
College Board Digital SAT Suite Concordance and Vertical Scaling Studies

The digital SAT launched internationally in March 2023. Rigorous concordance and vertical scaling studies confirm that scores from the digital SAT can be used in the same manner as scores from the paper and pencil SAT.

Introduction

The SAT Suite of Assessments is going digital, making the testing experience easier for educators and students.

College Board is transitioning the SAT® Suite of Assessments to be fully digital by spring 2024. The digital SAT is easier to give, easier to take, more secure, and more relevant for students. The digital SAT is almost an hour shorter, and students can test on their own devices using Bluebook™, College Board's new digital testing app. If students don't have a device, they can request to borrow one from College Board. The digital test is adaptive, so it can continue to measure the same core reading, writing, and math knowledge and skills much more efficiently, shortening the overall length of the test while also allowing students more time, on average, per question.

The SAT will still be scored on a 1600 scale, and scores from the digital SAT can be used in the same manner as scores from the paper and pencil exam for higher education admission, placement, scholarship, and for K–12 accountability purposes. A key goal in developing the digital SAT was to ensure this alignment between scores on the digital and paper and pencil versions of the test so users wouldn't need to use a table to compare scores.

After conducting two comprehensive concordance studies and vertical scaling studies for the digital SAT, we are pleased to confirm a straight-line concordance between our paper-based and digital assessment scores.

This brief details our approach to these studies.

What is concordance and why is it important?

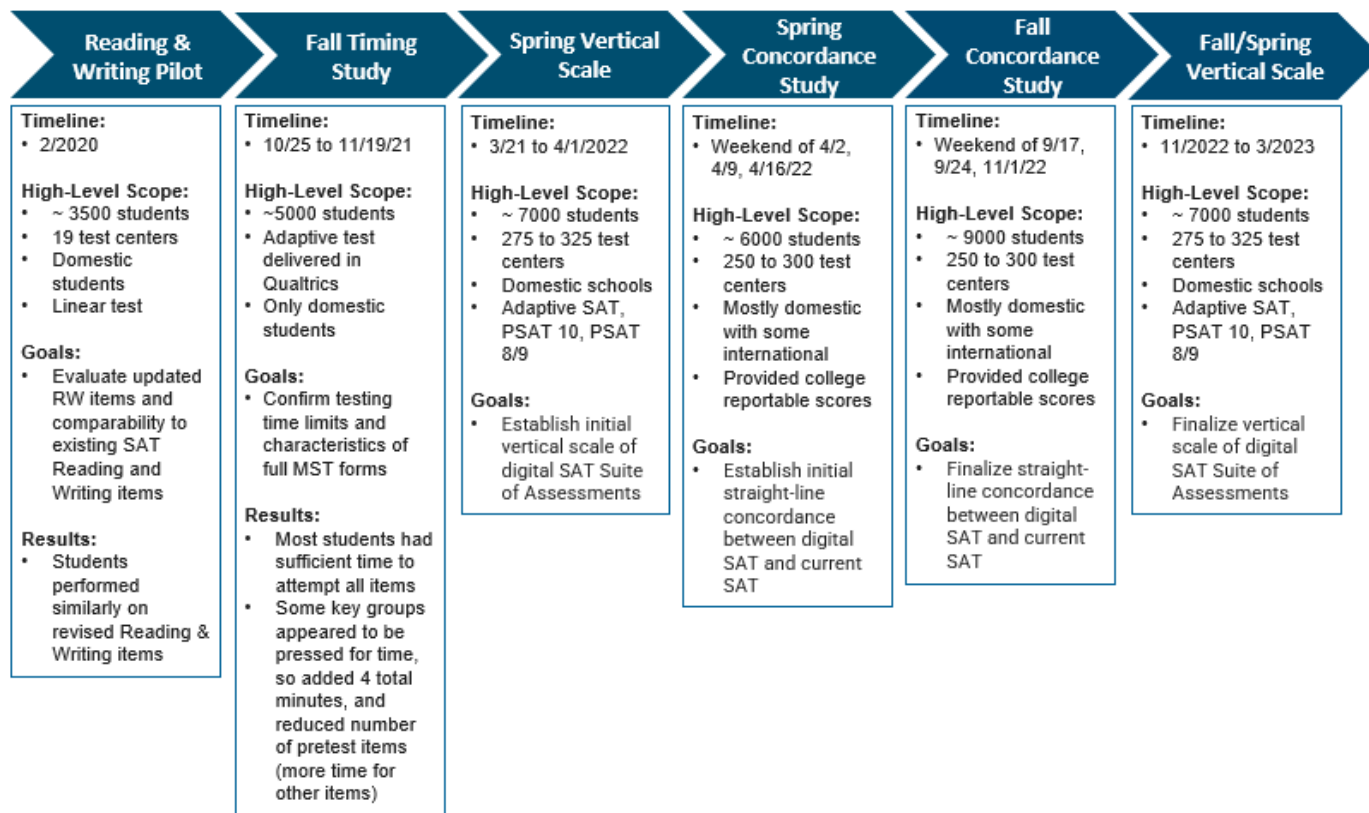
Concordance refers to how well scores from two versions of an assessment measure the same thing or construct. A high degree of concordance implies the two tests measure the same construct and can likely be used similarly, and a low degree implies the opposite. One of the most well-known concordances is between the SAT and ACT (see [here](#)). Additionally, developers of assessments also concord scores when they make notable updates to their tests.

