NBER WORKING PAPER SERIES

STANDARDIZED TEST SCORES AND ACADEMIC PERFORMANCE AT A PUBLIC UNIVERSITY SYSTEM

Theodore J. Joyce
Mina Afrouzi Khosroshahi
Sarah Truelsch
Kerstin Gentsch
Kyle Du

Working Paper 34975
http://www.nber.org/papers/w34975

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2026

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w34975

Standardized Test Scores and Academic Performance at a Public University System
Theodore J. Joyce, Mina Afrouzi Khosroshahi, Sarah Truelsch, Kerstin Gentsch, and Kyle Du
NBER Working Paper No. 34975
March 2026
JEL No. J13

ABSTRACT

Recent studies of Ivy-Plus institutions suggest that standardized test scores (SAT/ACT) are far better predictors of college success than high school grade point average (HS-GPA), prompting a return to the requirement that test scores be submitted for admission at elite colleges. We ask whether re-establishing the SAT requirement for admission at a large urban public university system would improve the predictability of academic outcomes. Using administrative data for the 2010-2019 first-year cohorts, we update earlier work of students from public universities as to the relative predictive power of HSGPA and SAT scores on first-year outcomes and graduation rates. Contrary to findings at elite private institutions, we find that HSGPA is the dominant predictor of academic success in this public system. A one-standard-deviation increase in HSGPA is four to six times more predictive of six-year graduation than a comparable increase in SAT scores. Out-of-sample forecasts for the post-COVID period (2020–2024) confirm that test-optional models relying only on HSGPA experience relatively little loss in predictive accuracy compared to models that include test scores. We conclude that HSGPA remains the most reliable signal of degree completion at broad-access public universities.

Theodore J. Joyce
City University of New York
Department of Economics and Finance
and NBER
theodore.joyce@baruch.cuny.edu

Mina Afrouzi Khosroshahi
City University of New Yorrk
mafrouzikhosroshahi@gradcenter.cuny.edu

Sarah Truelsch
Office of Applied Research, Evaluationand
Data Analytics
Sarah.Truelsch@cuny.edu

Kerstin Gentsch
Office of Applied Research, Evaluation
and Data Analytics
kerstin.gentsch@cuny.edu

Kyle Du
Office of Applied Research, Evaluation
and Data Analytics
kyle.du@cuny.edu

1 Introduction

During the COVID-19 pandemic, most colleges did not require the SAT/ACT for admission. Many colleges continued the policy after COVID or made submission of test scores optional (Avery et al., 2025). In 2024 several elite colleges re-instituted the SAT/ACT requirement based on empirical evidence that these standardized tests were a superior predictor of first-year academic outcomes than high school grade point average (HSGPA) (Cascio et al., 2024; Friedman et al., 2025a; Sacerdote et al., 2025).

The re-institution of SAT/ACT (hereafter SAT) for college admissions received widespread coverage in the popular press and re-ignited the long-standing debate over the role of standardized tests in college admissions (Deming, 2024; Leonhardt, 2024; The Economist, 2025; Lemann, 1999; Soares, 2011; Bowen et al., 2009; Kurlaender et al., 2020). What made the renewed debate novel was that the requirement of SATs for admission, instead of being optional, increased the likelihood that disadvantaged students would be accepted to an elite college (Sacerdote et al., forthcoming). In addition, the SAT was not biased against disadvantaged students because it did not underpredict their academic success at elite colleges (Friedman et al., 2025a).

In this study we analyze the relative predictive power of the SAT for student academic success at the eleven senior colleges in a large urban public university system (hereafter, the System). The System provides a useful contrast to the experience of Ivy-plus colleges analyzed in recent studies. Based on Barron's guide to college selectivity, the colleges in the System vary from very competitive to non-competitive and are comparable in selectivity to a wide range of public colleges in the US. [1] Like many peer institutions, these colleges required SATs prior to COVID, dropped the requirement for the entering cohorts of 2021-22 followed by a test-optional policy thereafter.

