The SAT and Admission: Racial Bias and Economic Inequality

Ethan Biamonte†

November 15, 2013

Abstract

Standardized testing is an important part of college admissions. The most commonly taken standardized test—the SAT—has been argued to have appreciable economic and cultural biases. This is important because the act of using the SAT as a college admission standard oppresses low-income, racial-minority, and female groups. I conclude that it is unethical to use the SAT for college admission based on evidence in income-SAT ratio, the effect of coaching programs, differential item functioning, and the SAT male-female Mathematics gap.

Keywords: Differential Item Functioning (DIF), socioeconomic status, test-prep coaching, undergraduate admissions, College Board

†A.B. in Mathematics and Economics (2016), University of North Carolina at Chapel Hill. Email: biamonte@live.unc.edu.
1 Introduction

The SAT, developed by College Board, is used to evaluate applicants in most U.S. colleges’ admissions. While admissions officers stress that the SAT is only one of many factors considered in admission, it is an important factor: the test is used to calibrate students’ grades (Ritger). This is because educational standards vary among states and schools; an A at Phillips Exeter Academy does not necessarily equal an A at a rural high school in Mississippi. Additionally, Dr. Michele Hernandez, a former admissions officer at Dartmouth College, said that low SAT scores “might mean you will not be in the range [for admission]” (Hernandez). Therefore, SAT scores have a measurable impact on students’ admissions.

The question arises: considering the SAT’s weight in admission, is it a fair assessment of students’ potential? I argue that it is unethical for the SAT be used in college admission because it has cultural and economic biases which oppress low-income groups, racial minorities, and females. First, I demonstrate that there is sufficient evidence to show that the test is biased against each of these groups. Next, I approach its oppressive effects from a theoretical perspective, given the current lack of research in this area (which most likely results from the absence of hard data). Combining its bias and resulting oppression, I argue that it is unethical to use the SAT as an admission standard.
2 Economic Bias in the SAT

One of the most controversial issues is whether the SAT is economically biased toward the wealthy. Collage Board’s publications tend to show that there is no appreciable economic bias. However, according to the aggregate evidence from independent research, students with higher family incomes tend to perform significantly better on the SAT. This is not so much a result of a more privileged upbringing, as some may argue, as it is access to test preparation materials such as coaching programs.

SAT Score v. Income Correlation (Rampell)

As one can see, all three components—critical reading (CR), math (M), and writing (WR)—show an approximately linear relationship with income. Critical reading and Writing in particular have a 130-point disparity between the top and bottom income brackets.
Greg Mankiw, the Chair of the Economics Department at Harvard University, criticized this very chart in his blog.

Suppose we were to graph average SAT scores by the number of bathrooms a student has in his or her family home. That curve would also likely slope upward. (After all, people with more money buy larger homes with more bathrooms.) But it would be a mistake to conclude that installing an extra toilet raises yours kids' SAT scores (Mankiw).

Mankiw goes on to say that “The key omitted variable here is parents' IQ. Smart parents make more money and pass those good genes on to their offspring” (Mankiw).

While SAT scores do show a moderate correlation with general intelligence, $g$, of 0.483 (Frey & Detterman), when this is combined with parent-child IQ heritability of $r = 0.5$ (Plomin et al.), the SAT’s coachability, and controversy over IQ tests’ validity as a measure of intelligence, this is an unconvincing argument.

More compelling is the fact that income itself does not directly cause SAT scores to rise. As Mankiw says, graphing the number of toilets in various households would also likely correlate with SAT scores. Rather, higher family income allows increased access to preparatory materials which do have a causal relationship with SAT scores. Coaching programs, for example, have been shown to significantly raise students’ scores. In the typical study, the effect of coaching was to raise achievement test scores by 0.25 standard deviations (Bangert-Drowns et al.)
Coaching’s $\Delta z = +0.25$ effect becomes elucidating with an example. The standard normal distribution function is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$ 

Thus, if we take a student with a 50th percentile composite SAT score, 1490, we would expect coaching to raise his score to the 60th percentile, 1570 (College Board). An 80-point score increase is quite significant. Given that IQ tests are not coachable, by definition and in practice, and the SAT is, this finding refutes Mankiw’s argument. This finding also refutes College Board’s study, which argues the effect is an increase of around 30 points (Powers & Rock).

High coachability reflects bias. *Ceteris paribus*, if a student from a wealthier family can gain access to ample coaching while a low-income student cannot, the low-income student is at a significant disadvantage.

Perhaps the strongest evidence of income bias in the SAT is that other measures of accomplishment—First-year GPA (FYGPA), high-school rank (HSR), and extracurricular accomplishments—show little to no correlation with family income.

**Correlations of High School Rank and SAT with Socioeconomic Background**

*(Crouse & Trusheim)*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Family Income</th>
<th>Father’s Occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT</td>
<td>0.286</td>
<td>0.296</td>
</tr>
<tr>
<td>HSR</td>
<td>0.029</td>
<td>0.043</td>
</tr>
</tbody>
</table>
High school rank (HSR), has a significantly lower correlation with family income than the SAT does—0.029 versus 0.286. The same pattern appears with father’s occupation.

For FYGPA, Astin found that “the income of a student’s parents has no relationship to [college] freshman GPA, either before or after controlling for high school grades, academic aptitude, and college selectivity” (Astin). Similarly, for extracurricular activities, ETS found that “students from families with different incomes did not significantly differ in the number or level of accomplishments they reported” (ETS).

That HSR, FYGPA, and extracurricular accomplishments have little to no relationship with family income strongly implies that the SAT is biased. FYGPA in particular provides compelling evidence that the more wealthy students do not necessarily have the higher aptitudes that their SAT scores show. If a low-income freshman is performing as well as a high-income freshman, but the high-income freshman has a significantly higher SAT score, the scores have been skewed by bias in the test—perhaps from coaching.