What is vertical scaling and why is it important?

Vertical scaling is a statistical method used to link similar assessments to the same reporting score scale so that students and educators can compare assessments designed for different grade levels and track student progress over time. College Board's vertical scale allows stakeholders to compare scores across the PSAT™ 8/9, PSAT™ 10, PSAT/NMSQT®, and SAT.

Timeline of major studies

College Board has undertaken several studies to support the concordance and vertical scales for the transition to the digital SAT Suite of Assessments.



Concordance sample description

We recruited 21,647 students to take the two assessments. Each student was required to take the digital SAT and a paper version within one month of their digital SAT. The paper-based version was administered during an SAT Weekend or as part of the SAT School Day. Students who could not meet this requirement were removed from the analysis sample.

Table 1 provides a breakdown of the percentage of students across background characteristics. Students who didn't meet our criteria were removed from the original recruitment sample. Despite this attrition, the remaining sample is still large and diverse enough to conduct the analysis accurately. We used statistical methods, such as power analysis, to obtain sample size requirements. For comparison, the concordance study for the last redesign of the SAT in 2016 included 8,677 students in the sample.

Overall, our sample generally represented our intended population of test takers. A few aspects to note:

- The sample was, on average, more able than a typical population of test takers. A few subgroups were overrepresented compared to a full year of test takers.
- Specifically, we noted more females and more students of Asian background participated in the studies than would typically take the SAT.
- We evaluated statistical methods to adjust for these differences. Those methods did not meaningfully alter the results.

Description of Concordance Results

As noted above, our goal was to concord the digital and the current paper-based SAT so that the scores could be used in the same manner by K-12 and higher education. Psychometrically, this means:

- On average, students would score the same on both assessments, and the variability in scores would be the same for both assessments.
- Both assessments would rank order students nearly identically, specifically with a correlation greater than 0.866.
- The statistics that reflect these criteria are the mean (average), standard deviation, and correlation.
- On average, students perform similarly for critical subgroups.

Table 2 includes the results from our studies, which clearly show the mean and standard deviation are essentially equal across the two assessment versions in both test sections. The correlation is very high, and there is only a slight difference in the standard deviation of scores. Because the total score of both assessments is a by-product of the section scores, the fact that they correlate so highly and the means are equal further supports that scores from the digital SAT can be used in the same manner as scores from the paper-based SAT.

Regarding subgroups, the average score for each subgroup tends to be similar across assessment versions in both sections.

Next steps

College Board plans to produce a complete technical report of the concordance study, to be available upon request later in 2023.

Table 1**Concordance Study Sample Size by Subgroup**

| | <i>N</i> | % | | <i>N</i> | % |
|----------------------|----------|-------|---|----------|-------|
| Grade | | | Ethnicity/Race | | |
| 11th Grade | 6,973 | 55.2% | Native American or Alaska Native | 46 | 0.4% |
| 12th Grade | 5,667 | 44.8% | Asian | 2,901 | 23.0% |
| Gender | | | Black or African American | 1,218 | 9.6% |
| M | 5,531 | 43.8% | Hispanic or Latino | 3,081 | 24.4% |
| F | 7,064 | 55.9% | Native Hawaiian or Other Pacific Islander | 10 | 0.1% |
| A | 45 | 0.4% | Two or more races, non-Hispanic | 455 | 3.6% |
| Best Language | | | White | 4,216 | 33.4% |
| English Only | 8,474 | 67.0% | No response | 713 | 5.6% |
| English and Another | 3,504 | 27.7% | Domestic/International | | |
| Another | 421 | 3.3% | Domestic | 11,519 | 91.1% |
| No Response | 241 | 1.9% | International | 1,121 | 8.9% |

Table 2**Average Scores and Variability Across Modes of Administration**

| | Paper-Based Admin. | | Digital-Based Admin. | | Correlation |
|---------------------|--------------------|-----------|----------------------|-----------|-------------|
| | <u>Mean</u> | <u>SD</u> | <u>Mean</u> | <u>SD</u> | |
| Reading and Writing | 602 | 94 | 602 | 94 | 0.885 |
| Math | 600 | 113 | 600 | 113 | 0.919 |
| Total Score | 1202 | 195 | 1202 | 192 | 0.940 |