

2013

Is the SAT the Root of all Evil? Reviewing the Evidence on Admission Policies and Diversity in Higher Education

Loris Fagioli

Claremont Graduate University, loris.fagioli@cgu.edu

Follow this and additional works at: <http://scholarship.claremont.edu/lux>

 Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Educational Methods Commons](#), [Elementary and Middle and Secondary Education Administration Commons](#), and the [Higher Education and Teaching Commons](#)

Recommended Citation

Fagioli, Loris (2013) "Is the SAT the Root of all Evil? Reviewing the Evidence on Admission Policies and Diversity in Higher Education," *LUX: A Journal of Transdisciplinary Writing and Research from Claremont Graduate University*: Vol. 2: Iss. 1, Article 15. Available at: <http://scholarship.claremont.edu/lux/vol2/iss1/15>

Is the SAT the Root of all Evil? Reviewing the Evidence on Admission Policies and Diversity in Higher Education

Loris P. Fagioli
Claremont Graduate University
School of Education

Abstract

It is imperative to achieve diversity in Higher Education. With affirmative action policies under fire, it is becoming difficult to enroll a diverse student body. Many critics see standardized tests, and the SAT in particular, as contributing to the problem. This paper reviews research on such criticism, about suggested alternative approaches, and regarding recommendations on how to improve the current situation. In general, this review finds little evidence against a judicious use of the SAT. Also, alternative approaches such as percent plans or abolishing the SAT have had little success in increasing diversity. However, most specialists agree that a comprehensive approach to college admissions is needed.

Diversity in higher education is a vital goal if higher education wants to fulfill its mission and have success in an increasingly diverse environment (Smith, 2009). However, policies currently in place (such as affirmative action) have come under fire (e.g., Proposition 209 in California, Initiative 200 in Washington, Hopwood case in Texas). Without such policies, providing equal access for higher education to all is very difficult. The former president of the University of California described the situation of diversity at the U.C. as being in great trouble after Proposition 209 was passed (Atkinson & Pelfrey, 2006). Thus, colleges and universities have had to come up with new ways to achieve a diverse student body.

Since the rise of meritocracy during the 1980s and 1990s, institutions have relied more and more heavily on test scores (Alon & Tienda, 2007). However, there are prevalent differences between populations on several academic measures (Camara & Schmidt, 1999). African Americans consistently score about 1 Standard Deviation (SD) below whites on the SAT, ACT, Graduate Record Examinations (GRE), employment tests, military tests, and general ability tests (Roth, Bevier, Bobko, Switzer, & Tyler, 2001). Similarly, Hispanics score about 2/3 SD below whites on the same tests. Differences with regard to females compared to males also exist, albeit to a much smaller degree (.1-.2 SD). This achievement gap has persisted over decades and efforts to reduce this gap have not been successful (Hedges & Nowell, 1998; Krueger, Rothstein, & Turner, 2006). The increasing importance of test scores and the prevailing mean differences between racial groups make the issue of enrolling a diverse student body challenging. It is thus not surprising that criticism of high-stakes testing is common and tests are often seen as either the root of the problem or at the least contributing to the achievement gap and the limited diversity in higher education (Crouse, 1988; Gould, 1996; Phillips, 2006; Roth *et al.*, 2001). This question of “are tests inherently evil?” (Sireci, 2007a) is hotly debated inside and outside the

educational research community. Unfortunately, the discussion is rarely systematic and has become more and more polemic. There are two extreme points and no middle ground, one side seeing tests as severely biased and the main problem of differences in achievement, and the other side seeing no fault with the current tests (Schellenberg, 2004). Cole (1981) noted that it is often very clear from the start to which camp the research author belongs. He called the one camp the “defenders”, who are convinced of no bias in tests and who defend the status quo of today’s testing practices. The “reformers”, on the other hand, believe that only through a radical reform of testing practices can the faults be remedied.

Differing opinions prevail and it is thus vital to approach this topic in a systematic way. Due to space constraints however, this literature review will not evaluate every facet of testing and its criticism but only focus on research pertaining to the SAT and its specific implications for diversity in higher education. This paper consists of four parts. The first part addresses the most commonly voiced arguments against the SAT and reviews the empirical literature on each argument. The second part discusses alternative approaches to admissions. The third part reviews suggestions from measurement specialists regarding how to improve testing and admission policies from a psychometric perspective. The paper ends in a summary with conclusions and recommendations for future research.

Common Criticism of the SAT and its Use in Admissions

It is ironic that in the beginning standardized tests were introduced in order to reduce differences and background influences. The French psychologist Alfred Binet utilized standardized testing to find children who needed help in the educational system (Gould, 1996) and James B. Conant – Harvard president in the 30s and 40s – introduced tests to limit wealth and make admissions more egalitarian for people with lesser means (Zwick, 2002a). Today however, standardized tests are often seen as biased against minorities (Crouse, 1988; Gould, 1996; Jenks, 1998; Jenks & Phillips, 1998; Phelps, 2009; Phillips, 2006; Rothstein, 2005) and as only a measure of wealth (Phillips, Brooks-Gunn, Duncan, Klebanov, & Crane, 1998; Wilson, 1998).¹ There are of course many more areas of criticism that can be identified, specifically, against the use of the SAT in admissions (Sackett, Borneman, & Connelly, 2008; Wightman, 2003; Zwick, 2002a). In this review however, I will focus only on the following three arguments: the SAT (1) is a bad predictor and only explains a low variance of college success; (2) does not predict beyond first year GPA and does not measure other important criteria; and (3) is biased against minorities.