The economic bias in the SAT has significant consequences for low-income groups. As shown earlier, the test carries appreciable weight in admission, and its bias against low-income students lowers their SAT scores relative to that of high-income students. Therefore, the SAT’s economic bias lowers low-income groups’ admission chances, and less will get a college education. According to the U.S. Census Bureau (2000), individuals with Bachelor’s Degrees earned $45,648 versus $23,233 for those with only high school diplomas. Thus, the SAT’s bias not only oppresses low-income groups
by denying them a college education; it also lowers their future standard of living. This is unethical.

3 Racial Bias in the SAT

Along with its economic bias, the SAT has a marked racial bias.

**Average 2009 SAT scores by race and ethnicity (College Board)**

<table>
<thead>
<tr>
<th></th>
<th>Critical Reading</th>
<th>Mathematics</th>
<th>Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>528</td>
<td>Asian 587</td>
<td>Asian 520</td>
</tr>
<tr>
<td>Asian</td>
<td>516</td>
<td>White 536</td>
<td>White 517</td>
</tr>
<tr>
<td>Overall</td>
<td>501</td>
<td>Overall 515</td>
<td>Overall 493</td>
</tr>
<tr>
<td>Am. Indian</td>
<td>486</td>
<td>Am. Indian 493</td>
<td>Am. Indian 469</td>
</tr>
<tr>
<td>Hispanic</td>
<td>455</td>
<td>Hispanic 461</td>
<td>Hispanic 448</td>
</tr>
<tr>
<td>Black</td>
<td>429</td>
<td>Black 426</td>
<td>Black 421</td>
</tr>
</tbody>
</table>

As observed from average 2009 SAT scores, White test-takers perform significantly better than all racial minorities with exception of Asians. African-Americans consistently perform most poorly on the SAT.

One possible explanation is bias in question construction. The SAT uses an experimental section to test questions that may be used on future tests. If they don’t test well, they are scrapped. If they do, they are placed on future tests. Jay Rosner analyzed 276 verbal and math questions from the 1998-2000 SATs. He discovered what he calls “Black questions,” which more Blacks than Whites answered correctly on the experimental sections. These questions never made it onto the scored sections of future
tests. Instead, the SAT contains “White questions” (Rosner). Rosner argues that the questions are geared toward Whites, as test developers are mandated to recreate the norm, and the norm is White males outperforming their peers.

Another form of racial bias that has been researched is Differential Item Functioning (DIF). A DIF question is one in which students “matched by proficiency” and other factors have variable scores, predictably by race, on selected questions. Santelices and Wilson found in 2010 that the SAT has these DIF questions, and favors certain ethnic groups over others (Santelices & Wilson).

While outdated, perhaps the most infamous example of DIF on the SAT is the oarsman-regatta analogy question:

RUNNER: MARATHON ::
A) envoy: embassy
B) martyr: massacre
C) oarsman: regatta
D) referee: tournament
E) horse: stable

The SAT replaced analogies and the Verbal section with Critical Reading and Writing in its 2005 restructuring. However, this still serves as an enlightening example of cultural bias in the test (and one that still remains according to Santelices & Wilson). To borrow the words of Robert Schaeffer who served as Public Education Director of FairTest, “That's incredibly culturally centered. You don't see a regatta in center-city L.A., you don't see it in Appalachia, you don't see it in New Mexico” (Pringle).
Thus, minorities are not necessarily performing worse on the test because they grow up in poorer neighborhoods on average; rather, the questions are racially biased against them. This is shown in DIF questions and is explained by test-takers attempting to recreate the norm as Rosner argues.

4 Gender Bias in the SAT

The second form of cultural bias in the SAT is gender bias. Females with similar ability levels to males tend to perform worse than males on the math section of the test. Hence, females’ abilities are underestimated.

\[
\text{Difference between Male & Female Math SAT Scores 1971-2008 (College Board)}
\]

As shown in the graph, \( \Delta Math \text{ score} = Male_{ave} - Female_{ave} \approx 38 \). This gap has persisted for many decades with little change. This gap is not indicative of males’ superior mathematical abilities. According to Kessel & Linn, who summarized more than a dozen studies of large student groups and colleges such as MIT, Rutgers, and Princeton, young women typically earn the same or higher grades as their male
counterparts in math and other college courses despite having SAT-Math scores 30-50 points lower (FairTest).

One possible explanation is that females are less confident in their answers: “Previous reviewers have suggested that women display lower self-confidence than men across almost all achievement situations” (Lenney). Because females are less self-confident, it follows that they are more likely to doubt their answers on the SAT. Rechecking answers is disadvantage on a test with as much time pressure as the SAT.

Hence, the structure of the SAT itself is biased against females. There is no other explanation given that young women earn higher grades in college math courses. While the overall female demographic is not underrepresented in undergraduate programs, it is unethical that there is a double standard for males and females on the SAT math section.

5 Conclusion

I have demonstrated that each component is unethical, but one might wonder how these components connect. They are clearly not independent of one another: this is a multidimensional issue. A test taker could be African-American, low-income, and female. This individual would be at a great disadvantage for college admission. This is all independent of her skill, too. She could have a high GPA and class rank—neither of which are significantly influenced by her family income or gender as we discussed before. A bright student could be overlooked; her potential wasted. It should be clear that this is unethical, but what are the broader implications of this? How is society affected when
colleges unintentionally turn away future great minds? And these minds are turned away by a test that is biased against them. In a broader sense, this is not only unethical in terms of admissions: it’s damaging to our society.
Works Cited


