

## Measurement and Evaluation for Higher Education Admissions: Findings from California

### Medición y evaluación para los procesos de admisión de la educación superior: Hallazgos desde California

Saul Geiser

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

#### Abstract

This article provides an overview of the measurement and evaluation efforts that have been at the heart of the University of California's admissions system throughout the last century, with a particular focus on developments over the past 30 years. It describes the admissions experience, tensions, and challenges presented by the use of standardized testing, as well as alternative admissions paths implemented more recently. The article ends with an overview of the challenges that lie ahead for the most selective higher education institution in California; these may also be relevant to other higher education institutions around the world.

**Keywords:** university admissions, admissions testing, SAT, socioeconomic status

---

#### Post to:

Saul Geiser  
Center for Studies in Higher Education  
University of California, Berkeley, USA.  
771 Evans Hall #4650, Berkeley, CA, USA  
Email: sgeiser@berkeley.edu

---

© 2016 PEL, <http://www.pensamientoeducativo.org> - <http://www.pel.cl>

ISSN: 0719-0409      DDI: 203.262, Santiago, Chile  
doi: 10.7764/PEL.53.1.2016.19

## Resumen

El presente artículo entrega una visión general de los esfuerzos de medición y evaluación que han conformado el núcleo del sistema de admisión de la Universidad de California a lo largo del siglo pasado, enfocándose particularmente en los avances de los últimos 30 años. Describe la experiencia, las tensiones y los desafíos del proceso de admisión que se asocian con el uso de pruebas estandarizadas, así como los caminos de admisión alternativos implementados más recientemente. El artículo concluye con una visión general de los desafíos futuros para la institución de educación superior más selectiva en California, los que también pueden ser relevantes para otras instituciones de educación superior de todo el mundo.

**Palabras clave:** admisión a las universidades, pruebas de admisión, SAT, estatus socioeconómico

Over the past few decades, the University of California (UC) has been at the forefront of the national discussion about the role of standardized testing in college admissions, spurred by the marginal contribution of test scores to the prediction of college outcomes and their significant adverse impact on minority students. Motivated by the sharp decline in admission of minority students as a consequence of the ban on affirmative action approved by California's voters in 1996, UC has spearheaded the implementation of new policies and admissions processes, exploring alternatives that may prove helpful to other institutions.

This article provides an overview of that policy discussion and the UC experience in opening up policy alternatives. The article first describes the history of standardized testing in the United States and then focuses on the experience of the University of California and the SAT, the growth of criterion-referenced tests, and the lessons to be learned from policy implementation post 2014. I draw on the results from the research agenda I have developed over the past fifteen years, first as the head of the Admissions Research Office at the University of California Office of the President and then as a researcher at the Center for Studies of Higher Education at the University of California, Berkeley (Geiser, 2009, 2014, 2015; Geiser & Atkinson, 2013; Geiser & Caspary, 2005; Geiser & Santelices, 2006, 2007; Geiser & Studley, 2002).

### Standardized testing

Standardized assessment for college admissions began in 1926 with the SAT, then known as the Scholastic Aptitude Test. Until then, admissions tests in the United States had involved intensive, written examinations in different subjects. The SAT offered something entirely new: an easily-scored, multiple-choice test for measuring students' general ability or aptitude for learning.

The SAT grew out of the experience with IQ testing of soldiers during World War I. The framers of both tests shared a number of assumptions that most now regard as problematic: that intelligence was a unitary, inherited trait; that it was not subject to change over a lifetime; and that it could be measured in a single number.

Yet, the SAT was attractive to American colleges and universities for many reasons. It was standardized in a way that high-school grades were not, and so could be used to compare applicants from different high schools. According to its promoters, it could identify promising students from poorer schools who might not otherwise be admitted. Above all, the SAT offered a tool for prediction, providing admissions officers with a means to distinguish between applicants who would perform well or poorly in college. It is easy to understand why the test became popular in the years after World War II.

In keeping with its origins in IQ testing, the SAT has long been known for tricky, puzzle-type items. Figure 1 shows a controversial type of item known as *verbal analogies*, which was used until just a few years ago.

TENET: THEOLOGIAN:

- (A) predecessor: heir
- (B) hypothesis: biologist
- (C) Recluse: rivalry
- (D) Arrogance: persecution
- (E) Guitarist: rock band

Answer: (B)

Figure 1. Sample SAT verbal item (pre-2005). Source: College Board.

This item-type was much criticized because students do not normally learn verbal analogies as part of their regular high-school coursework. Though intended to measure verbal reasoning, critics contended that it only measured students' ability to pay for test-preparation services, where verbal analogies were taught. This item-type was dropped from the SAT in 2005.

### Enter the ACT

The ACT—an acronym for American College Testing—was introduced in 1959 as a competitor to the SAT. The ACT embodied an alternative approach to admissions testing. Rather than assessing students' general ability or aptitude, the ACT was designed to measure student achievement, that is, how much they had learned in school and the extent to which they had mastered different subjects. The common perception is that while the SAT measures students' innate ability, the ACT measures how much they have learned in school.

Consistent with this philosophy, the ACT is more closely aligned with high-school curricula than the SAT. Test items are developed from surveys of high-school course offerings across the U.S. The ACT appears less coachable, and the consensus is that the ACT places less emphasis on test-taking skills and more on mastery of the curriculum. In line with the ACT's philosophy, test items tend to be more straightforward than the SAT. Figure 2 shows a representative item from the mathematics section of the ACT.

