RESEARCH REPORT 2016-17

CollegeBoard

Using Old and New SAT® Scores for Admission: A Closer Look at Concordant Scores in Predictive Models

RESEARCH

Jessica P. Marini is an associate research scientist in the College Board's research department.

Emily J. Shaw is a senior director in the College Board's research department.

Linda Young is a senior database statistician in the College Board's research department.

About the College Board

The College Board is a mission-driven not-for-profit organization that connects students to college success and opportunity. Founded in 1900, the College Board was created to expand access to higher education. Today, the membership association is made up of over 6,000 of the world's leading educational institutions and is dedicated to promoting excellence and equity in education. Each year, the College Board helps more than seven million students prepare for a successful transition to college through programs and services in college readiness and college success — including the SAT® and the Advanced Placement Program®. The organization also serves the education community through research and advocacy on behalf of students, educators, and schools. For further information, visit collegeboard.org.

RESEARCH

For more information on College Board research and data, visit research.collegeboard.org.

Contents

Tables

Introduction

During the transition period between the use of exclusively old SAT® scores and the use of exclusively new SAT scores, college admission offices will be receiving both types of scores from students. Making an admission decision based on new SAT scores can be challenging at first because institutions have methods, procedures, and models based on the use of old SAT scores. To ease this transition, admission offices can use a concordance table to translate new SAT scores to comparable old SAT scores to use with their existing decision methods. Concordance tables are carefully developed tools that allow test scores from different tests covering similar content to be comparable to each other (Dorans, 2004). For more detailed information on the use of concordance tables and concordant scores, please see Marini, Shaw, Young, and Walker (2016).

The current study continues to establish evidence for the appropriate and sound use of concordant scores in admission decisions. It builds upon recent applied concordance research that examined first-year grade point average (FYGPA) predictions made with native[1] and concordant scores using the old SAT scores and simulated ACT scores (Marini et al., 2016). This study found that comparisons made between predictions using native or concordant scores in native models were highly consistent within student and across student and institutional subgroups.

While the findings from Marini et al. (2016) are useful and provide initial support for the use of concordant scores in native score models, the ACT scores in that study were simulated for each student in the sample using the known relationship between SAT and ACT scores. To make the strongest argument for using and understanding both native and concordant scores in admission decisions, it would have been ideal to have access to native (actual) scores on both assessments studied. The current study design can now account for this and advance the findings of the previous study; it examines two sets of native scores linked in a publically available concordance table. Using data from a pilot study of the redesigned SAT, students in this sample have both native old SAT and native new SAT scores. This makes it possible to make FYGPA prediction comparisons between both native and concordant scores.

This research is important during the transition period between the use of old and new SAT scores in college admission decisions. For the first year or more of the administration of the new SAT, students may submit to college admission offices either old or new SAT scores, depending on when[2] a student took the SAT. Before an admission office can develop a new predictive model using new SAT scores (which is not possible until adequate admission and outcome data are collected), institutions will need to rely on the concordance tables provided by the College Board to transform new SAT scores to old SAT scores to use in preexisting models developed with old SAT scores. Institutions are understandably looking for evidence and reassurance that this practice is sensible and appropriate. The current study was undertaken to explore this issue.

Method

Sample

The data analyzed in this study are from the pilot predictive validity study of the new SAT (see Shaw, Marini, Beard, Shmueli, Young, & Ng, 2016) and include 2,050 students from 15

1. Native refers to an actual or nonconcordant score the student received when taking the test.

2. The new SAT was introduced in March 2016.

four-year institutions (see Table 1). Compared to the 2014 College-Bound Seniors Cohort[3] (the population), this study sample was relatively representative of African American students (13% for both), Hispanic students (17% sample, 18% population), and white students (46% sample, 49% population), but had more Asian students (20% sample, 12% population) and female students (64% sample, 53% population) than the College-Bound Seniors 2014 cohort (The College Board, 2014). Please see Shaw et al. (2016) for in-depth information regarding sample selection and data cleaning procedures.

Table 1.