Much of the literature relating SATs to student success has focused on first-year academic outcomes. This is understandable. First-year metrics provide a timely assessment of recent changes, especially at elite colleges, as they correlate strongly with later outcomes. But graduation is a more important marker of college success at less elite colleges and is worrisomely low (Bowen et al., 2009). The six-year graduation rate at all public universities in the US was 58.5 percent for the 2017 freshman cohort, but 96 percent at the eight Ivy-League colleges. [2]

To assess the relative contribution of SATs to predicting student academic success, we follow commonly used approaches that compare SATs to other predictors of academic outcomes. Data pertain to all first-year fall cohorts from 2010-2024 at the System's eleven senior colleges. In the first part of the analysis, we estimate the effect of a one-standard deviation

[1] Barrons guide ranges in 2011 ranged from Highly Selective which includes schools that are Most Competitive or Highly Competitive, Very Competitive, Competitive and Less Selective which includes schools that are Less Competitive or Non-competitive. Ninety percent of all public four-year colleges and universities ranged from Very Competitive to Less Selective. Two of the System's 11 colleges were considered Very Competitive, 4 Competitive, 1 Less Competitive and 3 non-competitive (Barron's Educational Series, 2011; see also StayInformed_DataMethodologyFall2021.pdf for a more updated description).

[2] Data extract from IPEDS online Trend generator, https://nces.ed.gov/ipeds/trendgenerator.

increase in HSGPA and SAT on academic outcomes from 2010-2019. We then estimate separate models for under-represented minorities and Pell grant recipients and test whether SATs or HSGPAs evidence calibration bias as predictors of academic outcomes across groups (Friedman et al., 2025a). Third, we compare the predictive power of SATs to that of a statewide exam in English Language Arts (ELA) administered to high school students primarily in the 11th grade. The advantage of SATs is that they provide a common metric across students. A criticism is that SATs do not measure student learning in high school (Kurlaender & Cohen, 2019). Comparing the predictive power of SATs to the statewide ELA exam addresses this issue.

In the second part of the analysis we use data from pre-COVID cohorts to "train" a model of student outcomes that includes both SATs and HSGPA. We use the model to predict the outcomes of students just before and during COVID. We then repeat the procedure for the post-COVID years when SATs were optional using only HSGPA to predict out-of-sample outcomes. The exercise simulates the task of admission staff when SATs are unavailable and quantifies the change in prediction errors when only HSGPA is used in admissions.

Methodologically, we take advantage of several procedures that have not been used in this literature. We apply the Gelbach decomposition which uses the omitted variable formula in linear regression to apportion the change in the coefficient on SAT scores, for example, to the inclusion of HSGPA or demographic characteristics of the student. With binary outcomes, we use the receiver operating characteristic (ROC) curve to assess the incremental effect of excluding the SAT on a model's sensitivity and specificity. We apply the Shapley decomposition to estimate the marginal increase in R^2 that is less sensitive to the order in which variables are added. We simulate the task of admission staff when SATs are not used to make out-of-sample forecasts of student academic success.

Our results underscore the dominance of HSGPA as the most important predictor of student success at this public university system. The effect of a one-standard deviation increase in HSGPA is four to six times larger than the effect of the same increase in SATs on six-year graduation rates. We find that the predictive power of HSGPA is remarkably robust to the inclusion of SAT scores, state-wide exams in ELA, and demographic characteristics of students. Exclusive use of HSGPA to forecast student academic success does not increase prediction errors when compared to models with SAT scores. Our results are remarkably close to those of Bowen et al. (2009) and Allensworth & Clark (2020) who made the same comparison across different samples of students at public colleges and universities.

2 Literature

The history of the SAT is elegantly described in Lemann's book, *The Big Test* (Lemann, 1999). Originally designed as a test of aptitude, the SAT has evolved in response to truth-in-testing legislation, large racial and ethnic differences in performance, the development of test-prep firms, and the need to more effectively assess college-readiness skills. Despite major revisions in 2006 and 2017, scores on the SAT continue to vary substantially by race/ethnicity and socioeconomic status. In response to these persistent concerns, a test-optional movement grew (Soares, 2011). By 2019, the percentage of 400 plus colleges tracked by the Compass Education Group that required the SAT stood at only 64 percent. The test-optional movement was supercharged by the disruptions of the COVID pandemic, which lead to canceled test dates and made it impossible or extremely burdensome for students to take the exams. Many colleges suspended their testing requirements during the COVID disruption. By 2023, only 4 percent of these same schools tracked by the Compass Education Group required standardized test scores. [3]

The move back to SATs from a test-optional policy began among elite colleges in 2023. In that year a high-profile study of students in the Ivy-plus colleges [4] showed that SATs were a much stronger predictor than HSGPA of several elite outcomes: reaching the top 1% of the income distribution, admission to top graduate schools and employment in prestigious firms by age 33 (Chetty et al., 2023, forthcoming). Using a similar sample of students from Ivy-plus colleges, researchers reported that SATs, as compared to HSGPA, were a much stronger predictor of first-year GPA than HSGPA for both advantaged and disadvantaged students (Friedman et al., 2025a). In addition, SAT scores did not underpredict the performance of students from disadvantaged high schools or under-represented minorities (URM) as would occur if SATs were biased against these groups. A third study by some of the same researchers showed that Dartmouth's test-optional policy diminished the likelihood that high-achieving students from disadvantaged backgrounds would be admitted (Sacerdote et al., 2025). The finding turned a major criticism of standardized testing on its head.