The SAT as a bad predictor

One of the most cited arguments against the SAT is its low predictive value. Commonly, SAT scores have been reported to correlate with college GPA between .25 to .35 (Bridgeman, McCamley-Jenkins, & Ervin, 2000; Sackett, Schmitt, Ellingson, & Kabin, 2001). When squaring these values the results suggest an explained variance of less than 10%. However, there are several problems with the interpretation and the calculation of these values.

As a first note, the interpretation of explained variance is already difficult to grasp. For this reason, alternative measures have been developed to make interpretation more intuitive. For instance, Lawshe, Bolda, Brune, and Auclair (1958) developed expectancy tables with odd ratios. They showed that with a correlation of .30 the top 20% of a population are twice as likely

to succeed compared to the bottom 20%. These calculations show that even low correlations still hold some value.

Nevertheless, the bigger problem is the strong underestimation of the reported numbers. There is compelling evidence that these correlations do not represent the true relationship between test scores and college success. Two statistical phenomena contribute to this underestimation. The professional literature calls the first issue underestimation of validity due to range restriction (Mendoza & Mumford, 1987). For instance, if a population with a correlation of .5 is split in half and correlations for each half are recalculated, the estimates drop to .33 (Zwick, 2002a). The lower variance of the restricted pool leads to an underestimation of the original correlation. The good news is that the estimates can be corrected (Hunter, Schmidt, & Huy, 2006; Sackett & Yang, 2000). In fact, the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) recommends the use of these corrective techniques.

The second problem that leads to further underestimation is the criterion (i.e. college GPA) itself. It is well known that grades are unreliable. Apart from variances in grading standards between instructors, grades vary considerably between field and rigor of the course. Failure to correct for this known measurement error or unreliability of the criterion will result in an underestimation of the values (Sackett, et al., 2001). Again, there are statistical methods to correct for this phenomenon (Stricker, Rock, & Burton, 1993; Young, 1991).

These two problems are well known in psychometrics and statistics. Unfortunately, many studies are still published without the use of these corrective techniques. Research critical of the SAT rarely includes any corrections (Geiser & Santelices, 2007; Geiser & Studley, 2002) which make accurate comparisons among results difficult. The following will highlight some studies that have used these techniques and show the differences between corrected and uncorrected estimates.

Bridgeman *et al.* (2000) analyzed data from 23 colleges and reported uncorrected correlations between SAT and Freshman Grade Point Average (FGPA) around .30. But when correcting for range restriction and course difficulty the estimates ranged from .47 to .58. Similarly, Burton and Ramist (2001) reviewed several studies that used corrected estimates. They showed that the correlations improved by about .30. When taking both SAT and High School Grade Point Average (HGPA) into account, correlations ranged from .64 to .76. The authors concluded that these estimates “can no longer be characterized as ‘small’ or even ‘moderate’. The corrected correlation of .76 (...) is *large*” (Burton & Ramist, 2001, p. 27).

It is unfortunate that due to complications in statistics and measurement many studies report values that underpredict the true nature of the relationship between tests and college success. It seems that while not a perfect measure of future success, SAT scores - especially in combination with HGPA - have a strong foundation of predictive power.

The SAT does not predict college success beyond Freshman GPA

At first glance it might be perplexing that most studies use the GPA score at the end of Freshman year as a measure for college success. Why are studies not using better measures of college success such as senior year GPA, graduation, or income after college? This question is indeed valid, but there seem to be very practical reasons for the use of FGPA. After the first year of college, course taking patterns become more varied and students take more specialized

courses. As a result, the reliability of senior year GPA is reduced (Bridgeman *et al.*, 2000). First year GPA, on the other hand, has a set of generally comparable courses, and more comparable grading (Zwick & Sklar, 2005). Additionally, FGPA is quickly available (Burton & Ramist, 2001) and suffers from fewer issues with missing data. Due to student drop-outs and transfers the issue of range restriction is heightened when estimating correlation coefficients with senior year GPA. While these arguments might be reasonable, they do not change the fact that FGPA is not an ideal measure of college success.

The issue of range restriction and unreliability of criterion measured is even more important when using measures beyond FGPA. In a review of studies that looked at measures other than first year GPA, Burton and Ramist (2001) noted that it was difficult to compare and review these articles. Only few use corrections in their estimates and post-hoc adjustments are difficult. Nevertheless, results showed that the combination of HGPA and SAT successfully predicted grades, honors, acceptance, and graduation rates in colleges, graduate schools, and professional schools (coefficients ranged from .29 to .33). In another review, Sackett *et al.* (2008) reported on several meta-analytic studies that found similar predictive results of GPA for all four years. In a longitudinal study of gifted children who took the SAT, results showed a positive relationship to getting a PhD, getting tenure, and having high job satisfaction, (Lubinski, Benbow, Webb, & Bleske-Rechek, 2006). There is also some indication for a positive relationship between the SAT and post college income and earnings (Bowen & Bok, 1998; Jenks & Phillips, 1998).