A car averages 27 miles per gallon. If gas costs \$4.04 per gallon, which of the following is closest to how much the gas would cost for this car to travel 2,727 typical miles?

- A) \$44.44
- B) \$109.08
- C) \$118.80
- D) \$408.04
- E) \$444.40

Answer: (D)

Figure 2. Sample ACT Math item. Source: ACT.

Answering the item requires only a straightforward calculation. The main challenge for students is what is called the *speededness* of the test, that is, the need to answer large numbers of such items in a limited amount of time.

The SAT has long been the dominant admissions test in the United States, but the ACT has made gains in recent years, and in 2012, for the first time, more students took the ACT than the SAT. The SAT is dominant on the West and East Coasts, as well as Texas, while the ACT is predominant in Midwestern

states. California is mostly an SAT state, although the University of California also accepts the ACT. Of all the states, California is the biggest user of the SAT.

### Convergence of the SAT and the ACT

Despite their different origins, the SAT and ACT have converged over time and become more alike. The SAT has dropped many of its trickier item-types, like verbal analogies and quantitative comparisons. Meanwhile the ACT has become more speeded, putting more emphasis on students' time-management skills and quick recall—and becoming more SAT—like in that respect.

Both have also added a writing test, in response in part to the research conducted at the University of California, which showed that the writing test was among the better predictors of college performance. Both tests have about the same predictive power, and almost all American colleges and universities now accept both tests and treat SAT and ACT scores interchangeably. Most important, both are norm-referenced tests.

### Norm-versus criterion-referenced tests

Norm-referenced tests are intended to rank students, that is, to compare students with others who take the same test. They are designed to produce a bell-curve distribution.

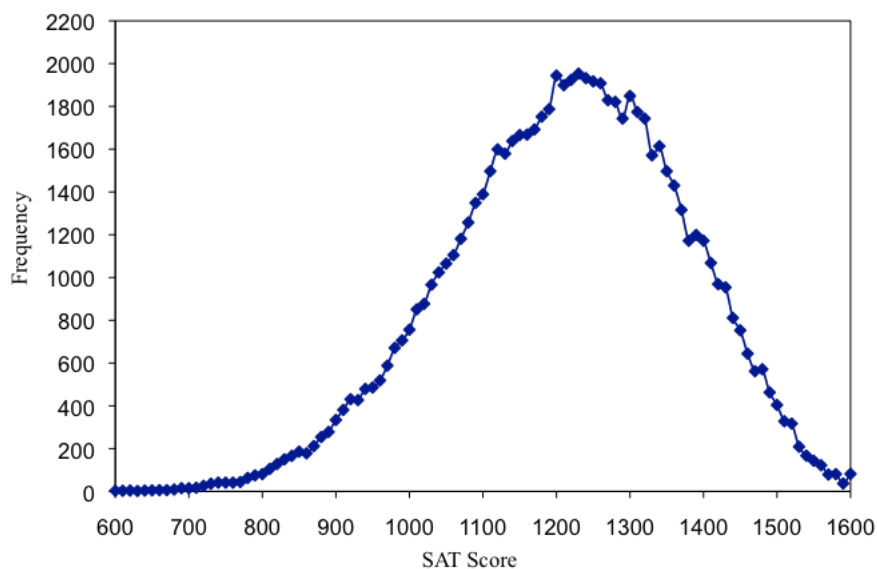


Figure 3. The bell curve: Distribution of test scores among Californian SAT takers.

The bell curve results from the way in which the test is constructed and scaled. In developing the test, the test maker discards items that too many students can answer correctly; the ideal item is one that divides the test population evenly. Then, in scaling the test, raw scores are converted into scaled scores in a manner that produces steep tails at both ends of the distribution.

There is an important difference between norm-referenced and criterion-referenced tests in this regard. Criterion-referenced assessments measure how much students know about a given subject, rather than how they compare with others. On a criterion-referenced test, it is possible for most or even all students to perform well, assuming they know the subject matter. This is not possible on a norm-referenced test because of the way exam performance is scored on a curve.

Some states in the United States use the SAT or ACT to measure educational achievement in their K-12 schools. However, because norm-referenced tests are designed always to produce a bell-curve distribution,

they are poorly suited for measuring changes in educational achievement within any population of students or schools. Test designers routinely eliminate items that too many students answer correctly, so that it is impossible for the tests reliably to detect improvement in educational achievement over time.

Despite the shortcomings of norm-referenced tests as a measure of what students have learned—as opposed to how they compare with others—colleges and universities have found them very useful for admissions purposes. This is especially true for highly selective colleges and universities that receive large numbers of applications. There, test scores provide a short-hand, seemingly objective measure of ability that makes it easy to compare and rank students. By design, norm-referenced tests highlight small differences at the upper end of the score distribution, from which universities select their students. This enables admissions officers to make fine distinctions within applicant pools where almost all students are high achievers.