Unlike the elite colleges, the University of California, for years a leader in higher public education, went in the other direction. After a prolonged debate over the role of SATs in its admission decisions, the President of the University of California empaneled the Standardized Testing Task Force (STTF) to assess whether the University should continue to use the SAT for admissions or modify its use, replace it with another standardized test, or eliminate it altogether. [5] The STTF's 200-plus page report recommended the University continue to use SATs in its admission process as part of a holistic review that "...compensate(s) for group differences in test scores" (STTF, p. 5). Using UC system data from 2010-2012, the Report's authors demonstrated that SATs were a stronger predictor of freshman GPAs than HSGPAs, but equivalent to predicting retention to the second year and graduation. This was true within demographic and socioeconomic groups.

The report garnered critical comment from leading academics (Kurlaender et al., 2020). Critics argued that SATs were a manifestation of society's inequities that diminished the proportion of students from underrepresented groups admitted to the UC system. They argued that a standard test more closely aligned with high school curriculum and college readiness should be considered in place of the SAT. They presented evidence that a well-regarded test of college readiness was as predictive of student success as the SAT (Kurlaender & Cohen, 2019). At the meeting of the Regents of the University of California in June of 2020, the President of the University recommended that the University suspend the requirement

[3] See Deming (2024). `https://forklightning.substack.com/p/should-selective-college-admissions` (last accessed November 13, 2025). For a more complete list see `https://www.fairtest.org/sites/default/files/Optional-Growth-Chronology.pdf` (last accessed October 28, 2025).

[4] Ivy-plus includes the 8 Ivy league colleges plus Duke, MIT, Stanford and the University of Chicago)

[5] `https://senate.universityofcalifornia.edu/_files/committees/sttf/sttf-report.pdf`

that SAT be submitted for undergraduate admissions through 2024. In 2025 the University of California dropped any use of the SAT in admissions.

Arguments against requiring SATs for admissions are that they favor more advantaged students who can afford private test tutors or test preparation courses and are more likely to take the SAT multiple times (Rothstein, 2004; Alon & Tienda, 2007; Goodman et al., 2020). The Dartmouth study clearly challenged this argument by analyzing actual admissions decisions. The study revealed that admissions officers viewed applicants' scores within the context of their circumstances and did not necessarily penalize disadvantaged applicants for having lower scores so long as those scores demonstrated ability to do rigorous work. Under Dartmouth's test-optional policy fewer disadvantaged students reported scores that would have greatly increased their probability of admission (Sacerdote et al., 2025).

The study by Friedman et al. (2025a) also challenged the assumption that SATs were biased against less advantaged applicants. If, for example, under-represented minorities (URM) performed better than non-URM with the same SAT score, then SATs would be a biased signal of URM's ability to succeed. This could occur if non-URM's access to test-preparation tutors and multiple test-taking opportunities inflated their scores compared to those of URMs of same academic potential. Friedman et al. (2025a) found the opposite. Using a larger sample of students from Ivy-plus colleges, URMs had lower first-year GPAs than non-URM at almost every level of SAT, although the test remained a strong predictor of each group's academic success.

The predictive power of SATs for student success looked dramatically different at public colleges and universities than at elite colleges. The seminal study by Bowen et al. (2009) compared the predictive power of HSGPA versus SATs on six-year graduation rates among public universities. The sample consisted of 150,000 first-year students from 21 flagship public universities and 47 state colleges from four states states Maryland, North Carolina, Ohio, and Virginia in 1999. They divided the schools into five clusters based on selectivity. They used standardized regression coefficients to compare the effect of a one-standard deviation increase in either HSGPA or SATs on the probability that a student graduated in six years. The authors found that HSGPA was 2.5 times more predictive than the SAT of six-year graduation rates at the most selective public universities and up to 10 times more predictive at the least selective schools.

Allensworth & Clark (2020) reached the same conclusion. They followed over 17,000 high school graduates from the Chicago Public Schools who enrolled in four-year colleges between 2006-2009. The average six-year graduation rate of the colleges attended was 47 percent. They found that the effect of a one-standard deviation increase in HSGPA was roughly six times larger than the same change in the ACT scores when each was included separately in the regression. The effect of ACT scores on graduation became negative when included with HSGPA and a random effects control for high school.