There is evidence that the SAT is predicting beyond just first year college GPA. However, most research shows a diminishing magnitude of association, especially for measures beyond GPA. The question is whether a test of developed abilities should be able to predict, for instance, college graduation or income (Sackett *et al.*, 2001). Many other factors play a role in student success in college. Financial considerations, family and life events, social environment, personal relationships, and many other factors play a role in college success (Geiser & Santelices, 2007). For instance, Fleming (1998) showed that institutional factors play a major role in students who drop out of Science Technology Engineering and Mathematics (STEM). The author showed that validity estimates can underestimate performance if adjustment to college is poor. Thus, even though several relevant criteria are not predicted well by the SAT, they do not constitute a case against its use, but a strong imperative for alternative measures that do touch on these important elements of college success. The SAT is only a measure of developed academic abilities; it is important to point out its limitations and to make sure admission decisions are not solely based on academic criteria or only one measure (see the third section of this paper for more information on this topic).

The SAT is biased against minorities

Test bias is maybe the most fiercely debated topic when it comes to standardized testing. As mentioned in the introduction, the achievement gap has a long history with standardized tests as well as other measures of achievement revealing underachievement of minorities and underprivileged groups (Cataldi, Laird, & Kewal-Ramani, 2009). Additionally, there is also no denial that many tests were severely biased against minorities in the past (Gould, 1996) and that there were people who supported tests who held racist, anti-feminist, or anti-Semitic views (Zwick, 2002a). Perhaps contrary to popular belief, the issue of test bias and especially bias against minorities is one of the most prominent research themes in the testing community (Cole,

1981; Hunter & Schmidt, 2000; Thorndike, 2005). Tests and test construction are not stagnant and have developed tremendously over the last few decades. Unfortunately however, a large gap exists between the scholarly writing in psychometrics and the opinions of people not familiar with testing practices (Cole, 1981). One reason given by Cole for this discrepancy is the different foci of concern. While technical researchers are mostly interested in test bias itself, the public and many educational researchers are generally more worried about bias in its social, economic, and political context. Another reason for the discrepancy could also be that noting mean differences between groups is not proof for bias (Sackett et al., 2008). There are several reasons why test scores can differ depending on experiences, background, and psychological state. In fact, *The Standards* (AERA et al., 1999) noted that “the idea that fairness requires equality in overall passing rates for different groups has been almost entirely repudiated in the professional literature” (p.74). There is a wide body of research on the topic of test bias. It is important to know how tests are developed and what steps are taken to assure the creation and use of a fair test (Messick, 1989; Schellenberg, 2004; Sireci, 2004, 2007a, 2007b; Thorndike, 2005; Zwick, 2002a). However for this review, I will focus specifically on the SAT.

While mean differences between groups on a test should raise concern, it is not enough evidence that bias exists. *The Standards* mentioned above presents the most widely accepted idea of how to test for fairness: “examinees of equal standing with regard to the construct the test is intended to measure should on average earn the same test score regardless of group membership” (AERA et al., 1999, p.74). That is, a test should predict a criterion equally well for people with the same score. For instance, a group of students with the same SAT score should on average have a similar college GPA. If the SAT were biased, we would expect one group consistently having a higher GPA compared to others with the same SAT score. This phenomenon is called “underprediction” and is evidence of negative bias (*i.e.*, puts a group at a disadvantage). In underprediction the SAT would falsely suggest that some students had a lower GPA than they could actually achieve. On the other hand, “overprediction” would be positive bias since the test predicts a higher GPA than is actually achieved.

Extensive research has been conducted on the topic of over/underprediction of minority group performance on the SAT. Large-scale studies² such as Ramist, Lewis, and McCamley-Jenkins (1994), Bridgeman, McCamley-Jenkins, and Ervin (2000), and Mattern, Patterson, Shaw, Kobrin, and Barbuti (2008); as well as reviews such as Linn (1973), and Young and Kobrin (2001), all come to similar conclusions. Even though there is variation, and not all studies have the same groupings, the general consensus of these authors is that there is overprediction with regard to racial minorities taking the SAT. Generally, African Americans, Hispanics, and Native Americans score .1 to .3 points below their expected GPA based on their SAT score. However, the exact reasons and causes for this overprediction are not well understood yet (Young & Kobrin, 2001).³

More evidence exists for the underprediction of females. Female GPA scores are, on average, higher than expected based on predictions from their SAT scores. However, the deviation is smaller, ranging from 0 to .10. Some suggest that this difference is due to course-taking patterns (Ramist, Lewis, & McCamley, 1990) but this and other explanations do not hold up across studies (Zwick, 2002a). More research is needed to understand this phenomenon. Additionally, little is yet known about English Language Learners and their performance (Sireci, Han, & Wells, 2008; Zwick, 2002a, 2007).

The reviewed literature suggests that there is little evidence of bias against racial minorities on the SAT. On the contrary, most research shows that college GPA is overpredicted

for African Americans, Hispanics, and American Indians. On the other hand, some evidence does exist of a small bias against women. However, even though these results are well established, the underlying causes are not yet fully understood and more research is needed in this area. As Schellenberg (2004) noted, psychometrics may detect but cannot explain racial bias. Bias is not only inherent in a test, but there can be bias in its application, its use, and its interpretation (Jenks, 1998; Thorndike, 2005). Already Messick (1975, 1989) noted that there is a technical side to bias and an ethical side as well (some authors prefer the term "fairness" for the discussion of the latter). The third part of this review will thus look more closely at some suggestions of proper use of SAT scores.