### The University of California and the SAT

In 1996, UC became engulfed in a major political crisis that led it to reevaluate the use of standardized tests. Californians voted to end affirmative action in UC admissions. Until then, UC had considered race as one factor in admissions as a way of expanding enrollment for minority groups with historically low rates of attendance. After affirmative action was eliminated, enrollment of Latino and African American students dropped sharply at Berkeley, UCLA, and other UC campuses— at a time when these groups were rapidly on their way to becoming a majority of California's K-12 school population. As a public institution whose mission is to serve all Californians, it came under intense pressure to reexamine its admissions criteria. What was found challenged many established beliefs about admissions tests and ultimately set in motion a number of important admissions reforms.

One surprise was that the best predictor of college outcomes was not standardized tests but rather high-school grades. College outcomes can be measured in many ways. The measure used here is first-year grades at UC, but our findings were very similar for other outcomes, such as graduation. Whatever the outcome measure used, high-school grades accounted for most of the explained variance.

| <u>Predictors</u> | <u>Explained variance</u> |
|-------------------|---------------------------|
| HSGPA             | 15.4%                     |
| SAT               | 13.3%                     |
| HSGPA + SAT       | 20.8%                     |
| SAT increment     | + 5.4%                    |

Figure 4. Prediction of UC freshman grades. Source: Geiser and Studley, 2002.

Many are surprised that high-school grades are a better predictor than test scores, since grading standards differ from one high school to another. But test scores are based on a single sitting of only 3 or 4 hours. High-school grades, on the other hand, are based on repeated sampling of student performance over a period of years.

Two other findings were noteworthy. First, the overall level of prediction is relatively modest. Taken together, high-school grades and test scores accounted for only about 21% of the variance in college performance— leaving almost 80% unexplained. Our ability to predict college outcomes is relatively limited. Second, test scores add little to the prediction of college performance after high-school grades are taken into account. Adding test scores improved the prediction by an increment of only about five percentage points.

### Prediction error

Because of their limited predictive power, predictions based on test scores are subject to considerable error. The figure below compares two students who are identical in all respects —same high-school grades, same socioeconomic background, etc.— except for test scores. The data are not hypothetical but are rather based on a sample of almost 80,000 UC students. Student A, with an SAT score of 1200 points, is predicted to earn a grade-point average of 3.0, a B average during her first year at UC. Student B is otherwise identical to Student A except that she scored 1300 on the SAT, 100 points higher than Student A. Her predicted GPA at UC is 3.13.

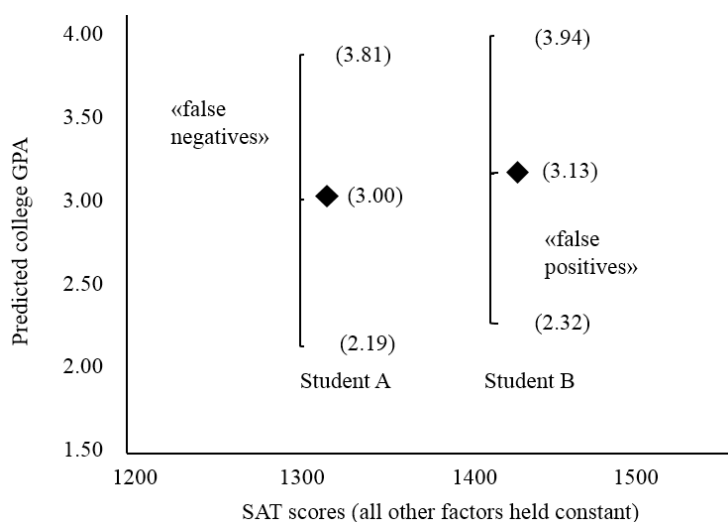


Figure 5. Prediction errors. Source: Calculated from regression analysis of 77,893 freshmen entering UC from 1996 to 1999.

Most admissions officers would automatically select Student B over Student A. Test scores are commonly used as a tiebreaker to choose between students who are equally qualified, and in this case the tie would go to Student B.

What this method ignores, however, is the error bands around the predictions. Because test scores account for such a small fraction of the total variance in college outcomes, the error bands around the predictions are quite large. In this sample of 80,000 students, the error bands were plus or minus 0.81 grade points at the 95% confidence level. This means that, for both students, the error bands are quite broad, ranging from an A average, or outstanding performance, to a C average, or relatively poor performance in college.

And here is the problem: Because the confidence intervals are so broad, predictions based on test scores are subject to errors of two kinds. First, are false positives. In a large proportion of cases, the admitted student will in fact perform worse than those denied admission would have performed. Second, are false negatives: admissions officers may deny admission to students who would have succeeded at the university and performed better than those admitted. Both kinds of error are inevitable when the predictive power of tests is low and score differences are small —as is often the case in admissions at selective colleges and universities, where almost all applicants have high scores.

### Impact of socioeconomic status on test scores

Another surprise in our research at UC was the extent to which test scores act as a barrier to admitting students from underprivileged circumstances. When the SAT was first introduced, one of its attractions was its claim to identify talented students from poorer backgrounds who might not otherwise be admitted at top universities. Many defenders of the SAT continue to believe this claim.

Yet, our data showed that SAT scores were correlated much more closely with students’ socioeconomic background than were high-school grades or other admissions criteria. Among UC applicants, the correlation between SAT scores and family income was about 0.3, and correlation with parents’ education was about 0.4. In contrast, the correlation of high-school grades with both measures was less than 0.1.

|                 | Family | Parent’s education |
|-----------------|--------|--------------------|
| SAT scores      | .32    | .40                |
| High-school GPA | .04    | .06                |

Figure 6. Correlation of SAT scores and HSGPA with socioeconomic status. Source: Geiser and Santelices (2007).