Characteristics of Sample

Student Characteristics (*n* = 2,050)		
Gender	Female	64
	Male	36
Race/Ethnicity	American Indian or Alaska Native	<1
	Asian, Asian American, or Pacific Islander	20
	Black or African American	13
	Hispanic	17
	White	46
	Other	3
	No Response	<1
Institutional Characteristic (*n* = 15)		
Control	Private	33
	Public	67
Admittance Rate	Under 50%	40
	50% to 75%	40
	Over 75%	20
Undergraduate Enrollment	Small	0
	Medium	33
	Large	13
	Very Large	53

Measures

Native New SAT Scores. New SAT scores were obtained for each student in the study sample in a special administration of a pilot form of the new SAT in the fall of 2014. These new SAT scores include two section scores, three test scores, two cross-test scores, and seven subtest scores. For this study, we were interested in the following native scores:

- Two section scores (200 to 800 scale) — Evidence-Based Reading and Writing (ERW); Math (MS).
- Three test scores (10 to 40 scale) — Reading (R); Writing and Language (WRLA); and Math (MT).

3. College-bound students in the class of 2014 who took the SAT or SAT Subject Tests™ at any time during high school.

Native Old SAT Scores. The most recent old SAT scores were obtained from the College Board for each student in this sample. The old SAT includes three sections, Critical Reading, Mathematics, and Writing, and the score scale range for each section is 200 to 800.

Concordant Old SAT Scores. Concordant old SAT scores were mathematically arrived at for each student in the sample using the student's native new SAT score in conjunction with the concordance table linking new SAT scores to old SAT scores. Note that this concordant old SAT score is not an actual score that a student earned but an estimate of a comparable old SAT score based on a student's performance on the new SAT. Concordant score pairs were as follows (native to concordant):

- New SAT math section (MS) to old SAT mathematics section (M)
- New SAT reading test (R) to old SAT critical reading section (CR)
- New SAT writing and language test (WRLA) to old SAT writing section (W)

High School GPA. Self-reported high school grade point average (HSGPA) was obtained from the SAT Questionnaire when students had taken the old SAT and is constructed on a 12-point interval scale, ranging from 0.00 (F) to 4.33 (A+).

First-Year GPA. Each participating institution supplied first-year grade point average (FYGPA) values for the students included in this sample.

Analyses and Results

All students in the sample had to have native old SAT scores and native new SAT scores in the study file. Concordant old scores were calculated for native new SAT scores using the concordance table linking old and new SAT scores developed by the College Board. In order to begin prediction analyses, each student had to have three types of scores — native old SAT, native new SAT, and concordant old SAT scores (new SAT scores changed to old SAT scores).

There were two main comparisons of interest. The first set of analyses was designed to compare whether there were meaningful differences in how the native and concordant old SAT scores predicted a student's FYGPA. The second set of analyses compared the prediction accuracy between old SAT and new SAT score predictions of FYGPA, as well as concordant and native score predictions with a student's actual FYGPA.

Research Question One: If an institution has preexisting prediction model built with old SAT scores and receives new SAT scores which are concorded to old scores to place in the model, are the FYGPA predictions highly similar to those made with native old SAT scores in that same model?

During the transition period covering the simultaneous use of old and new SAT scores in admission, institutions with prediction models built using old SAT scores need to rely on the concordance tables (translating new SAT scores to old SAT scores) until their decision models can be analyzed and updated to solely use new SAT scores.

Imagine two hypothetical institutions, A and B, each with an existing prediction model that uses old SAT scores and HSGPA to predict how well a student will perform during the first year of college. Institution A uses all three sections of the old SAT (CR, M, and W), as well as HSGPA to predict FYGPA. Institution B uses only two sections of the old SAT (CR and M), as well as HSGPA, to predicted FYGPA. These institutions begin receiving new SAT scores

from applicants, but cannot use these scores in their existing models without concording them to old SAT scores. Both institutions use the concordance tables and translate the new SAT scores they received into concordant old SAT scores. However, they want to know how "good" the predictions of FYGPA are using the concordant old SAT scores in their models. Are they as good or accurate as they would have been if the student had taken/submitted native old SAT scores?