Alternatives to the SAT in Admission Decisions

The previous section showed considerable support for a judicious use of the SAT. However, with increased opposition against affirmative action, it is becoming more and more difficult to enroll a diverse student body (Atkinson & Pelfrey, 2006). Several other avenues have been suggested to increase diversity either in addition or as a substitute to the SAT. This section will look at the following three suggestions: percent plans, high school GPA, and comprehensive measures.

Percent plans

Due to legal challenges of affirmative action, several states were forced to look at other criteria besides race to ensure diversity in higher education. California, Texas, and Florida introduced "percentage plans" where the top 4 percent (California), 10 percent (Texas), or 20 percent (Florida) of a graduating class is ensured enrollment at a state college or university. The problem with this policy is, of course, immediately apparent. It only works under the assumption that high schools are segregated. This could lead to an incentive for schools not to desegregate or to improve academically if their top students are guaranteed enrollment (Zwick, 2002a). Furthermore and more importantly, several studies have been published that show little to no effect of percent plans in improving diversity. The Civil Rights Project at Harvard University conducted a series of studies and concluded that percent plans in the three states have very little effect on diversity (Horn & Flores, 2003). Similarly, an evaluation of the U.S. Commission on Civil Rights noted that the Commission was "deeply concerned" about percentage plans used for diversity (Office of Civil Rights, 2002). Their study showed that percentage plans alone do not foster diversity. Data from the University of California Santa Barbara revealed that from 77 students considered under the percentage plan in a year, 37 (48%) were minority students, none of whom were admitted (Zwick, 2007). Simulations of percentage plans also showed no effect on diversity (Carnevale & Rose, 2003). Why do percentage plans have such little impact? There are, of course, several reasons and the percentage plans vary considerably across the three states (Office of Civil Rights, 2002). Most likely, however, the top students considered under the plan are already eligible without the plan (Horn & Flores, 2003).

High school GPA as a sole measure

High School GPA is often the most important criterion in college admissions, while standardized tests are reported as the second most important (Hawkins & Lantz, 2007). Several

studies report that HGPA is consistently the single best predictor for college GPA (Bridgeman et al., 2000; Burton & Ramist, 2001; Geiser & Santelices, 2007; Geiser & Studley, 2002). Additionally, supported by the recent criticism of the SAT by Richard Atkinson - the former President of the University of California - some have suggested abolishing the SAT in favor of using only HGPA. However, there are several reasons against this implementation. First, Atkinson never intended to abolish standardized testing (Atkinson, 2002). On the contrary, he is in favor of the use of tests, but suggested to move away from an aptitude orientation to a focus on achievement (a road which the new SAT has taken) (Zwick, 2004). Secondly, even though studies have shown HGPA to be a better predictor than the SAT by itself, the combination of the two significantly increases accuracy of prediction (Burton & Ramist, 2001; Geiser & Santelices, 2007). But what would happen if the SAT were abolished in favor of only HGPA? Some research suggests that it would increase admissions for white students and decrease it for blacks (Zwick, 1999). Zwick (2007) argued that abolishing the SAT might work for smaller colleges (see below for comprehensive measures) but it would be very challenging for larger institutions. Using only HGPA would simply substitute a professionally developed test with a teacher test. But more importantly, the achievement gap is not restricted to standardized tests but is evident in all measures related to success in school, including GPA (Cataldi et al., 2009). It is thus unlikely that using HGPA which has similar patterns of achievement as the SAT, would produce dramatic changes in college admissions (Zwick, 1999). This point is further supported by evidence from the University of California. Only 2.5% of students were not eligible due to low test scores, but 67% were not eligible due to a lack of course-taking background (Zwick, 2002a). This evidence suggests that test scores are not the main reason for the limited enrollment of minorities in the case of the University of California.

Comprehensive assessment

Affirmative action policies are the most effective at ensuring a diverse student body (Bowen & Bok, 1998). Due to increased opposition towards these practices, other avenues need to be explored. Research suggests that neither percentage plans, nor the abolishment of standardized tests will achieve the desired results. There is some evidence, however, that more comprehensive approaches to admissions can support diversity in higher education. For instance, the University of California adopted several measures including: outreach to lowest performing 20 percent of schools, weighting of achievement over aptitude, comprehensive review of applications, and guaranteed access if certain grades or courses were taken in a community college. Because of these measures, minority enrollment was increased from 15% to 18% (compared to 21% under affirmative action) (Atkinson & Pelfrey, 2006). Similarly, there are several smaller liberal arts colleges that have very comprehensive application processes with sometimes optional SAT scores. Extracurricular activities, letters of recommendation, personal essays, and personal interviews are usually important elements. Bates College, which introduced such admission criteria about 20 years ago, reported that enrollment increased for women, minorities, and international students during that time period (Syverson, 2007). Similar experiences are reported of Reed College (Diver, 2005) and Providence College (Shanley, 2007). However, completely abolishing the SAT might not be feasible for larger institutions (Zwick, 2007). Nevertheless, comprehensive measures should always be a priority. Even the College Board (the association that administers the SAT) suggests that the SAT should always be used in conjunction with other measures such as portfolios, writing samples, personal statements, school

records, counselor recommendations, and other criteria (College Board, 2002). Also *The Principles of Good Practice* published by the National Association for College Admission Counseling advises against the use of a minimum test score as a sole criterion for admission (NACAC, 2009).