The similarity of the results in Bowen et al. (2009) and Allensworth & Clark (2020) underscore the consistency of HSGPA as the strongest predictor of college completion at public universities. The sample by Bowen et al. (2009) is unique in its breadth of colleges whereas Allensworth & Clark (2020)'s use of Chicago public high school graduates is similar to those in our study. Despite the differences in samples and cohorts, their results are remarkably similar. Lastly, Bowen et al. (2009) and Allensworth & Clark (2020) focus on six-year graduation rates. Many studies that compare the predictive power of HSGPA and SATs focus on first-year GPA, credits earned and first-year retention (Kobrin et al., 2008; Mattern & Patterson, 2014; Westrick et al., 2019; Soares, 2011). As Bowen et al. (2009) emphasize, a college degree is the relevant goal of incoming students.

Why does the predictability of HSGPA and SAT for academic success differ so much between elite and non-elite colleges? As noted by previous researchers, HSGPA suffers from range restriction at elite colleges, which attenuates its correlation to outcomes (Rothstein, 2004). This is particularly acute at Ivy-plus colleges in which the variation in HSGPAs among applicants is limited (Friedman et al., 2025a). The outcomes of interest may also be different. Elite colleges already boast extremely high graduation rates, which do not present much variation in outcomes. As a result, recent research on the Ivy-plus colleges have employed more selective outcomes, such as position in the top 1% of earners. But for non-elite colleges, graduation is not a given. It is both a real concern for college admission officers, and presents enough variation for meaningful analysis.

For applicants to public colleges, the HSGPA captures a set of skills that are arguably more relevant for college graduation than the academic skills measured by standardized tests. In an important study of high school seniors, authors found that a composite measure of self-regulation was the strongest predictor of both HSGPA and college graduation (Galla et al., 2019). They write, "*Our findings suggest that report card grades provide information about self-regulation not captured by admissions test scores. In particular grades assigned to students by their high school teachers reveal quite a bit about their capacity to resist momentary temptations, regulate emotions, and sustain effort across days, months and years in pursuit of important goals*" (Galla et al., 2019, p. 2103). These are the kinds of skills required to persist in the pursuit of a college degree, especially for students facing more obstacles and enjoying fewer resources than those at elite colleges.

A common concern with high school grades is that they are not comparable across high schools that vary in quality and standards. Admission staff may have difficulty assessing the comparability of high school grades among applicants from school for which they historically receive few applicants. Thus, adjusting for differences in high-school quality is important for assessing the predictive power of HSGPA and SATs. When Bowen et al. (2009) and Allensworth & Clark (2020) included high school fixed or random effects in their regression models, the effect of HSGPA increased while the effect of SAT turned negative. In an analysis of first-semester GPAs at four Texas public universities, the inclusion of high school fixed effects doubled the magnitude of the coefficient on high school class rank and eliminated or reversed black-white differences (Fletcher & Tienda, 2010). Even in the recent study of the Ivy-plus colleges, the inclusion of high school fixed effects increased the coefficient on HSGPA in a regression of first-year GPA by over 50 percent (Friedman et al., 2025a, Table 1).

Our study advances the literature in several ways. We provide an important update on the work by Bowen et al. (2009) and Allensworth & Clark (2020) using more recent data that reflect major revisions to the SAT. As with both studies, we assess the relative importance of

SATs for a large public university system across campuses of varied selectivity. Methodologically, we apply the Gelbach and Shapley decompositions as new ways to assess the relative importance of HSGPA and SATs. We follow Friedman et al. (2025a) and test for calibration bias with the SAT and HSGPA. Finally, we also cover the period pre- and post-COVID. Many schools stopped requiring or made optional the SAT for admissions in wake of COVID. We simulate the challenge facing admission staff by estimating the out-of-sample prediction errors when SATs are not used in a model.

3 Methods

3.1 Data

We use enrollment data from the fall cohorts of first-time students from 2010 to 2024 at all 11 of the System's four-year colleges. Data are linked to the National Student Clearinghouse (NSC) to capture retention and graduation of those who leave the System after enrollment. We analyze first-year grade point average (GPA) and first-year retention. If students drop out after the first semester, we use their first-semester GPA as a proxy for their cumulative first-year GPA to prevent sample attrition. First-year retention is a dichotomous indicator of whether a student enrolled for classes in the fall of their second year whether at a campus in the System or a college external to the System. We also evaluate four- and six-year graduation rates. Again we use data from the NSC to follow students who leave the System and graduate from another college.