As a result, when used to rank students, test scores have a much more negative effect than other criteria on the admission of minority students, who disproportionately come from underprivileged backgrounds. This is demonstrated in Figure 7.

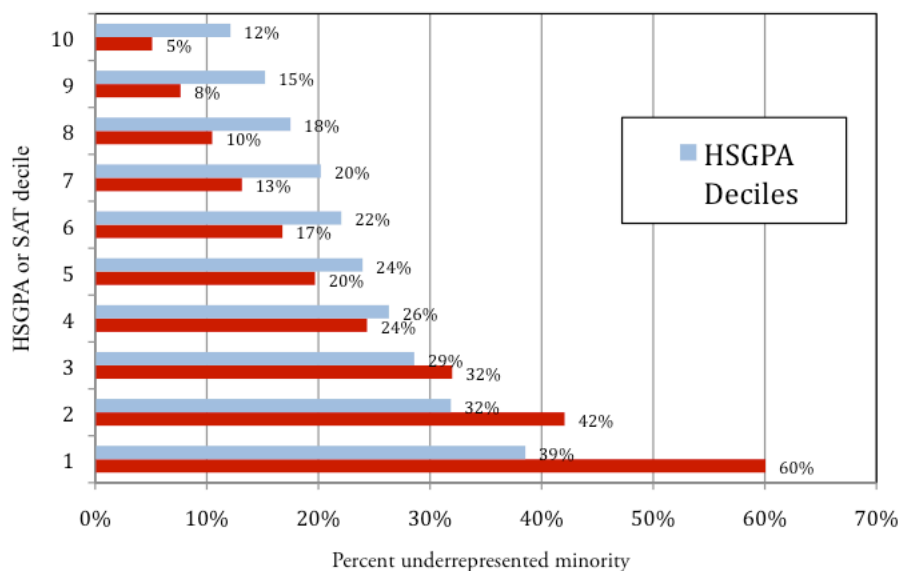


Figure 7. Distribution of underrepresented minority applicants by SAT vs. HSGPA deciles, 1994 to 2011. Source: UC Corporate Student System data on CA resident freshman applicants, 1994 to 2011.

The lighter bars show the percentage of minority students —Latinos, African Americans, and American Indians— within each high-school grade-point average (HSGPA) decile, from the top 10% to the bottom 10% of all UC applicants ranked by high-school grades.

The darker bars show their percentage within each SAT decile. Bear in mind that these are the same students, only ranked on different criteria. Though minority students tend to cluster toward the bottom of both distributions, the racial stratification produced by SAT scores is far more extreme.

Furthermore, recent findings based on the population of California residents who applied for admission to the University of California from 1994 through 2011 show that socioeconomic background factors —family income, parental education, and race/ethnicity— account for a large and growing share of the variance in students’ SAT scores over the past twenty years. More than a third of the variance in SAT scores can now be predicted by factors known at students’ birth, up from a quarter of the variance in 1994. Of those factors, moreover, race has become the strongest predictor. Rather than declining in salience, race and ethnicity are now more important than either family income or parental education in accounting for test score differences.



### «Signaling effects»

Another feature of admissions tests that is equally important but harder to measure is their *signaling effect*—the message that the tests send to students and teachers in the schools. Students often view the SAT, for example, as a test of intelligence, rather than as a test of what they have learned in school. This is increasingly true of the ACT as well, given its speededness and emphasis on quick recall of material. Low scores on either test are often interpreted as meaning that a student is lacking in intellectual ability or potential.

Criterion-referenced tests, on the other hand, send a very different message. A low score on a criterion-referenced test signifies simply that the student has not learned the content being tested. This may be due to any number of factors, such as poor teaching, lack of instructional resources, or even lack of hard work on the part of the student. Criterion-referenced tests call attention to determinants of performance that are alterable, and so are better suited to stimulate educational improvement and reform.

### Growth of criterion-referenced tests

Criterion-referenced testing has grown enormously in the United States over the past two decades as part of the effort to establish curriculum standards for K-12 schools in each state. Criterion-referenced tests are also known as standards-based assessments, and the terms are used interchangeably. Standards-based testing is associated with state-level efforts to establish curriculum standards for what students are expected to learn in each subject area.

The key idea underlying standards-based assessment is alignment. The goal is to align teaching, learning, and assessment by establishing clear standards for what students are expected to learn, teaching to the standards, and then testing students against those standards. Students' scores reflect their level of mastery of a given subject, rather than how they compare with others. The standards movement has swept secondary education in the United States over the last 20 years, but has yet to take hold at the university level, where norm-referenced tests still prevail. Schools in the United States are well ahead of colleges and universities in this regard.

Below is an example of the State of California's curriculum standards, in this case, for biology. The standards define what high-school students are expected to know about genetic coding and the sequence of amino acids encoded in DNA and RNA. As this example makes evident, the standards require in-depth knowledge of the subject matter.

Genes are set of instructions encoded in the DNA sequence of each organism that specify the sequence of amino acids in proteins characteristic of that organism.

As a basis for understanding this concept:

- a. *Students know* the general pathway by which ribosomes synthesize proteins, using tRNAs to translate genetic information in mRNA.
- b. *Students know* how to apply the genetic coding rules to predict the sequence of amino acids from a sequence of codons in RNA.

