

Validity of the SAT®

In March 2005, the College Board introduced a revised SAT, with an additional writing section and minor changes in content to the verbal and mathematics sections. Colleges and universities across the United States provided first-year performance data for the fall 2006 entering cohort of first-year, first-time students to validate the use of the SAT in college admissions.

- The final sample included 151,316 students attending 110 colleges and universities across the U.S.
- On average, the sample for this study performed better on the SAT—around 60 points higher per section—and had a higher mean high school grade point average (HSGPA)—around 0.3 higher—than the 2006 college-bound seniors cohort, as expected, given that all of the students in the sample were enrolled in four-year colleges.

Differential Validity and Prediction of the SAT

The purpose of this study was to assess the differential validity and differential prediction of the revised SAT for gender, racial/ethnic, and best language subgroups. Differential validity exists if the magnitude of the test-criterion correlation varies by subgroup. Differential prediction occurs when a test systematically over- or underpredicts the criterion—e.g., first-year college GPAs (FYGPAs)—by subgroup. The results are similar to prior research indicating that changes to the SAT did not substantially impact the differential prediction and validity of the test. A summary of key findings is presented below:

Differential Validity:

- For gender, the results show that the individual sections of the SAT as well as the combination of all three sections are more predictive of FYGPA for females ($r = 0.52$ to 0.58) than males ($r = 0.44$ to 0.50).
- For HSGPA, there is a similar pattern with a larger correlation for females ($r=0.54$) as compared to males ($r=0.52$), although the differences are smaller.
- The results for race/ethnicity show that for the individual SAT sections, the SAT is most predictive for white students with correlations ranging from 0.46 to 0.51, as compared to underrepresented groups with correlations ranging from 0.40 to 0.46.
- Likewise, the multiple correlation for all three sections of the SAT combined is higher for white students than for underrepresented groups ($r=0.53$), with the exception of American Indian students ($r=0.54$).
- For HSGPA, the same pattern emerges with higher correlations for white students ($r=0.56$) as compared to underrepresented groups (correlations ranging from 0.44 to 0.49).
- For best language, the correlations between the individual sections of the SAT, as well as the combination of all three sections, and FYGPA are the highest for students whose best language is English ($r = 0.47$ to 0.54), in the middle for students who select English and another language ($r = 0.41$ to 0.50), and weakest for students whose best language is not English ($r = 0.28$ to 0.42).
- Similarly for HSGPA, the correlations are highest for students whose best language is English ($r=0.55$) and lowest for students whose best language is a language other than English ($r=0.35$).

Differential Prediction:

- The individual sections of the SAT, as well as the combination of all three sections, underpredict FYGPA for females with mean standardized residuals ranging from 0.10 to 0.17. Conversely, male performance is overpredicted with mean standardized residuals ranging from -0.11 to -0.20.
- A similar pattern emerges for HSGPA as well as the combination of HSGPA and SAT, with females being underpredicted (0.07 and 0.09, respectively) and males being overpredicted (-0.08 and -0.10, respectively).
- American Indian, African American, and Hispanic students are overpredicted by all measures and combination of measures. African American students' FYGPA tends to be the most overpredicted with mean standardized residuals ranging from -0.32 to -0.17.

- HSGPA tends to result in the most differential prediction for most racial/ethnic groups while the combination of SAT and HSGPA results in the least differential prediction.
- For best language, the results for the SAT, by section and for the combination of all three sections, show that students whose best language is English are accurately predicted (mean standardized residuals ranging from 0.00 to 0.01), whereas students whose best language is not English are underpredicted by the critical reading and writing sections of the SAT (mean standardized residuals of 0.40 and 0.37, respectively) and are accurately predicted by the mathematics section (mean standardized residual = 0.00).
- For students whose best language is English and another language, the individual SAT sections, as well as the combination of all three sections, overpredict their FYGPAs (mean standardized residuals ranging from -0.09 to -0.02).