The key independent variables are high school grade point average (HSGPA), total SAT scores, and a statewide exam in English Language Arts (ELA). Both the HSGPA and the ELA exam are measured on a scale of 0 to 100. The University System used only college preparatory courses taken in high school to determine an applicant's HSGPA from 2010 to 2018. Since 2019 the System has used grades in all high school classes to determine HSGPA. The total SAT and ACT scores are expressed on the same scale (400 to 1600) via a standard crosswalk from ACT to SAT. As noted, we use SAT to include either the SAT or ACT. The SAT underwent a major redesign that was deployed in March of 2016. The College Board published a conversion table so that scores from previous years could be compared to scores from the re-designed SAT. We use the conversion data to link the total SAT scores from 2010-2016 to the redesigned series. [6] The third key predictor of student success is a statewide ELA exam. The ELA exam is designed to measure student grasp of the common core material in English, Language and Arts in grades 11 and 12. The ELA exam is administered three times a year and most students take it in their junior year. We view the ELA exam as a possible substitute for the SAT because unlike the SAT, the ELA exam assesses students' learning in high school and provides a common metric by which to evaluate within-state applicants (Kurlaender & Cohen, 2019; Kurlaender et al., 2020; Deming, 2024).

[6] The System converted scores from the re-designed SAT to the old SAT in 2017 and 2018. We regressed the re-designed scores on the old scores for those two years and used the coefficients from the model to harmonize the scores from 2010-2016 to the redesigned scores from 2017.

3.2 Statistical Analysis

We divide the empirical analysis into two parts. In the first we use data from the 2010-2019 fall cohorts to assess the relative importance of HSGPA and SAT in predicting student outcomes with and without controls for student characteristics. In the second part, we simulate the task of an admission office by predicting first-year outcomes for the 2019-2024 incoming cohort using only high school grades and test scores to evaluate applicants.

3.2.1 Part I

We follow the general literature and use linear regression to estimate the relative contribution of HSGPA and SATs to academic outcomes (Friedman et al., 2025a):

$$Y_i = \beta_1 \text{HSGPA}_i + \beta_2 \text{SAT}_i + \mu_s + \tau_t + \lambda_h + \omega X_i + e_i \tag{1}$$

Let Y be the academic outcome for student (i); HSGPA and SAT are the high school GPA and SAT scores for each student. We normalize both HSGPA and SAT so that β_1 and β_2 show the effect of a one standard deviation increase on Y. Let μ_s and τ_t be college (s) and year (t) fixed effects and λ_h be high school (h) fixed effects. Let X be a vector of student characteristics such age, gender, whether the student is from underrepresented minority (URM), and whether the student received the maximum or partial Pell grant. [7]

We estimate equation (1) separately for URM and non-URM as well as for students who receive the maximum Pell versus some Pell or none. In some specifications we use non-parametric representations of SAT and HSGPA following Friedman et al. (2025a).

A common practice in this literature is to show how the coefficient on SAT changes, for example, when HSGPA or other variables are added to the model. However, the order in which variables are added can affect their relative contribution especially if they are correlated. The Gelbach decomposition uses the omitted variable formula to apportion the change in a coefficient to the inclusion of other variables that are invariant to the order in which they are added (Gelbach, 2016). Specifically, we estimate equation (1) with only SAT. Refer the coefficient as β_{base}. We then add the rest of regressors in equation (1) and apportion the change in the coefficients on SAT between HSGPA, the other variables, and fixed effects.

Another common approach to assessing the relative contribution of a regressor is to display the marginal increase in the R-squared (R^2) as new variables are added to the model. As is well-known, the order in which the variables are added affects their marginal contribution to R^2. To overcome this limitation, we apply the Shapley decomposition. The approach runs separate regressions based on every permutation of variable ordering. A variable's contribution to R^2 is the average of all its marginal contributions across all permutations.

[7] We use the maximum Pell grant as a proxy for a low-income household. We use Pell in place of Adjusted Gross Income (AGI) because the latter is only available after 2014. However, for the years in which we have both, substituting AGI for Pell had little impact on the coefficients of SAT or HSGPA, the primary focus of the analysis.