Figure 8. Sample California content standards for biology/life sciences. Source: California Department of Education.

At the university level, the two most prominent examples of standards-based tests in the United States are the SAT Subject Tests and the Advanced Placement (AP) Exams. Unlike the SAT or ACT, these tests are subject-specific and measure student mastery of biology, chemistry, foreign languages, history, physics, and so forth.

The SAT Subject Tests were introduced in the 1930s and are now offered in about 20 subject areas. They are hour-long, multiple-choice tests. SAT Subject Tests are currently required for admission mainly at top private universities like Harvard.



The AP Exams are used more widely. These tests were introduced in 1955. Their original purpose was to allow students to earn college credit while still in high school. The AP exams take 2 to 3 hours and include a combination of multiple-choice, free-answer, and essay questions. They are now offered in over 30 subject areas. Though originally intended for purposes of awarding college credit, the AP exams have come to play an increasingly important role in college admissions, especially at top universities.

Below is a sample item from the AP biology exam. Compared to the SAT or ACT, the AP exams are much more rigorous, and require more intensive knowledge of the subject matter.

If a segment of DNA is 5'-TAC GAT TAG-3', the RNA that results from the transcription of this segment will be:

(A) 3'-TAC GAT TAU-5'  
 (B) 3'-ATG CTA ATA-5'  
 (C) 3'-UAC GAU UAG-5'  
 (D) 3'-AAC GAU UAA-5'  
 (E) 3'-AUG CUA AUC-5'

Answer: (E)

Figure 9. Sample AP biology item. Source: College Board.

And here is a typical score distribution for an AP exam, continuing with the example of biology. Clearly this is not the bell-curve distribution associated with the SAT or ACT.

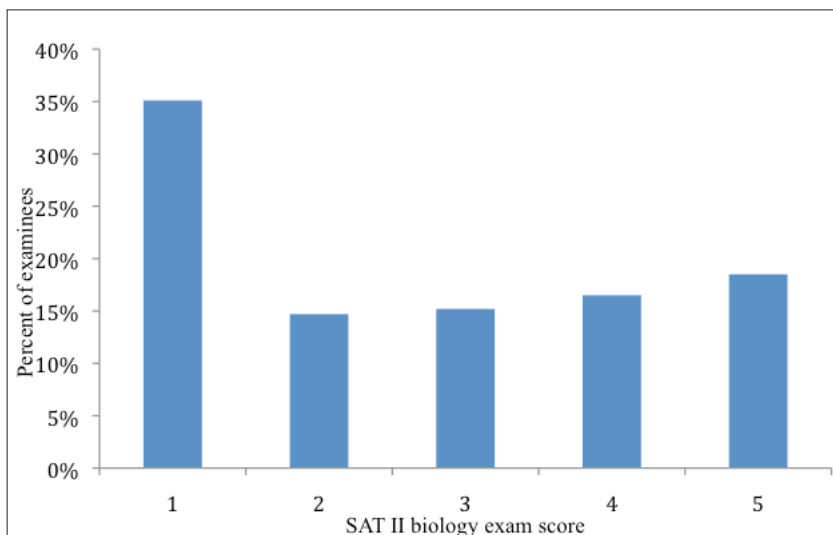


Figure 10. Distribution of AP biology exam scores. Source: College Board, 2010.

At most universities, an AP score of 4 or 5 is usually required for a student to be deemed qualified or proficient in a given subject, although a score of 5 is sometimes required at top institutions. This is usually referred to as a *cut score*. But a cut score on a criterion-referenced test has a different meaning than a cut score on a norm-referenced test. On norm-referenced tests, the cut score represents a student's percentile rank compared to other test takers. On criterion-referenced tests, the cut score represents the student's level of proficiency in a given subject.

### Predictive validity of criterion- vs. norm-referenced tests

The University of California is one of the few institutions in the United States that has used both AP Exams and Subject Tests, as well as the SAT and ACT, and UC thus has an extensive database with which to compare the two types of tests. One of the biggest surprises from our research was that criterion-referenced tests predict college outcomes at least as well as norm-referenced tests. Although the AP exams and Subject Tests are not designed for prediction, they were among the better predictors of how students will perform in college.

| Predictor variables     | Standardized weight |
|-------------------------|---------------------|
| HSGPA                   | .29                 |
| AP exams                | .14                 |
| SAT subject tests       | .09                 |
| Parents' education      | .09                 |
| SAT (or ACT equivalent) | .06                 |
| Family income           | .03                 |
| Explained variance      | 25.1%               |

Figure 11. Relative weight of different variables in predicting UC freshman grades. Source: Geiser and Santelices, 2006.

The strongest predictor of college performance was high-school grades. The predictive weight for high-school GPA in our sample was .29. What this means is that, for every one standard-deviation increase in high-school GPA, college GPA increased by .29 of a standard deviation, when all other factors were held constant.

AP exams were the second best predictor, followed by the Subject Tests. Parents' education was next. Finally, SAT and ACT scores ranked behind parents' education and only slightly ahead of family income in predicting college outcomes.

The main lesson to draw from these findings is not that criterion-referenced tests are preferable to norm-referenced tests because they are better predictors. In fact, the ability to predict student success in college is limited. Taken together, all of these predictors accounted for only about a quarter of the variance in UC freshman grades, leaving three-quarters unexplained. The statistical differences among the various tests are too small to dictate adoption of one test over another.

