AUTHOR

Joseph A. Soares, Ph.D.
Wake Forest University

CORRESPONDENCE

Email
soaresja@wfu.edu

For Tests that are Predictively Powerful and Without Social Prejudice

In Philip Pullman's dark matter sci-fi trilogy[1], there is a golden compass that in the hands of the right person is predictively powerful; the same was supposed to be true of the SAT/ACT – the statistically indistinguishable standardized tests for college admissions. They were intended to be reliable mechanisms for identifying future trajectories, not unlike a meritocratic fortune telling device. In Pullman's novels, the compass works; however in the real world, the predictive accomplishments of the SAT/ACT are sadly less dramatic.

Pullman's novels also posit the existence of multiple parallel universes where enlightenment and love struggle against dogma and hate. If multiple universes exist, surely some of them by now have worked out how to make college admissions meritocratic, for even we are approaching that goal in the early part of our twenty-first century. We have labored long, misdirected by an old-compass admissions system, designed in the heyday of eugenics, which worked more effectively to exclude social "undesirables" than to include those who were academically fit. In the last two decades, however, nearly a third of our four-year-degree-granting institutions have gone "test-optional" breaking in part or whole with the old-compass camp. New tools, often called non-cognitive tests, which statistically outperform previous tests and do so without transmitting social disparities, have been used by thousands of students at universities as diverse as Tufts, DePaul, and Oklahoma State. Today, there are good reasons to be optimistic about the progress being made in the real world.

Test Scores Add Little to High School GPA

What I am referring to here as the old-compass admissions system "is the 20th century formula of looking at high-school record and one of two standardized tests, either the SAT or ACT, in order to predict grades in the first year of college" (Soares, 2012b, p.

66). The scientific prowess of the old method was never found to be very great, predicting at best, according to the test makers, about 21% of the variance in college grades (Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008). The contribution of each part of the old system, however, high school grades and test scores, was far from equal.

Though many parents and academics are surprised by this, it remains true that high school grades have always done a better job in predicting college grades than test scores. As the Emeritus President of the University of California, Richard Atkinson, and Berkeley statistician Saul Geiser, remind us, "Irrespective of the quality or type of school attended, cumulative grade point average (GPA) in academic subjects in high school has proved to be the best overall predictor of student performance in college. This finding has been confirmed in the great majority of 'predictive-validity' studies conducted over the years, including studies conducted by the testing agencies themselves" (Atkinson & Geiser, 2012, p. 24). In technical articles, for statistical *cognoscenti,* the College Board concedes that high school grades matter most, but for *hoi polloi* of the press, they go "truth optional" and unabashedly claim that the test predicts best (Kobrin et al., 2008; Morgan, 1989; for press coverage, see: http://thechoice.blogs.nytimes.com/2011/11/09/sat/).

Because the SAT and ACT tests are less predictive than the high-school record, the real question is, how much value do they add? Youths and their families should not have to suffer through the time, expense, and effort to take a test that stands outside the high school curriculum, unless it raises to a higher level our ability to identify college-ready talent. When statisticians attempt to model outcomes such as SAT scores, which range from 200 to 2800, they use linear multiple regressions to measure the contribution that each variable makes to the explanatory power, or R-square, of the model. The test industry claims to find an 8-point boost, raising one's R-square from 13% with high school GPA alone, to 21% with the SAT (Kobrin et al., 2008). Independent researchers, however, most often find an increase of merely 2 points (Soares, 2012a). As one can see from examples in my book, *SAT Wars,* institutional validity studies show that the SAT increased Johns Hopkins' R-square by two percentage points, raising their models' explanatory punch from an R-square of 0.18 to 0.20; at the University of Georgia it added one percent, raising their R-square from 0.30 to 0.31; and at DePaul the ACT was found to contribute one percent, raising their R-square from 0.19 to 0.20. Independent scholars found that neither the SAT nor the ACT adds more than a few percentage points to what is already known from high school GPA. For a billion-dollar industry, this is pretty pathetic value added for the money.

If the SAT/ACT improves one's predictive model by just one or two percentage points, how could that be worth the costs? Those tests do not lift college admissions out of the realm of practical wisdom into the realm of applied science. When 70 to 80 percent of the variance in college grades is left unexplained by our best statistical models, it is time again to acknowledge that admissions professionals do not have a golden compass; they are making decisions that remain more art than science. A false sense of scientific precision is one type of collateral damage done by the test industry. When test scores are used to set floors below which admissions staff will not go, we are doing an injustice to thousands of students; and when we decide between students based on a test score difference, we are relying on a compass that cannot find true north.