To answer this question, two regression prediction models were created for each institution in the study using native old SAT scores — one using the Critical Reading section, Mathematics section, Writing section, and HSGPA (three-section model) and the other using the Critical Reading section, Mathematics section, and HSGPA (two-section model). Each student's predicted FYGPA produced by each model was saved for further analysis. Then, concordant old SAT scores (new SAT scores concorded to old SAT scores) were plugged into the three-section model and then the two-section model to produce two additional predicted FYGPAs for each student.

Ultimately, each student had four predicted FYGPAs: a native and concordant score prediction from the three-section model and a native and concordant score prediction from the two-section model. Then, the native old and concordant old score prediction within a given model was compared for each student using the threshold of being within ±0.165. As in Marini et al. (2016), the threshold of ±0.165 was used to establish predictions that were highly similar. As explained in the previous study, this threshold was chosen as it is half of 0.33, the grade change amount to place a student into a different letter grade category (e.g., typically a grade of A = 4.00, while a grade of A– = 3.67) (Marini et al., 2016). The results of the comparisons mentioned above are listed in Table 2.

Table 2.

Comparison of Predicted FYGPAs for Pairs of Native and Concordant Scores for a Given Student

	Highly Similar Estimate (within ± 0.165)	Over-estimate	Under-estimate
Old SAT three-section model (CR, M, W, HSPGA)	76%	14%	10%
Old SAT two-section model (CR, M, HSGPA)	85%	8%	7%

Note. All comparisons were computed by subtracting the predicted FYGPA using the concordant scores from the predicted FYGPA using the native scores. An overestimate indicates the predicted FYGPA using concordant scores was larger than the predicted FYGPA using native scores. An underestimate indicates the predicted FYGPA using concordant scores was smaller than the predicted FYGPA using native scores.

As the table shows, native and concordant scores produce highly similar estimates of FYGPA under the three-section model (76%) and two-section model (85%). The process of linking or concording scores introduces error to prediction just by its very nature so that one would not expect 100% of students to fall within the highly similar category. Also, the three-section model has three scores with some "concordance error" in the regression equation, whereas the two-section model has one fewer. This is likely accounting for the higher percentage of students with highly similar predicted FYGPAs when using the two-section model. It is also possible that the two math sections/tests and two reading sections/tests are more similar to each other than the two writing sections/tests are to each other. In addition, the percentage of students with an over- or underestimated predicted FYGPA is relatively similar within a model.

Research Question Two: Does using either native old SAT scores or concordant old SAT scores (based on new SAT scores) more accurately predict a student's actual FYGPA?

Once the predicted FYGPAs were compared to each other, it was important, based on previous findings in this study and Marini et al. (2016), to see if native and concordant scores produced similarly accurate predictions of FYGPA, as would be expected. To do this, the differences between each predicted FYGPA and the student's actual FYGPA were examined to see how similar or different they were from each other. The predicted FYGPAs from research question one (the three-section model with native and concordant scores and the two-section model with native and concordant scores) were used, and three new models using native new SAT scores (two-section model; three-test-score model; and two-test-score model) were created. These three models using new SAT scores were created to correspond to the three-section and two-section old SAT models to allow for comparions between predicted FYGPAs from native and concordant old SAT scores. Studying predicted FYGPAs from these new SAT score models allows for greater context regarding the classification values for native and concordant scores. Also, for comparison purposes, a model predicting FYGPA from only HSGPA by institution was created.

Ultimately, each student had eight predicted FYGPAs: four from old SAT models varying native or concordant scores and three or two test sections, three from new SAT models all with native scores, and one from the HSGPA-only model. Once again, the threshold of ±0.165 was used to signify a highly similar estimate. The results of these comparisons are shown in Table 3.

Table 3.

Comparison of Predicted FYGPAs to Actual FYGPA for a Given Student

		Highly Similar Estimate (within ± 0.165)	Over-estimate	Under-estimate
No SAT	HSGPA-only model	24%	31%	45%
Old SAT	Three-section model, native scores (CR, M, W, HSPGA)	28%	29%	43%
	Three-section model, concordant scores (CR, M, W, HSPGA)	27%	27%	47%
	Two-section model, native scores (CR, M, HSGPA)	28%	29%	43%
	Two-section model, concordant scores (CR, M, HSGPA)	26%	27%	46%
New SAT	Section-level model (ERW, MS, HSGPA)	27%	30%	43%
	Three-test model (MT, R, WRLA, HSGPA)	27%	30%	43%
	Two-test model (MT, R, HSGPA)	27%	30%	43%

Note. All comparisons were computed by subtracting the predicted FYGPA from the actual FYGPA. An overestimate indicates the predicted FYGPA was larger than the actual FYGPA. An underestimate indicates the predicted FYGPA was smaller than the actual FYGPA. Also, rows might not add to 100% due to rounding.