Rather, the lesson to draw is this: Because the statistical differences are so small, the choice of tests to be used in college admissions must be driven by considerations other than prediction. It is here that the advantages of criterion-referenced tests are most apparent.

### Essential features of college-admissions tests

Based on the UC research, here are the most important features that should be required of admissions tests:

- The primary focus should not be how an applicant compares to others, but whether he or she has mastered the foundational knowledge and skills needed for college.
- Admissions tests should exhibit face validity as well as predictive validity, so that the link between the knowledge and skills being tested and those needed for college is transparent.

- Admissions tests should align with high-school coursework and reinforce teaching and learning of a rigorous curriculum in our schools.
- Admissions tests should minimize the need for test preparation by rewarding content knowledge over test-taking skills, so that the best test preparation is regular classroom instruction.
- Finally, admissions tests should send a clear signal to students that working hard and mastering academic subjects in high school is the most direct route to college.

Standards-based or criterion-referenced tests have significant educational advantages over norm-referenced tests in all of these respects.

To summarize the research conclusions at the University of California, the findings show that admissions criteria that measure mastery of curriculum content —such as high-school grades and standards-based tests— predict student success in college at least as well as norm-referenced tests such as the SAT or ACT. Such measures —especially HSGPA— are also fairer and more inclusive of low-income and minority applicants.

Equally important, these criteria have a broader and more beneficial signaling effect throughout all of education. They send a strong message to students that working hard and mastering academic subjects in high school —not innate ability— is key for admission to college. In this way, they help to align teaching, learning, and assessment all along the pathway from high school to college.

#### **Postscript 2014: Lessons from policy implementation**

The previous section presented the main conclusions of my research as of 2001. The findings helped spark one of the most remarkable and sustained periods of policy reform in the University of California's history. The reforms included, first, UC's Top 4% Plan, which guaranteed admission to top students from each California high school. Students' rank in school was determined solely by their grades, thus reducing the importance of test scores as an admissions criterion. Second, UC introduced holistic review, an admissions process that takes into account a breadth of factors such as socioeconomic disadvantage and opportunity to learn. This reform, too, was designed to reduce the role of tests. Third, UC initiated a massive program of university outreach to low-performing schools throughout California, aimed at improving college preparation among minority and low-income students. And finally, under the leadership of UC president Richard Atkinson, the university mounted a major challenge to the then-dominant national admissions test, the SAT itself.

Some of those reform efforts proved successful, others much less so. In the final portion of this article, I shall briefly reflect back from the vantage point of 2014 on each of the reforms and the lessons we learned in the course of policy implementation.

#### **The Top 4% Plan**

UC introduced the Top 4% Plan in 2001. The policy proved quite successful. The policy was intended to increase admission of students from schools that historically had sent few students to the university. And as shown here, it had precisely that effect.

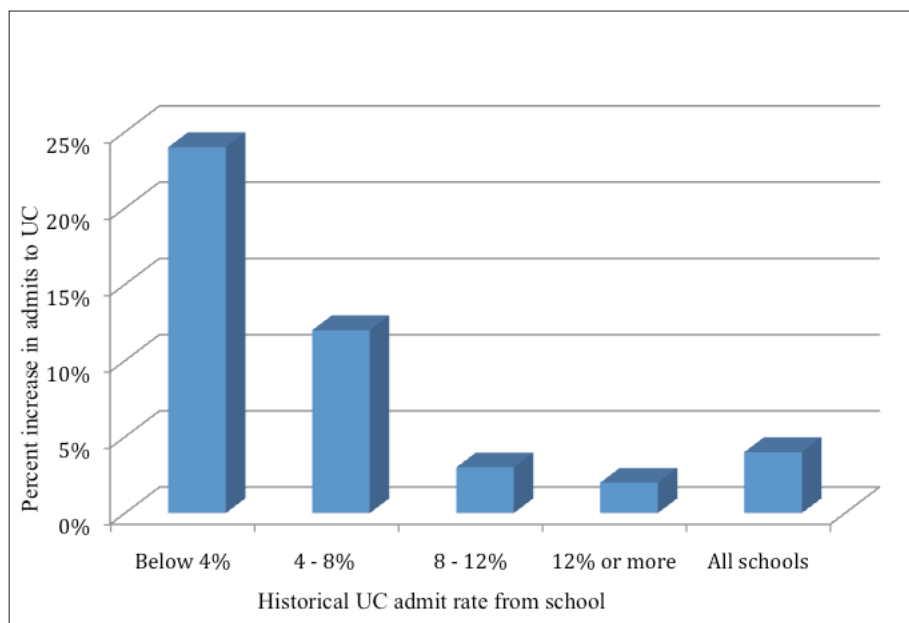


Figure 12. Impact of Top 4% Plan on schools with historically low UC admit rates. Source: University of California.

Another strength of the Top 4% Plan what may be called its policy narrative. The narrative or story used to explain a policy initiative to the public can be an important factor in its ultimate success or failure. In the case of the Top 4% Plan, the idea of admitting top students from each California high school was a powerful combination of both egalitarian and meritocratic values. It was egalitarian insofar as it extended eligibility for admission to students in every high school in the state. But it was also meritocratic insofar as it limited eligibility only to the very top students in each school. The policy proved extremely popular for this reason and illustrates the importance of a strong policy narrative for effective policy implementation.