Some Tests Calcify Social Disparities

In addition to being largely redundant with information provided by the high school transcript, these particular tests are discriminatory. Not all tests disguise social selection as academic selectivity, but the SAT and ACT do. Admissions by the old-compass method "narrows the socioeconomic and racial diversity of one's pool and yield. The more one relies on SAT/ACT/LSAT-type standardized tests, the more social disparities unfavorable to racial minorities, women, and low SES youths are passed along" (Soares, 2012b, p. 66). Those tests tell us that women are less quantitative than men, because females score on average 33 points lower than males on math sections. Hispanics/Mexican Americans and Blacks are "dumb and dumber," with the former falling 219 points, and the latter 303 points, on average behind Whites.

When 70 to 80 percent of the variance in college grades is left unexplained by our best statistical models, it is time again to acknowledge that admissions professionals do not have a golden compass; they are making decisions that remain more art than science.

Test score disparities by gender and race do not end the list of demographic problems with the test. Family income has a strong linear relation to test score: the higher one's family's income, the higher the average test score. In fact, test scores correlate more strongly with family income than with high school grades. Students from poor families, those earning less than 20 thousand dollars annually, score 100 points lower than students from families earning near the median range in America, between 40 and 50 thousand dollars per year; and further, those students from median income families score 200 points behind students from families earning over 100 thousand dollars annually.

Some researchers have expressed the concern that HSGPA might be more correlated with family socioeconomic status (SES) than is the case for SAT scores (Sternberg, Bonney, Gabora, & Merrifield, 2012). An argument used to defend the SAT/ACT is that these tests level the playing field, providing for a nationally-normed test that reduces disparities among high schools due to the property values of the neighborhood and the SES composition of the student body. But University of California researchers found the opposite. Geiser and Santelices (2007) "reported that the SAT-V correlated at the .32 level with family income, and at the .39 level with parents' education; similarly, SAT-M scores correlated respectively at .24 and .32, but HSGPA correlated with family income at the .04 level, and with parents' education at the .06 level" (p. 2). If Geiser and Santelices are right, HSGPA is far from being a proxy for social class. Since HSGPA retains its punch, without conveying social disparities, then why not save money, energy, and incalculable family anxiety by dropping the SAT/ACT? (Soares, 2012b).

One indirect effect of the SES selection accomplished by using these tests is an economic payoff for institutions in higher education. Colleges can balance their budget with full-fare paying families if they can advertise high average test scores for admitted students. The higher the college's average score, the more economically affluent the next year's applicant pool. Prospective students will self-select away from or toward institutions based on test scores, and in doing so assure that very selective colleges are economically homogenous and privileged. Bank accounts, not brains, determine which birds flock together. Needs-blind admissions furthers the SES charade, because only the most economically exclusive colleges can afford to bank on an applicant pool so affluent that it never risks admitting more needy students than it can afford to cover (Soares, 2007). Some will say, if these tests select for youths from families with higher incomes, and against women, Hispanics, and Blacks, is that just a reflection of our society's inequalities in academic preparation? Is it not the case that White males from affluent families are going to receive the most resources and attention from their families and schools? Perhaps, the test is fair and the group disparities it displays are just a measure of life's unfairness?

I have already offered for your consideration Geiser and Santelices' (2007) finding that family income and parents' education correlate with test scores but do not correlate with grades earned in high school. From their work, one can see that SAT selection promotes social disparities not captured by selection based mainly on HSGPA. Selection by test scores stratifies higher education into a class system: the higher one's college's selectivity, the higher the SES composition of one's student body (Soares, 2007). Evidence is also available from the University of Texas, where the natural experiment of admitting all students in the top ten percent of each high school class were admitted, which enabled racial and social class diversity, without detriment to the students or the university.

One indirect effect of the SES selection accomplished by using these tests is an economic payoff for institutions in higher education. Colleges can balance their budget with full-fare paying families if they can advertise high average test scores for admitted students.

Table 1
Percent of Each Higher Education Tier Occupied by Each SES Quartile

SES Quartiles	% of Tier 1	% of Tier 2	% of Tier 3	% of Tier 4	% of Tier 5	% of Tier 6	% of Tier 7
Top	79	64	51	37	23	36	22
Upper Middle	16	19	24	27	28	21	25
Lower Middle	3	9	14	23	28	24	27
Bottom	2	7	10	13	20	19	25

Source: National Educational Longitudinal Survey, 1988-2000. U.S. Department of Education. Restricted Access Data License Control Number: 06011044 (as cited in Soares, 2007).