Affirmative action policies are the best way to ensure a diverse population in light of the reality that students do not have the same educational opportunities. Unfortunately, quick solutions such as percentage plans or abolishing the SAT do not have the same effect. Research has shown that only through increased efforts in outreach, informing students and parents, and a comprehensive review of applicants, can diversity in higher education be improved. More research is needed to find new ways of increasing diversity despite restrictions on affirmative action policies (Studley, 2004).

Technical and Practical Suggestions for Standardized Tests in College Admissions

Reviewing the professional literature revealed that commonly voiced criticism against the SAT of limited predictability and racial bias do not hold up in most circumstances. The reviewed evidence thus supports a judicious use of the SAT. The SAT was also found not to contribute or diminish the diversity in higher education. However, diversity in higher education is still an issue and needs to be at the forefront of an institution's mission and goals (Smith, 2009). As noted before, a fair test does not ensure its results are used appropriately. Neither does the use of the SAT in admissions assure a fair admissions' process. Hence, the last part of this literature review will summarize recommendations from measurement specialists on how to improve the SAT as well as go over suggestions on its interpretation and use in college admissions.

Even though this literature review found that there is considerable support for the SAT, it should not be taken as evidence that all standardized tests are good. There are bad tests, as well as good tests that are being used inappropriately (Sireci, 2007a). Additionally, to date there is no agency that overlooks companies who develop tests and that monitor testing practices. So far, test agencies seem to be self-regulated and have a minimum motivation to follow good practices as outlined in *The Standards* (AERA et al., 1999) in order to avoid lawsuits. It would thus add credibility and support efforts in fairness if such an agency were created (Zwick, 2002a).

There are several recommendations on how to improve the SAT itself as well as research on the SAT. More research is needed on how to test for validity (Schellenberg, 2004; Sireci, 2004), on expanding item analysis and formats (Banks, 2006; Hambleton & Murphy, 1992), and on improving the practices of detecting bias (Jenks, 1998; Sireci & Khaliq, 2002). More up-to-date research is needed on the SAT in general. Published validity studies that look beyond Freshman GPA are limited and not up to date. Also, more research is needed on the under-prediction of females and little is yet known about the performance of English Language Learners (Zwick, 2004, 2007).

More practical recommendations suggest that there should be a discussion on the amount of testing. Currently, the number of tests a student has to take and the amount of pressure that students are under should be reconsidered (Sireci, 2007a). Also, test preparation programs should be offered by testing companies to reduce fear and anxiety and could be a platform to give information to students and parents and potentially improve motivation (Sackett *et al.*, 2001).

In terms of the use of the SAT in admissions, several suggestions have been made. As Sireci (2007a) noted, there is no inherently good or evil test. However, the use of a test in an appropriate context determines its validity. It is thus paramount that the SAT should be used

according to its purpose. It was already noted that for small colleges that have the luxury of performing in depth analyses of applicants, the SAT might not be necessary. A candidate should be evaluated according to the goals of the institution. Hence, a high test score should not always be preferred over a lower test score (*e.g.*, for artists, athletes, or musicians) (Sackett *et al.*, 2008). In general, there is a consensus among several authors that the SAT and the GPA should be a “bare minimum” of admission criteria (Burton & Ramist, 2001; Sackett *et al.*, 2008; Zwick, 2002a). Especially students in “the gray area” should not be evaluated on only academic measures, but on additional criteria as well as the institution’s overall goal (Burton & Ramist, 2001). Other measures can include personal statement, letters of recommendations, or non-cognitive factors (Robbins *et al.*, 2004; Sackett *et al.*, 2001, 2008). While experts agree that additional measures are needed, it seems that the design and creations of additional measures are difficult. Non-cognitive questionnaires have been proposed and are in use (Sedlacek, 1994, 1997) but a meta-analytic review of 47 studies has found no validity of such a test (Thomas, Kuncel, & Credé, 2007). Since standardized tests do not measure other important criteria for admissions (Sackett *et al.*, 2008), research on these additional measures is vital. The SAT is a test of developed abilities and is based on academic performance and cognitive components. It is well known however, that background variables do affect a student’s academic experience and success (Saegert *et al.*, 2006; Zwick, 2002b). Social environment and involvement, motivation, satisfaction, and institutional factors play an equally important role in college success which is not covered by the SAT (Fleming, 1998; Sackett *et al.*, 2008). In short, the SAT can predict whether someone has the academic capability for college, but other factors play a major role as well for a successful college completion. Admission decisions should thus not be based solely on academics.

Summary

Achieving diversity in higher education is imperative for a successful future. The number of minority students has gone down considerably in states where affirmative action policies have been prohibited. In order to increase these numbers, admission policies need to be carefully reviewed. This literature review looked at one aspect of this admission process. High stakes admission tests such as the SAT are under scrutiny and are sometimes seen as contributing to the problem of diversity in higher education.

