



College Board Response to *Harvard Educational Review* Article by Santelices and Wilson

June 2010

The *Harvard Educational Review* has published a research article by Maria Veronica Santelices (Pontificia Universidad Católica de Chile) and Mark Wilson (University of California, Berkeley) that is critical of the Differential Item Functioning (DIF) analyses used in the construction of the SAT[®]. Unfortunately, this work is deeply flawed. It utilizes only partial data sets, focuses on a student sample that lacks representation and diversity, and draws conclusions that do not match the data. Simply stated, this research does not withstand scrutiny.

The SAT is a fair assessment, and many years of independent research support this. It is the most rigorously researched and designed test in the world and is a proven, reliable measure of a student's likelihood for college success regardless of student race, ethnicity or socioeconomic status. There is no credible research to suggest otherwise. While a few critics have promoted the notion that the test results indicate bias in the tests themselves, this theory has been by and large debunked and rejected by the psychometric community.

In reviewing this article, our researchers identified a number of fundamental flaws in the data analysis, and they also expressed serious concerns about the conclusions reached by the authors. Key concerns with this study include the following:

- The researchers present results across what they label as four “SAT test forms.” However, what they acknowledge only through a footnote in the report is that they actually used only two distinct SAT forms (the other two forms were duplicate forms, with the exact same test questions in a different order). Furthermore, their discussion of the results focuses on the one “form” that showed results that best support their conclusions, while ignoring the results from other forms, which did not support their findings and/or which showed contradictory findings.
- The researchers focus discussion on two test items in particular that they say have serious values of DIF. What the researchers fail to mention is that their work showed the same two items not in the “serious” range, when they studied the identical items in another test form containing the exact same questions in a different order. This means that the researchers are presenting inconsistent findings as conclusive results. Furthermore, they fail to mention that their own analysis demonstrates that across the entire score scale, there is no accumulation of DIF of any consequence. This suggests selective reporting of findings.
- The researchers used a sample that is not representative of the entire population of SAT takers and that is too narrow to suggest meaningful results. The sample represents California public school students who graduated in 1995 and 2000. The sample is further restricted by using only students who also were accepted and enrolled in University of California schools. In contrast, the SAT conducts DIF analyses on samples that are representative of the entire population of SAT takers. It would not be appropriate to attribute results from the sample used in the analyses conducted by Santelices and Wilson to the overall SAT population.

The College Board and external researchers perform rigorous and ongoing research concerning all aspects of the SAT, including test and item development policies and procedures. Every SAT question goes through an extensive review with a diverse group of educators from around the country to ensure that each question measures the material taught in high school classrooms. Every SAT question is thoroughly pretested and analyzed for bias. Any item that does not clear the educator reviews or that exhibits different performance across subgroups during pretesting is removed before it appears on an official scored section of the SAT.

The College Board is not involved in the admission process at any college or university nor does it play any role in determining the admission policy of any college or university. It was highly irresponsible of the authors to suggest in their research that admission officers might base decisions “exclusively or predominantly” on SAT scores. We are not aware of a single college in the United States that uses SAT scores as the sole measure for admission, and the College Board’s own score usage guidelines specify that SAT scores should be used in conjunction with other indicators, such as secondary school record (grades and courses), interviews, personal statements, writing samples, portfolios and recommendations, to evaluate the applicant’s admissibility at a particular institution.

As an education organization, the College Board encourages open scholarship regarding the SAT, and we are always open to external review of our work. The College Board has provided data to many external researchers who have requested data on the SAT, including the authors of this study. We are disappointed in the decision by the *Harvard Educational Review* to release this article, which was accepted and published without the benefit of peer review or a differing point of view.

The College Board has made — and will continue to make — improvements to the SAT based on the needs of our members and the students we serve. We are proud of the SAT, the important role it plays in helping students prepare for college and its value as an admission tool.