As described in my book, *SAT Wars,* "The Vice-Provost [at the University of Texas in Austin] for admissions, Dr. Bruce Walker, has released multiple reports on the 10% solution showing how high school ranking is an excellent and reliable predictor of college GPA and graduation" (Soares, 2012a, p. 203). Class rank reduces, rather than passes along, SES disparities. "Being in the top 10 percent of any high school graduating class, allows a youth to overcome the disadvantages of coming from a low income family; of having parents without high school degrees; and of attending a low performing high school. Top 10 percent youths from families with the lowest incomes, below $20,000 per year, and from the least desirable high schools, those officially ranked "low performing," do better academically at the University of Texas than youths below the top 10 percent from "exemplary" high schools, who are from high-income families, and with college-educated parents" (p. 203).

If school grades and class rank are less influenced by SES than the SAT/ACT, the absence of SES effects on high school grades could be due to the stratification that divides students by race and class into different high schools in the first place. Again, one may still argue that the test is just a reflection of life's inequalities. But there is another, more sinister possibility. What if the test has social discrimination built into it? What if the questions used on the test systematically favor some groups over others?

Test Question Selection and Social Bias

In *SAT Wars*, Jay Rosner, the vice-president of the Princeton Foundation, offers shocking evidence of systematic bias in the SAT's logic of question selection. The questions that count on each year's version of the SAT are drawn from experimental questions that are pretested in previous years. Each test combines questions that will count for that year's scores and experimental questions that are being vetted to see how they perform for future use. The difference between a good experimental question and a bad one is whether it retains the bell curve shape of test score results. The SAT has retained the same bell curve distribution ever since 1926, which some take as a measure of its validity, rather than as an indicator of its role in transmitting social disparities. Working with two years of national SAT data, Rosner (2012) found there are few "neutral" test questions, in the sense that both men and women, Blacks and Whites, all perform equally well or equally poorly on them. Rather, all but one or two questions in each section of the real test are questions that, when they were rolled out in the experimental section of previous tests, students performed differently on those experimental questions based on their demographic profile: race, gender, and family income. Students taking the test are invited, for reasons of research, to voluntarily provide demographic information on themselves. Rosner presents in chapter six of SAT Wars examples of math questions that women, and verbal questions that Blacks and Hispanics, outperform males and Whites on, respectively.

Here is one example of a verbal sentence completion question that produces a racial test score gap: "The actors bearing on the stage seemed _____; her movements were natural and her technique _____." Rosner then provides the five possible word-combination answers that were used on the SAT, tells the reader that the correct answer was "(C) unstudied … uncontrived" and invites the reader to guess whether this was a question Whites outperformed Blacks on or the reverse. One may think this looks like a good question, using terms that belong in a college student's vocabulary, but that is incorrect. Rosner informs us that this is a Black advantage question, on which Black youths outperform Whites; and because of that, it does not make it onto next year's SAT exam. This question never counted. Rosner finds that out of the 156 verbal questions on two years of the SAT that counted, zero were questions like the one above, on which Blacks scored better than Whites (Rosner, 2012). All verbal questions on the SAT have been White advantage questions. I am not going to provide additional examples of racial bias or gender bias in the question selection step for the SAT, because I would like you all to read Rosner's contribution. But I will say that if I were able to pick next year's questions, rather than rely on a statistical algorithm that retains a bell curve, I could eliminate the test's gender gaps on math scores and racial gaps on verbal scores.

> **What if the test has social discrimination built into it? What if the questions used on the test systematically favor some group over others?**

Non-Cognitive Tests Predict Better, and Without Social Prejudice

To those who still believe that test score disparities by demographic groups are just a reflection of life's unfairness, I would point to the chapter in *SAT Wars* written by Robert Sternberg. Sternberg was dean at Tufts University when that institution adopted the Kaleidoscope project to assess applicants' creative and practical problem solving ability. They found these "non-cognitive" tests performed statistically better than the SAT in predicting grades and college retention; and they did so without any gender or racial test score gaps. The January-March 2012 issue of *Educational Psychologist* provides case studies of non-cognitive tests for undergraduate admissions and for law school admissions that are simultaneously predictively more powerful and without transmitting the social disparities of the SAT or LSAT. There are tests that predict without prejudice. We are not inescapably compelled to transmit society's previous social inequalities.

Checkered History of Admissions Tests

Tests and college admissions have a century-long troubled history. Public universities, roughly between the 1890s and the late 1950s, used to admit everyone with a high school degree from a certified public high school. Then, in the 1950s, mid-western public universities developed the ACT as an alternative to the SAT; once the University of California, under Kerr's presidency, wanted to compete with Harvard, it signed up for the SAT in 1968, against the recommendations of every study produced by the university (see John Douglass' account in chapter 3 of *SAT Wars*), making standardized testing rather than high school grades the passkey to higher education. The direct link between public universities and public high schools was cut mid-century.