The first part of this paper looked at three common criticisms of the SAT. However, little evidence was found to support the three arguments. The SAT is a good predictor of college achievement when adjusting for range restriction and unreliability of the criterion (Bridgeman *et al.*, 2000; Burton & Ramist, 2001). Unfortunately, studies that are critical of the SAT do not adjust for these effects, making comparisons between articles difficult. The issue of racial bias is one of the most important topics in research on testing with several hundred studies published on the topic. Most agree that the SAT over-predicts college achievement for minority students (with under-prediction being evidence of bias) (Bridgeman, *et al.*, 2000; Mattern *et al.*, 2008; Ramist *et al.*, 1994; Young & Kobrin, 2001). The exact reasons for this over-prediction are however not clearly understood. Additionally, there is some evidence of under-prediction for women and little is yet known about English language learners (Zwick, 2002a). The SAT was also found to be predictive of all four years of college GPA (Sackett *et al.*, 2001). However, the SAT does less well in predicting measures that are not directly linked to academic achievement such as graduation or income.

Summarizing the results from this first part underscores the general validity of the SAT but also points to its limitations. The SAT is a test of academic abilities and should only be used in that context. It does not predict other important elements of the college experience such as satisfaction, integration, or perseverance. These limitations are important. Only the correct use and interpretation of a test assures an application without bias.

The second part of the paper looked at alternatives to the SAT. In general, the reviewed research showed that abolishing the SAT in favor of other measures showed little success and often increased the problem of diversity. Percent plans have had little success in increasing minority admissions (Horn & Flores, 2003; Office of Civil Rights, 2002). Similarly, using only HGPA does not solve the problem but replaces a professionally developed test with a teacher test. Simulations have shown that only using GPA does not change the number of minorities admitted to college (Zwick, 1999). The only practice that is encouraged by measurement specialists and is supported by research evidence is a comprehensive approach to admissions. HGPA and SAT scores should be the bare minimum in an admission decision (Burton & Ramist, 2001; Sackett et al., 2008; Zwick, 2002a); portfolios, letters of recommendation, interviews, and non-cognitive elements (e.g. persistence, communication, enthusiasm), give important insight on a student in addition to academic potential. Some small liberal arts colleges that can afford an intensive review of all applicants have been successful in achieving fair admissions with such a comprehensive approach and without the use of SATs (Diver, 2005; Shanley, 2007; Syverson, 2007).

Finally, recommendations from measurement specialists were summarized. It is important to note that the reviewed evidence supporting the SAT does not imply that there are no necessary improvements in testing practices. Several recommendations from experts were listed in all stages of testing: development, analysis, and use. Even though there is a large body of research available, not all research is up-to-date. More studies are needed that reflect the current circumstances and state of tests. In general however, the main recommendations for the current situation can be summarized as an urgent call towards comprehensive evaluations in admission. Stating a minimum SAT score will not achieve positive results. Clearly defined admission criteria that are in line with an institution's goal and mission are vital. Finally, Atkinson and Pelfrey (2006) called for an expansion of the definition of merit and urged to move away from a narrow focus and understanding of who is allowed to go to college.

Conclusions

The educational system has an inherent problem. Demand for higher education is higher than the supply, especially at more prestigious institutions. But basing admissions on past educational achievement will always reward the ones with better educational opportunities (Bakst, 1998). Therefore, the regularly observed mean differences of racial groups on all educational indicators guarantee a discrepancy in enrollment (Sackett *et al.*, 2008). When subgroups have lower mean scores, it becomes more likely that a smaller proportion of this group will be admitted to college (Sackett & Wilk, 1994). The ideal situation would be to fix the educational pipeline and provide equal opportunity to all. However, affirmative action policies would be a more immediate solution to ensure diversity in higher education in the current situation (Bowen & Bok, 1998). The very least that institutions can do is ensure a comprehensive evaluation of each student beyond simply academics and increase their efforts to reach out to

minority populations. More research is needed however to find new ways to increase diversity in higher education.

I believe that a polemic discussion on tests and testing only diverts attention from the real problem. Tests need to be continually improved, but to think that the large educational problems “are essentially issues of test bias is to be deceived” (Cole, 1981, p. 1075). A fixation on blaming tests and seeing the abolishment of testing as the solution to the problem of diversity in higher education is misguided. Equally dangerous is an uncritical acceptance of tests without an understanding of their design and limitations. Some authors have compared the use of tests to the use of thermometers (Sireci, 2004). The “thermometers” consistently point to a “fever” in the educational system. It is time to reduce the fever.

