Board wanted to have its cake and eat it too.

A fundamental question is what, exactly, the new test is designed to measure. Although the inclusion of the writing test and some higher-level math items are 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 critical reading and math sections continue to include items that are remote from what students encounter in the classroom, and the College Board has been at pains to demonstrate 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).

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. 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.

In our paper, there is a section devoted to the ACT College Admissions Test. The best I can do here is provide a brief summary. The ACT was introduced in 1959 as a competitor to the SAT. Its founder, E.F. Lindquist, made important contributions to the development of test theory; his conception of the ACT is captured in the following quotation:

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 seems remarkably fresh and prescient. However, as the ACT evolved into a national test it strayed from its founder's vision and in its current form falls short in several ways. It 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 quickly reading charts and tables can do well on this section—unlike the SAT Subject Tests or AP exams, 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, the ACT has become more “SAT-like” in other ways, such as the premium it places on the students’ time-management skills. It is not surprising that almost all colleges and universities now accept either test and treat ACT and SAT scores as interchangeable.

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 is to require several achievement tests in different subjects. The assessments most often used are the SAT Subject Tests and Advanced Placement exams.

As noted earlier, during the 1930s, the College Board developed achievement tests in various subject areas in addition to the SAT. These tests 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 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. Test-prep services such as the Princeton Review advise students that the most effective way to prepare for subject exams is through coursework. In a telling departure from its usual services, the Princeton Review offers content-intensive coursework in mathematics, biology, chemistry, physics, and U.S. history to help students prepare for these tests.

There is growing awareness of the value of subject tests within the national admissions community. The National Association for College Admissions Counseling (NACAC) 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 was chaired by William Fitzsimmons, dean of admissions at Harvard. The report is unusually thoughtful and worth quoting:

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 and achievement test scores. In contrast, the use of ... the 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 high school students (NACAC, 2008).

The main counter-argument to expanding use of such tests in college admissions is the fear that they might harm minority and/or low-income students from schools with less rigorous curricula. 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 Subject Tests that UC required. Low-income and minority students performed at least as well or better on the three SAT Subject Tests, than they did on the SAT or ACT. Further, scores on the SAT Subject Tests were *less* closely correlated with students' socioeconomic status than SAT or ACT scores (Geiser with Studley 2002).

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.

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. The preoccupation with prediction is gradually given way to another idea. Namely, the assessment of achievement as an alternative paradigm for admissions testing.

In fact, our ability to predict a student's college performance based on factors known at point of admission remains surprisingly limited. After decades of predictive-validity studies, our best prediction models (using not only admissions test scores but high-school grades and other academic and socioeconomic factors) still account for only about 25 percent of the variance in outcome measures such as college GPA. This means that some 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 a reliable guide, especially where score differences are small.

For the many reasons already mentioned, I believe that prediction will recede in importance, and other test characteristics will become more critical in designing standardized admissions tests. 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, 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 six characteristics that should be kept in mind as new admissions tests are developed:

- 1) To the extent possible, admissions tests should be *criterion-referenced* rather than norm-referenced: principal consideration should be whether an applicant demonstrates sufficient mastery of college-preparatory subjects to succeed in college.
- 2) Admissions tests should have *diagnostic utility*: Rather than a number or a percentile rank, tests should provide students with information about areas of strength as well as 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 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 college-preparatory courses in our high schools.
- 5) 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 and not test-taking skills, so that the best test-prep is regular classroom instruction.
- 6) 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 path to college.

I believe these six principles—and the idea of achievement testing that unites them—will be useful and relevant as a guide for evaluating new kinds of assessments that will emerge in the future. No existing admissions tests satisfy all of these principles. My

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. Today, more than at any time in recent history, American colleges and universities seem open to the possibility of a fresh start in college admissions testing.

I began this talk with a reference to my granddaughter. She was in the first group of high school students to take the New SAT. As a high school sophomore she took the PSAT—a test preparatory to taking the SAT—and did very well. She was worried that I had complicated her future by advocating for a change in the SAT. But, she did equally well on the New SAT and is now a junior in college. Her high school quickly adjusted to the proposed changes, and had students writing an essay once a week in preparation for the new test.

One of the clear lessons of history is that colleges and universities, through their admissions requirements, strongly influence what is taught in the schools. From my viewpoint, the most important reason for changing the SAT is to send a clear message to K-12 students, their teachers and parents that learning to write and mastering a solid background in mathematics is of critical importance.

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.