Private institutions launched the College Board in 1900 to set common exams on academic subjects that would give bragging rights to the private sector. Private colleges did not accept just any high school graduate, but only those who could do college level work in a particular subject as signified by their College Board exam scores. Then the Jewish community in New York blew by that academic hurdle, creating at Columbia University a Jewish foothold on the college/social mobility ladder. Fearing a Jewish invasion, anti-Semitic Yale and Princeton wanted an I.Q. test that would show, in the words of the Princeton psychologist who oversaw the design of the test, the superiority of their Nordic youths over inferior racial stock: the Alpine, Mediterranean (including Jews), and Negro (Soares, 2007). In the 1920s, I.Q. eugenics were not just an intellectual sub-culture, but rather the reverse - they were the law of the land with "separate but equal," forced "three generations of imbeciles are enough" sterilization, and strict immigration quotas. When the SAT was introduced in 1926 it was supposed to be an IQ test that would measure intrinsic intellectual aptitude, not academic subject mastery; it was supposed to help sort between the gems in the Nordic race from the subject-test grinds in the "Jewish race". It did not work to exclude Jews, but other tactics introduced in the 1930s of requiring mother's maiden name and place of birth, were more effective toward that goal. It also did not work to predict grades. Yale and Princeton knew that as early as 1930 (Soares, 2007). But the private sector clung to the test, first for the invidious distinction over public universities of requiring a national normed measuring stick, later because of the convenient way it disguised SES selection as academic selection, paying the bills along the way.

The lasting legacy was a pseudo-IQ test that sorted students by family income, opening or closing doors to colleges and careers in the process. We have traveled some considerable distance since then. The SAT's owners long ago discontinued using the name and the claim that it measured scholastic aptitude. Now the letters "SAT" do not reference anything, and the College Board only really claims the test predicts first year grades, which it does, but not well. There are significant defections even among the ranks of those who continue to embrace IQ bell curves. I take some considerable pleasure that Charles Murray, an author of the highly controversial *Bell Curve*, a man who believes firmly in the importance of IQ, joins me in calling for the abolition of the test. As Murray says in *SAT*

I find myself arguing that the SAT should be ended. Not deemphasized, but no longer administered... [T]he SAT score, intended as a signal flare for those at the bottom, has become a badge flaunted by those on the top.

Wars, "The evidence has become overwhelming …. [S]o I find myself arguing that the SAT should be ended. Not just deemphasized, but no longer administered. … [T]he SAT score, intended as a signal flare for those at the bottom, has become a badge flaunted by those on the top" (Murray, 2012, p. 69). I also agree with Murray that the test will end when any of the top colleges, such as Harvard or Stanford, break with the farce. Murray wrote, "If just those two schools took such a step, many other schools would follow suit immediately, and the rest within a few years. …Admissions officers at elite schools are already familiar with the statistical story … They know that dropping the SAT would not hinder their selection decisions" (Murray, 2012, p. 80). It is high time for higher education to set aside the old golden compass, and to strike out for admissions tools worthy of the 21st century.

Test-Optional Admissions: Theory and Practice

> **Our non-test-score undergraduates perform academically as well as our test score submitters. We have not suffered any lowering of academic standards from the new policy; rather, there is considerable evidence of the reverse.**

In *SAT Wars* there is a chapter jointly authored by two Princeton academics, a statistician, Chang Young-Chung, and sociology professor, Thomas Espenshade, which uses national data to model the impact on academic excellence and social diversity of doing admissions without relying on the SAT or ACT (Espenshade & Chung, 2012). Espenshade and Chung found results differed by type of institution. Private colleges were best served by going "test-optional." In their statistical simulation, private colleges got more racial and SES diverse and academically stronger students, as judged by high school grades and AP exam scores, by going test-optional. Public universities, on the other hand, did best by an admissions policy they dubbed "don't ask, don't tell," where the institution would not even look at test scores. State universities got academically stronger students, and more social diversity when they admit without any reference to test scores. It is a lesson reinforced by the findings on high school grades and standardized tests from the University of Georgia in chapter 8 of *SAT Wars* and by the findings of Bowen, Chingos, and McPherson's *Crossing the Finish Line: Completing College at America's Public Universities* (2009). Public universities waste taxpayers' money, distract students from focusing on learning the curriculum, and practice social discrimination when they require SAT/ACT scores.