The main weakness of the policy was its limited scale. The new policy generated only about 2,500 new admits because most students in the top 4% of their high schools were already eligible under UC's regular admissions criteria. To address this problem, UC expanded the policy from the top 4% to the top 9% of each high school in 2012.

### Holistic review

Holistic review was a second major reform. Its purpose is to evaluate applicants on a broader set of factors than tests and grades alone. The key idea is to assess achievement in context, that is, the extent to which students have taken advantage of their opportunity to learn. For example, an SAT score of 1400 for a student from a poor high school represents more of an achievement than the same score for a student from a top school. Until 2001, holistic review had been used only at highly selective private institutions. Its adoption at UC was one of the first cases where it was employed at a major public university with a large volume of applicants.

Holistic review has been much more controversial than the Top 4% Plan. Under this approach, reviewers are trained to consider all of the material in an applicant's file and then assign an overall score. It is labor-intensive and therefore expensive. It requires at least two, and sometimes three or more, reviews of each application by admissions readers in cases where the first two reviews differ. Admissions staff must be continuously trained and retrained to ensure consistent results across different reviewers—a process known as *norming*. In addition to the cost and effort involved, critics of holistic review dislike the lack of transparency it introduces into the admissions process, and they are suspicious of possible bias for or against different racial groups.

Another criticism concerns the assessment methodology. Because the reviewer combines the various admissions factors in their mind to produce an overall score, the weight assigned to individual factors may vary from one case to the next. But the holistic approach is not the only method for comprehensive review of applicant files. An alternative is the *fixed-weight* or *point system* approach shown below, from UC San Diego.

| <u>Academic factors:</u> | <u>points</u> | <u>Additional factors:</u> | <u>points</u> |
|--------------------------|---------------|----------------------------|---------------|
| High school GPA          | 4,500         | Disadvantaged school       | 300           |
| Test scores              | 3,200         | Low family income          | 300           |
| Extra classes            | 500           | First-generation college   | 300           |
| Top 4% /ELC              | 300           | Demonstrated leadership    | 300           |
|                          |               | Special talents            | 300           |
|                          |               | Community service          | 300           |
|                          |               | UC outreach participant    | 300           |
|                          |               | Special circumstances      | 300           |

Figure 13. An alternative to holistic review: «Fixed-weight» systems. Source: UC San Diego.

Point systems take into account all of the same factors as holistic review. But each factor is rated separately and has a fixed weight, and the ratings are summed to produce an overall score. There is a considerable body of research showing that a fixed-weight approach is more reliable and transparent, as well as less costly, than holistic review (Grove, 2005; Meehl, 1954).

In the final analysis, the key question is whether comprehensive review of either type —holistic or fixed weight— produces different, and better, admissions decisions. An early pilot study at Berkeley found that comprehensive review changed the outcome in only about 5 to 6% of all admissions decisions. It has had a measurable impact on the diversity of the student body, but that effect is relatively limited and likely to remain so as long as grades and test scores continue to be the primary criteria for university admissions.

### The rise and fall of UC outreach

UC's third major initiative was a massive expansion university of outreach programs for disadvantaged students and schools in California beginning in 2001. By 2003, UC outreach programs were serving over 300,000 students and 70,000 teachers and principals. The university had entered into partnerships with 300 of the lowest-performing schools in the state, located primarily in poorer, urban areas in the vicinity of each UC campus.

The expansion of university outreach grew out of a recognition that admissions reforms, by themselves, could have only a marginal effect in expanding access for minority and low-income students. Admission of these groups would continue to be limited as long as these students remained concentrated in California's lowest-performing schools. A major effort was needed to improve teaching and learning in those schools.

UC had long provided one-on-one tutoring and academic enrichment programs for individual students at the K-12 level. What was new was the effort to intervene in the schools themselves. The state legislature provided substantial new funding that enabled each UC campus to form partnerships with low-performing schools in their region. UC began offering professional development programs for teachers and principals in those schools. UC also offered subject-matter programs aimed at upgrading the curricula. It was understood that this was necessarily a long-term project. Some of the UC partners were middle and primary schools, so it would be some time before results would be forthcoming.

But the legislature's time frame was much shorter. They wanted immediate results in boosting minority admissions at UC. When this failed to occur, funding was withdrawn, and school-centered outreach came to almost a complete halt by 2005. Looking back a decade later, UC outreach is best viewed not as a policy failure, but rather as an unfinished experiment. The key question remains unanswered. Can universities successfully intervene to improve teaching and learning in the poorest schools?

### **Challenge to national admissions tests**

UC's last major policy initiative was its challenge to the nation's leading college-admissions test at that time, the SAT. (Since then, the SAT has been surpassed by the ACT as the most widely taken admissions test in the US.). With respect to that initiative, the final verdict is mixed. The efforts to change the tests proved partially successful in several ways. In 2005, both the SAT and ACT added a writing test in response to my and other researchers' findings that, of the various subject tests, writing was among the better predictors of college performance. The SAT dropped verbal analogies and quantitative comparisons, which we had criticized as a holdover from that test's IQ tradition. And in general, the College Board has taken steps to move the content of the SAT closer to the high school curriculum. For example, quantitative comparisons —another tricky item-type not normally taught in school— were also dropped in 2005 and replaced with more straightforward items covering second- and third-year math. These changes are all steps in the right direction. And both the SAT and ACT have recently announced changes that will move those tests still further in the same direction.