3.2.2 Part II

In this part we estimate equation (1) for cohorts just before and after COVID. We use the model to predict the subsequent years' outcomes and compare the actual to the predicted values. We estimate equation (1) but exclude all demographic characteristics in an effort to simulate the information admissible for use by an admission staff. We estimate models with and without SATs for the years in which SATs were required (2017-2020). For years in which the SAT was optional (2021-2024), we use only HSGPA along with year, college and high school fixed effects to predict first-year outcomes. We eschew use of the SAT because less than 7 percent of applicants submitted SATs during the test optional period.

We use the mean absolute error and the rank sign test to compare predicted to actual values for continuous outcomes and the Receiver Operating Characteristic (ROC) curve for binary outcomes. For example, we use a probit model with both HSGPA and SAT to explain first-year retention for the years 2017-2018. We use the coefficients from model to predict first-year retention in 2019. We use the ROC curve to evaluate the out-of-sample sensitivity and specificity in 2019. We repeat the procedure but exclude SAT from the "training" model in 2017-2018 and use the ROC curves to assess the change in sensitivity and specificity from excluding the SAT. [8]

We repeat the exercise but use the 2017-2019 cohorts to predict 2020 outcomes. The fall 2020 cohort is the first to experience a complete year under COVID and the last one in which SAT were required for admissions. We use the 2020 cohorts to predict 2021, the 2020-2021 cohorts to predict 2022 and so on through 2024. We compare pre- and post-COVID prediction errors with and without SATs in the model.

4 Results

Table 1 displays key summary statistics stratified by the pre- and post-COVID cohorts. A number of changes pre and post COVID are noteworthy. The proportion of underrepresented minorities increases from 42 percent among the 2010-2019 freshman cohorts to 50 percent from 2020-2024. Average high school GPAs (HSGPA) increases by 2 points on a scale of 0 to 100 while the proportion of students that receive the maximum Pell award or a lesser

[8] A ROC displays the tradeoff between the sensitivity of the prediction against the false positive rate (1-specificity). Specifically, the ROC curve takes each predicted value of an outcome, such as first-year retention, and compares it against a designated threshold. Every student with a predicted probability of retention greater than the threshold is considered as having been retained and coded as 1, and every student whose predicted value is below the threshold is considered not to have dropped out and coded as zero. Using this classification, we compute the sensitivity and false positive rate at that threshold. The algorithm repeats this calculation for every possible threshold from the set of predicted values and plots the estimated tradeoff between sensitivity and the false positive rate at each point. As an example, consider a threshold for first-year retention of 0.1. Every student whose predicted probability is greater than or equal to 0.1 is assigned a 1 and all predicted values below 0.1 are assigned a zero. At a threshold of 0.1, the model will have high sensitivity since the mean retention rate is roughly 86 percent, but it will misclassify students as having continued who did not and thus a high false-positive rate. The optimal threshold or cut point is where the difference between the sensitivity and false positive rate is greatest. Given the many observations in our sample, the ROC displays as a continuous curve (see Fawcett 2006; Nahm 2022)

Pell award increases from 59 percent to 67 percent over the same period. Among academic outcomes, first-year retention falls 4 percentage points while first-year GPA and four-year graduation rates change a little.

4.1 Part I: 2010-2019 Cohorts

Table 2 shows the effect of a one standard deviation increase in HSGPA, SAT and the ELA exam score on first-year outcomes among the pre-COVID cohorts. Each column exhibits estimates from a separate regression and all models include college- and year-fixed effects as well indicators of Pell support, underrepresented minorities (URM), age and male gender. As is standard in the literature, we estimate models with HSGPA and SAT separately (columns 1 and 2) and then combined (column 3). The estimates in column 4 add high school fixed effects. In columns 5 and 6 we replace SAT scores with a student's score on the ELA exam.

High school GPA is the most important predictor of both first-year GPA and first-year retention (Table 2, Panels A and B). A one-standard deviation increase in HSGPA raises first-year GPA by between 0.36 to 0.41 points over a mean of 2.86. Not only is the coefficient stable across specifications, but the inclusion of high school fixed effects increases the importance of HSGPA and diminishes the contribution of both standardized tests, the SAT and ELA exam (columns 4 and 6). The same pattern pertains to retention to the second year (Table 2, Panel B). A one-standard deviation increase in HSGPA raises retention by 6 percentage points in models that include both SAT scores and high school fixed effects (Table 2, Panel B, column 4) while the effect of the ELA exam also becomes negative (Table 2, Panel B, column 6).