There are greater differences between actual FYGPA and predicted FYGPA for models using concordant scores than for models using native scores. For example, you can review the difference between actual and predicted FYGPA for the three-section model. For native scores, 28% of students had a predicted FYGPA that was highly similar to their actual FYGPA, whereas 27% of students had a predicted FYGPA that was highly similar when concordant scores were used. The same pattern is evident for the old SAT two-section model. When looking at models created with native new SAT scores, the rate of highly similar estimates is similar to the old SAT models. Further, for comparison purposes, predicted FYGPAs were

modeled using only HSGPA. In the HSGPA-only model, 24% of the predicted FYGPAs were highly similar to the actual FYGPA. This indicates that although using concordant scores in a native model does not produce identical results to using native scores, it is better to use the concordant scores in the native model as opposed to no scores at all.

Discussion

The purpose of this study was to add to the collective understanding of how well a student's native (actual) admission test scores (received by an institution) predict their FYGPA when compared to the use of concordant scores for that same student. This study focused on the concordance between old and new SAT scores to explore the impact of concording new SAT scores to old SAT scores and using that concordant old SAT score in pre-existing admission models. Two types of analyses were performed.

The first set of analyses examined differences in the predicted FYGPA by student when using a native old SAT score versus a concordant old SAT score. The second set of analyses compared predicted FYGPAs to actual FYGPAs for native and concordant old SAT scores, new SAT scores, and HSGPA alone. These analyses showed that although the process of concordance adds some error into the prediction model, this error is not substantial. This practice is also an improvement over excluding concordant scores in the prediction model. Predictions are more accurate when concordant scores are used than when they are excluded. Although concordant score predictions were not identical to native score predictions, they were very close.

A practical message from these findings is that institutions can feel comfortable using the SAT concordance tables for their predictive models during the transition period. This is the interval in which they receive both old and new SAT scores, prior to having the outcome data to build new models based on the new scores. Taking a new SAT score submitted by a student and using the concordance table to find the corresponding old SAT score and "plugging" this concordant old SAT score into an existing model does not disadvantage students when predicting their FYGPA. Further, results showed that using concordant scores in the native score models produced more accurate results in the prediction of FYGPA than when omitting the concordant scores from the model altogether.

References

Dorans, N. J. (2004). Equating, concordance, and expectation. *Applied Psychological Measurement, 28* (4), 227–246.

Marini, J. P., Shaw, E. J., Young, L, & Walker, M. (2016). *How Does the Use of Concordant Scores Impact Admission Decisions?* (College Board Research Report No. 2016-3). New York: The College Board.

Shaw, E. J., Marini, J. P., Beard, J., Shmueli, D., Young, L., & Ng, H. (2016). *The Redesigned SAT Pilot Predictive Validity Study: A First Look* (College Board Research Report No. 2016-1). New York: The College Board.

The College Board. (2014). *2014 College-bound seniors: Total group profile report. New York: The College Board*. Retrieved from: https://secure-media.collegeboard.org/digitalServices/pdf/sat/TotalGroup-2014.pdf.

The Research department actively supports the College Board's mission by:

- Providing data-based solutions to important educational problems and questions
- Applying scientific procedures and research to inform our work
- Designing and evaluating improvements to current assessments and developing new assessments as well as educational tools to ensure the highest technical standards
- Analyzing and resolving critical issues for all programs, including AP®, SAT®, PSAT/NMSQT®
- Publishing findings and presenting our work at key scientific and education conferences
- Generating new knowledge and forward-thinking ideas with a highly trained and credentialed staff

Our work focuses on the following areas

Admission	Measurement
Alignment	Research
Evaluation	Trends
Fairness	Validity

Follow us online: research.collegeboard.org

CollegeBoard

CollegeBoard

00402-005