The problem, however, is that both tests remain norm-referenced assessments. They are fundamentally out of sync, in that respect, with the standards-based movement in the rest of U.S. education over the past two decades. It is probably inevitable that U.S. admissions tests will eventually evolve into standards-based assessments, although it will probably take some time. Until then, the SAT and ACT will continue to fuel the obsession with high-stakes testing —and indeed the whole test-prep industry— in U.S. college admissions. The importance of small differences in test scores that bear little relationship to college outcomes will continue to be exaggerated.

But there is a clear alternative. In 2008, the National Association for College Admissions Counseling convened a blue-ribbon commission on admissions testing in the United States. Chaired by William Fitzsimmons, dean of admissions at Harvard, the commission came to very much the same conclusion presented in this article:

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 benefitting millions of American students who do not enroll in selective colleges and positively affecting teaching and learning in America's schools (National Association for College Admission Counseling, 2008).

To be sure, curriculum-referenced or standards-based tests are no panacea that can make college admissions completely fair and rational. The inequalities in our secondary schools are too deep and pervasive for that. But standards-based tests can help make the process fairer and more rational than it is now. First, at the university level, they can help level the playing field for those who cannot pay for expensive test-prep services —even in the poorest schools, students who work hard and perform well in their regular coursework could improve their chances of admission.

And second, at the school level, criterion-referenced tests can reinforce teaching and learning of a rigorous college-preparatory curriculum. The reinforcement they provide may be even more important in low-performing schools than others. Experience in implementing state curriculum standards in the United States suggests that a strategy of setting clear content standards, teaching to the standards, and assessing students against those standards may produce the greatest benefits within the most disadvantaged schools. By adopting standards-based assessments, our colleges and universities can have a broader and more beneficial effect throughout all of education.

### Discussion

The experience from UC may be important for other selective institutions, both in the United States and abroad, that rely heavily on standardized test scores in their admissions process. Criterion-referenced or standards-based tests have clear advantages over norm-referenced tests for purposes of college admissions, as well as school improvement, more broadly. In addition, the findings about prediction error, the relationship between test scores and socioeconomic status, and the signaling effects of the admission indicators suggest that standardized test scores should be used only as one, among several other, admissions criteria. Test scores should be complemented by other sources of information about students' profiles, much in line with the Holistic Review implemented by UC. College Outreach remains as a potential alternative to increase the diversity of the institution's student body, but it needs to be understood as a long-term strategy that requires persistent institutional funding and commitment. The UC experience in implementing the Top 4% Plan suggests that admissions based on high-school class rank would have considerable scope to diversify college admissions without significant declines in college graduation. New and different admissions policy options must be considered if the goal is to increase the representation of minority groups in higher education, especially at highly selective institutions.

The original article was received on March 15<sup>th</sup>, 2015

The revised article was received on January 21<sup>st</sup>, 2016

The article was accepted on March 14<sup>th</sup>, 2016



### References

- Geiser, S. (2009). Back to the basics: In defense of achievement (and achievement tests) in college admissions. *Change: The Magazine of Higher Learning*, 41(1), 16-23.
- Geiser, S. (2014). *Back to the future: Freshman admissions at the University of California, 1994 to the present and beyond*. Research and Occasional Paper Series: CSHE.4.14. Berkeley, CA: Center for Studies in Higher Education, University of California, Berkeley. Retrieved from <http://www.cshe.berkeley.edu/publications/back-future-freshman-admissions-university-california-1994-present-and-beyond>
- Geiser, S. (2015). *The growing correlation between race and Sat Scores: New findings from California*. Research and Occasional Paper Series No. 10.15. Berkeley, CA: Center for Studies in Higher Education, University of California, Berkeley. Retrieved from <http://www.cshe.berkeley.edu/publications/growing-correlation-between-race-and-sat-scores-new-findings-california-saul-geiser>
- Geiser, S., & Atkinson, R. (2013). Beyond the master plan: The case for restructuring baccalaureate education in California. *California Journal of Politics and Policy*, 4(1), 67-123.
- Geiser, S., & Caspary, K. (2005). 'No show' study: College destinations of University of California applicants and admits who did not enroll, 1997-2002. *Journal of Educational Policy*, 19(2), 396-417.
- Geiser, S., & Santelices, M. V. (2006). The role of Advanced Placement and honors courses in college admissions. In P. Gandara, G. Orfield, & C. Horn (Eds.) *Expanding opportunity in higher education: Leveraging promise* (pp. 75-114). Albany, NY: State University of New York Press.
- Geiser, S., & Santelices, M. V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes*. Research and Occasional Paper Series #CSHE 9.07. Berkeley, CA: Center for Studies in Higher Education, University of California, Berkeley.
- Geiser, S., & Studley, R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, 8(1), 1-26.
- Grove, W. M. (2005). Clinical versus statistical prediction: The contribution of Paul E. Meehl. *Journal of Clinical Psychology*, 61(10), 1233-1243.
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota.
- National Association for College Admissions Counseling (2008). *Report of the commission on the use of standardized tests in undergraduate admissions*. Arlington, VA: Author.