In *Crossing the Finish Line* Bowen et al. (2009) underscored the importance of finishing and not just starting college. In Table 3 we contrast the effect of HSGPA, SAT and the ELA exam on four- and six-year graduation rates. Both HSGPA and SAT are important predictors of four-year graduation rates (Table 3, Panel A). But as with the first-year outcomes, the effect of a one-standard deviation increase in HSGPA raises the probability of graduation in four-years by between 12 and 14 percentage points over a mean of 35 percent. SAT scores are statistically a significant predictor of four-year graduation rates, but their effect is less than 25 percent of the effect of HSGPA with the inclusion of high school fixed effects (Panel A, column 4).

The difference is even more extreme with six-year graduation rates. The effect of SAT and the ELA exam is roughly a tenth of the effect of HSGPA (Panel B, columns 3 and 5), but their effects become negative with the inclusion of high school fixed effects (Panel B, columns 4 and 6). Our results are remarkably close to those of Bowen et al. (2009). Not only is the effect of HSGPA an order of magnitude larger than that of SATs, but the inclusion of high school fixed effects turns the coefficient on SAT scores from positive to negative. [9]

As noted in the Methods sections, the Gelbach decomposition (Gelbach, 2016) provides a systematic way to attribute the change in a linear regression coefficient to the inclusion of other covariates. In Panel A of Table 4 we illustrate the decomposition with first-year GPA. In a simple bivariate regression, a one standard deviation increase in SAT scores increases

[9] Contrast estimates in Table 6.1 to those in Table 6.5 of Bowen, Chingos and McPherson (2009).

first-year GPA by 0.225 points (column 1). Once we include the full set of regressors, the coefficient on SAT falls to 0.042, a change of 0.183 points (columns 2 and 3). High school GPA accounts for 0.213 of the decline and the complete set of other controls accounts for -0.030 (columns 4 and 5). In the second row of Panel A, we reverse the order. A one standard deviation increase in HSGPA raises first-year GPA by 0.427 points in the simple bivariate regression (column 1). Inclusion of all the other covariates changes the coefficient on HSGPA by a meager 0.018 points of which SAT account for 0.015 points of the decline.

The pattern repeats with six-year graduation rates (Table 4, Panel B). A one-standard deviation increase in the SAT increases the probability of graduating in six years by 0.061 in the simple bivariate regression (Panel B, column 1). Inclusion of the other controls decreases the coefficient on SAT by 0.084 of which 0.075 is explained by the inclusion of HSGPA alone. Again, we reverse the order and regress six-year graduation rates on HSGPA. The probability of graduating in six years improves by 0.157 for a one-standard deviation increase. Inclusion of all other covariates does little damage to this association as the coefficient on HSGPA falls by an inconsequential 0.01 (Panel B, column 3).

In column 6 of Panel A and B we use the Shapley decomposition to show the proportion of the R^2 accounted for by SAT scores and HSGPA for each outcome. High school GPA explains 4 times more of the variation in six-year graduation rates than do SAT scores (70.2% vs 15.5%).

4.2 Subgroup Analysis and Robustness checks: 2010-2019 Cohorts

We present results in the Appendix stratified by two subgroups: under-represented minorities (URM) and those awarded a full Pell grant as compared to those with partial Pell support or no Pell support. We show first-year outcomes and graduation rates by URM in Appendix Tables A1 and A2 and the same outcomes by Pell status in Appendix Tables A3 and A4. To summarize, there are no substantive differences between the subgroups. The effect of HSGPA on all outcomes dominates the effect of SATs or the ELA exam. The concordance between subgroups is clearest is the specifications that include high school fixed effects (columns 4 and 6 in each Table).

Although HSGPA is a robust predictor of student success, it could still under or over predict student outcomes. Friedman et al. (2025a), for example, found that at each level of SAT scores, predicted first-year GPAs were lower for URM relative to non-URM. They concluded SATs did not evidence calibration bias because they did not underpredict a URM's success in college.[10] However, calibration bias is more commonly defined as $E(Y|S, G = 1) \neq$

[10] This interpretation of calibration bias is specific to the concern that SATs lessen acceptance probabilities of disadvantaged applicants at highly ranked colleges. The argument is that higher SATs may reflect society's inequities by favoring better resourced applicants who can hire test-prep tutors or enroll in test-prep classes, in addition to taking the SATs multiple times. The lower SAT scores of less advantaged applicants may not reflect less academic potential, but rather less supportive resources. If true, then the lower scores of less advantage students will underpredict their performance relative to more advantaged students and possibly lessen their likelihood of being accepted. Friedman et al. (2025) find the opposite to be true and conclude that SATs are not biased against disadvantage applicants at the Ivy-plus colleges.