In *SAT Wars*, I show how Wake Forest University's experience, now four years old, of conducting test-optional admissions has confirmed the statistical forecast offered by Espenshade and Chung (2012). In the academic year after the May 2008 announcement of Wake Forest's test-optional policy,

> Our applicant pool, even in the worse economic year in recent history, went up by 16%; our minority applicants went up by 70%. As reported in the *Journal of Blacks in Higher Education*, 6% of Wake Forest's senior cohort were minorities of color before the policy change; in the two [now three] cohorts admitted thus far as test-optional, the percentage of Black and Hispanic has gone up to 23. Asian student numbers have increased to 11%. First-generation youths, where neither parent went to college, jumped to 11%; Pell Grant youths, whose families earn near the poverty line, nearly doubled to 11%. In 2009, 78% of WFU undergraduates came from outside North Carolina (Soares, 2012a, p. 207).

Our academic strength has grown as well, as measured by entering students from the top ten percent of their high school class, which has gone up from 65 percent in 2008 to 83 percent in 2011 (Soares, 2012a). For research purposes and to monitor the test-optional policy, Wake Forest requires everyone admitted without a test score to send one before he or she arrives on campus. Accurate scores are reported to ratings publications, so no one can accuse the university of using this policy to artificially inflate our standing in the ratings game. Matriculating students are examined each semester to determine whether there are any differences between students who do or do not submit test scores. Wake Forest looks at course enrollment patterns, withdrawals from classes or from college, and grades achieved. As reported in detail in my conclusions in *SAT Wars*, we have found no statistically significant differences. Our non-test-score undergraduates perform academically as well as our test-score submitters. We have not suffered any lowering of academic

standards from the new policy; rather, there is considerable evidence of the reverse. We have found, along with the percent of our students from the top 10% of their high school classes having gone up dramatically, that library usage has increased as well. "Librarians are marvelous for keeping track of their domain, and from them we learned that library usage went way up: 63% increase in personal research sessions; 55% increase in instructional library sessions; 26% increase in credited library instructional classes; daily average visits went up by 10%; daily unique library web site visits went up by 62%" (Soares, 2012a, p. 209). Campus life, in and out of the classroom, looks and feels more diverse, more stimulating, and more engaging than ever before. When Wake Forest went test-optional, there were about 775 higher education institutions in that camp; today our ranks number 856. With nearly a third of all four-year degree granting institutions already with some form of test-optional admissions, the tipping point to push past the SAT/ACT is within sight.

References

Atkinson, R., & Geiser, S. (2012). Reflections on a century of college admissions tests. In J. A. Soares (Ed.), *SAT wars: The case for test–optional college admissions* (pp. 23-49). New York, NY: Teachers College Press.

Bowen, W., Chingos, M., & McPherson, M. (2009). *Crossing the finish line: Completing college at America's public universities.* Princeton, NJ: Princeton University Press.

Espenshade, T., & Chung, C. Y. (2012). Diversity outcomes of test-optional policies. In J. Soares (Ed.), *SAT wars: The case for test–optional college admissions* (pp. 177-200). New York, NY: Teachers College 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 indicator of four–year college outcomes* (Occasional Paper Series, 6.07). Berkeley, CA: Center for Studies in Higher Education.

Kobrin, J., Patterson, B., Shaw, E., Mattern, K., & Barbuti, S. (2008). *Validity of SAT for predicting first year college grade point average* (Report No. 2008-5). New York, NY: College Board.

Morgan, R. (1989). *Analyses of the predictive validity of the SAT and high school grades from 1976 to 1985* (Report No. 89-7). New York, NY: College Board.

Murray, C. (2012). Abolish the SAT. In J. Soares (Ed.), *SAT wars: The case for test–optional college admissions* (pp. 69-81). New York, NY: Teachers College Press.

Rosner, J. (2012). The SAT: Quantifying the unfairness behind the bubbles. In J. Soares (Ed.), *SAT wars: The case for test–optional college admissions* (pp. 104-117). New York, NY: Teachers College Press.

Soares, J. A. (2007). *The power of privilege: Yale and America's elite colleges.* Stanford, CA: Stanford University Press.

Soares, J. A. (Ed.). (2012a). *SAT wars: The case for test–optional college admissions.* New York, NY: Teachers College Press.

Soares, J. A. (2012b). The future of college admissions: Discussion. *Educational Psychologist, 47*(1), 66-70.

Sternberg, R.J., Bonney, C., Gabora. L., & Merrifield, M. (2012). WICS: A model for college and university admissions. *Educational Psychologist, 47*(1), 30-41.

Note: The Golden Compass, NY, Knopf, 2002; The Subtle Knife, NY, Knopf, 2001; The Amber Spyglass, NY, Knopf, 2003.