Reference List

- Alon, S., & Tienda, M. (2007). Diversity, opportunity, and the shifting meritocracy in higher education. *American Sociological Review*, 72(4), 487–511.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- Atkinson, R. C. (2002). Achievement versus aptitude in college admissions. *Issues in Science and Technology*, 18(2), 31–36.
- Atkinson, R. C., & Pelfrey, P. A. (2006). Opportunity in a democratic society: Race and economic status in higher education. *Proceedings of the American Philosophical Society*, 150(2), 318–332.
- Bakst, D. L. (1998). *Hopwood, Bakke, and beyond: diversity on our nation's campuses*. American Association of Collegiate Registrars and Admissions Officers, Washington, DC.
- Banks, K. (2006). A comprehensive framework for evaluating hypotheses about cultural bias in educational testing. *Applied Measurement in Education*, 19(2), 115–132.
- Bowen, W. G., & Bok, D. C. (1998). *The shape of the river: Long-term consequences of considering Race in college and university admissions* (1st ed.). Princeton University Press.
- Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). Predictions of freshman grade point average from the revised and recentered SAT I: Reasoning Test. *New York: College Entrance Examination Board*.
- Burton, N. W., & Ramist, L. (2001). Predicting success in college: SAT studies of classes graduating since 1980. *New York: College Entrance Examination Board (= College Board Research Report, 2)*.
- Camara, W. J., & Schmidt, A. E. (1999). *Group differences in standardized testing and social stratification*. New York: College Entrance Examination Board.
- Carnevale, A. P., & Rose, S. J. (2003). *Socioeconomic status, race/ethnicity, and selective college admissions*. Century Foundation.
- Cataldi, E. F., Laird, J., & KewalRamani, A. (2009). *High school dropout and completion rates in the United States: 2007*. Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009064>.
- Cole, N. S. (1981). Bias in testing. *American Psychologist*, 36(10), 1067–1077.

- College Board. (2002). *Guidelines on the uses of college board test scores and related data*. New York: The College Board.
- Crouse, J. (1988). *The case against the SAT*. Chicago: University of Chicago Press.
- Diver, C. (2005). Is there life after rankings? *Atlantic Monthly*, 296(4), 136.
- Fleming, J. (1998). Correlates of the SAT in minority engineering students: an exploratory study. *Journal of Higher Education*, 69(1), 89–108.
- 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. *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.
- Gould, S. J. (1996). *The mismeasure of man* (2nd Revised ed.). New York: W. W. Norton.
- Hambleton, R. K., & Murphy, E. (1992). A psychometric perspective on authentic measurement. *Applied Measurement in Education*, 5(1), 1–16.
- Hawkins, D. A., & Lutz, J. (2007). State of college admission. *National Association for College Admission Counseling, Alexandria, VA*.
- Hedges, L. V., & Nowell, A. (1998). Black-white test score convergence since 1965. In C. Jenks & M. Phillips (Eds.), *The black-white test score gap* (pp. 149–181). Washington, D.C.: Brookings Institute.
- Horn, C. L., & Flores, S. M. (2003). *Percent plans in college admissions: A comparative analysis of three states' experiences*. Civil Rights Project, Harvard University.
- Hunter, J. E., & Schmidt, F. L. (2000). Racial and gender bias in ability and achievement tests: Resolving the apparent paradox. *Psychology, Public Policy, and Law*, 6(1), 151–158.
- Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, 91(3), 594–612.
- Jenks, C. (1998). Racial bias in testing. In C. Jenks & M. Phillips (Eds.), *The black-white test score gap* (pp. 55–85). Washington, D.C.: Brookings Institute.
- Jenks, C., & Phillips, M. (Eds.). (1998). *The black-white test score gap: An introduction*. Washington, D.C.: Brookings Institute.

- Krueger, A., Rothstein, J., & Turner, S. (2006). Race, income, and college in 25 years: Evaluating Justice O'Connor's conjecture. *American Law and Economics Review*, 8(2), 282–311.
- Lawshe, C. H., Bolda, R. A., Brune, R. L., & Auclair, G. (1958). Expectancy charts II. Their theoretical development. *Personnel Psychology*, 11(4), 545–559.
- Linn, R. L. (1973). Fair test use in selection. *Review of Educational Research*, 43(2), 139–161.
- Lubinski, D., Benbow, C. P., Webb, R. M., & Bleske-Rechek, A. (2006). Tracking exceptional human capital over two decades. *Psychological Science (Wiley-Blackwell)*, 17(3), 194–199.
- Mattern, K. D., Patterson, B. F., Shaw, E. J., Kobrin, J. L., & Barbuti, S. M. (2008). Differential validity and prediction of the SAT (College Board Research Report 2008-4). *New York, NY: The College Board*.
- Mendoza, J. L., & Mumford, M. (1987). Corrections for attenuation and range restriction on the predictor. *Journal of Educational and Behavioral Statistics*, 12(3), 282–293.
- Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, 30(10), 955–966.
- (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5–11.
- National Association for College Admission Counseling (NACAC). (2009). Statement of principles of good practice. The National Association for College Admission Counseling.
- Office of Civil Rights. (2002). *Beyond percentage plans: the challenge of equal opportunity in higher education*. DIANE Publishing.
- Phelps, R. P. (2009). *Correcting fallacies about educational and psychological testing* (1st ed.). Washington, DC: American Psychological Association.
- Phillips, M. (2006). Standardized tests aren't like t-shirts: One size doesn't fit all. *Multicultural Education*, 8(2), 52–55.
- Phillips, M., Brooks-Gunn, J., Duncan, G. J., Klebanov, P., & Crane, J. (1998). Family background, parenting practices, and the black-white test score gap. In C. Jenks & M. Phillips (Eds.), *The black-white test score gap* (pp. 103–148). Washington, D.C.: Brookings Institute.
- Ramist, L., Lewis, C., & McCamley, L. (1990). Implications of using freshman GPA as the criterion for the predictive validity of the SAT.

- (1994). Student group differences in predicting college grades: Sex, language, and ethnic groups (College Board Report No: 93-1, ETS No. 94.27). *New York: College Entrance Examination Board.*
- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*(2), 261.
- Roth, P., Bevier, C., Bobko, P., Switzer, F., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54*(2), 297–330.
- Rothstein, R. (2005). The limits of testing: Standardized tests are fine for some purposes, but they can't assess creativity, insight, or many other important traits. *American School Board Journal, 192*(2), 34–37.
- Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist, 63*(4), 215–227.
- Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist, 56*(4), 302–18.
- Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist, 49*(11), 929.
- Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*(1), 112–118.
- Saegert, S. C., Adler, N. E., Bullock, H. E., Cauce, A. M., Liu, W. M., & Wyche, K. F. (2006). Report of the APA task force on socioeconomic status. Retrieved on July, 18, 2007.
- Schellenberg, S. J. (2004). Test bias or cultural bias: Have we really learned anything? National Association of Test Directors 2004 Symposia. Retrieved from <http://www.natd.org/2004Proceedings.pdf>
- Sedlacek, W. E. (1994). Issues in advancing diversity through assessment. *Journal of Counseling & Development, 72*(5), 549–53.
- (1997). Admissions in higher education: Measuring cognitive and noncognitive variables. *Minorities in higher education, 98*, 47–71.
- Shanley, B. J. (2007). Test-optional admission at a liberal arts college: A founding mission affirmed. *Harvard Educational Review, 77*(4), 429–434.

- Sireci, S. G. (2004). How psychometricians can help reduce the achievement gap: Or can they? National Association of Test Directors 2004 Symposia. Retrieved from <http://www.natd.org/2004Proceedings.pdf>
- (2007a). Are educational tests inherently evil? *Nonpartisan Education Review*, Nonpartisan Education Review, 3(4). Retrieved from <http://ideas.repec.org/a/teg/journal/v3y2007i4p1-5.html>
- (2007b). On validity theory and test validation. *Educational Researcher*, 36(8), 477–481.
- Sireci, S. G., Han, K. T., & Wells, C. S. (2008). Methods for evaluating the validity of test scores for English language learners. *Educational Assessment*, 13(2/3), 108–131.
- Sireci, S. G., & Khaliq, S. N. (2002). NCME members' suggestions for recruiting new measurement professionals. *Educational Measurement: Issues and Practice*, 21(3), 19–24.
- Smith, D. G. (2009). *Diversity's Promise for Higher Education: Making It Work*. The Johns Hopkins University Press.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), 613–29.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811.
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 379–440). San Diego: CA: Academic Press.
- Stricker, L. J., Rock, D. A., & Burton, N. W. (1993). Sex differences in predictions of college grades from scholastic aptitude test scores. *Journal of Educational Psychology*, 85(4), 710.
- Studley, R. E. (2004). Inequality, student achievement, and college admissions: A remedy for underrepresentation. In R. Zwick (Ed.), *Rethinking the SAT: The Future of Standardized Testing in University Admissions* (1st ed., pp. 321–344). Routledge.
- Syverson, S. (2007). The role of standardized tests in college admissions: Test-Optional admissions. *New Directions for Student Services*, 2007(118), 55–70.
- Thomas, L. L., Kuncel, N. R., & Credé, M. (2007). Noncognitive Variables in College Admissions. *Educational and Psychological Measurement*, 67(4), 635 –657.

- Thorndike, R. M. (2005). *Measurement and evaluation in psychology and education* (7th ed.). Upper Saddle River, NJ: Pearson.
- Wightman, L. F. (2003). Standardized testing and equal access: A tutorial. In M. J. Chang (Ed.), *Compelling interest: examining the evidence on racial dynamics in colleges and universities*. Stanford University Press.
- Wilson, W. J. (1998). The role of environment in the black-white test score gap. In C. Jenks & M. Phillips (Eds.), *The black-white test score gap* (pp. 501–510). Washington, D.C.: Brookings Institute.
- Young, J. W., & Kobrin, J. L. (2001). Differential validity, differential prediction, and college admission testing: A comprehensive review and analysis. *College Entrance Examination Board*.
- Young, John W. (1991). Improving the prediction of college performance of ethnic minorities using the IRT-based GPA. *Applied Measurement in Education*, 4(3), 229.
- Zwick, R. (1999). Eliminating standardized tests in college admissions. *Phi Delta Kappan*, 81(4), 320.
- (2002a). *Fair game? The issue of standardized admissions tests in higher education* (1st ed.). Routledge.
- (2002b). Is the SAT a “wealth test”? *Phi Delta Kappan*, 84(4), 307–11.
- (2007). College admissions in twenty-first-century America: The role of grades, tests, and games of chance. *Harvard Educational Review*, 77(4), 419–428.
- (Ed.). (2004). *Rethinking the SAT: The future of standardized testing in university admissions* (1st ed.). Routledge.
- Zwick, R., & Sklar, J. C. (2005). Predicting college grades and degree completion using high school grades and SAT scores: The role of student ethnicity and first language. *American Educational Research Journal*, 42(3), 439–464.