$E(Y|S, G = 0)$ where E is expected value operator, S is some score or index used to predict Y and G is membership in a group (Obermeyer et al., 2019).

We undertake a similar exercise to Friedman et al. (2025a) with binned scatter plots of six-year graduation rates against HSGPA stratified by URM and whether the student received the maximum Pell grant (Figures 1a and 1b). We find that URM are less likely to graduate in six years at every level of HSGPA than are non-URM. Despite being the dominant predictor of graduation rates, HSGPA will overestimate, on average, a URM's likelihood of graduating relative to a non-URM in a model that includes all applicants. Consequently, any calibration bias in HSGPA operates *in favor* of URM students, as the metric implies a higher probability of degree completion than is realized. A possible explanation is the quality of a student's primary and secondary schools in terms of resources and rigor that are not captured by high school fixed effects. In contrast, we find no similar difference when we stratify the plot by whether the student received the max Pell grant or not, a strong proxy for a low-income household (Figure 1b).

The clear takeaway from the analysis of the 2010-2019 cohorts is the dominance of HSGPA in predicting student academic success. The results are in direct contrast to recent high-profile analyses of students from Ivy-plus colleges that demonstrated the superiority of SAT over HSGPA in predicting not only first-year academic outcomes, but post-college success as well (Friedman et al., 2025a; Chetty et al., forthcoming). Our results are remarkably similar to those of Bowen et al. (2009) and Allensworth & Clark (2020) who analyzed different samples of public university students from disparate cohorts.

4.3 Part II: 2020-2024 Cohorts

By 2021 many colleges had made SAT scores optional for admissions. The policy increased the importance of HSGPA along with other parts of an applicant's portfolio. However, public colleges with small admission staff and large applicant pools most likely relied on HSGPA even more. In this section, we estimate the out-of-sample prediction errors associated with models that use only HSGPA to forecast student academics outcomes.

As noted in the Methods' section, we use the 2017-18 cohorts to predict the outcomes of the 2019 cohort both with and without SAT scores. The results are shown in Panel A of Table 5. The mean absolute error with both HSGPA and SAT in the model is 0.611 first-year GPA points or 21 percent of the mean first-year GPA in 2019 of 2.92 (Table 5, Panel A, row 1). The sign rank is positive in which 65% of the actual first-year GPAs exceed the predicted. Little changes when we drop SAT from the model (row 2). However, if we exclude HSGPA from the predicting equation, the MAE rises from 0.611 to 0.694 first-year GPA points and the model underpredicts the actual first-year GPA 67.7 percent of the time.

In Panel B, we repeat the exercise but change the target year to 2020 and the "training years" to 2017-2019. The MAE is larger in 2020 than in 2019, 0.677 vs 0.611 (Table 5, Panel B). In models with only the SAT, the MAE rises to 0.773 in 2020 or 27 percent of the mean. As before, the model underpredicts the actual first-year GPA. Panel C shows the same result for the years after 2020. Of note is the sign rank is close to 50% and the z-score is negative and significant in 2021 and 2022 while the MAE is modestly smaller than in Panels A and B.

In Table 6 we present results from the same exercise but with year-one retention as the outcome. Because the outcome is binary, we use an ROC curve to assess the sensitivity and specificity of the model. Based on the estimates in panel A, excluding SAT from the model has no effect on the area under the ROC curve (AUC) or the sensitivity and specificity of the model. However, if we exclude HSGPA, the AUC falls from 0.702 to 0.64 while the sensitivity of the model falls from 0.67 to 0.60. Figure 2 provides a visual presentation of the ROC curve. The 45-degree line indicates a one-for-one tradeoff between sensitivity and the false positive rate (1-specificity). The ROC curves moves closer to the 45 degree line with the exclusion of HSGPA from the predicted model while dropping SAT has no effect on the ROC curve.

The results in panel B and C offer the same conclusion as with first-year GPA. Models become less predictive with only HSGPA of first-year retention among the 2021-2023 cohorts relative to the 2019 cohorts but the differences are not large. In figure A1 we show ROC curves for first-year retention for 2021-2023.

5 Conclusion

Prior to COVID, most colleges and universities required the SAT for admission. By 2022, the vast majority of colleges made the SAT optional (Avery et al., 2025). The dramatic shift re-ignited a long-standing debate about the usefulness of the SAT as a predictor of an applicant's readiness for college and the fairness of the test given societal inequities linked to race and class. Many of the country's most selective colleges have returned to requiring the SAT. The justification is based on evidence that the test not only predicts student success, but facilitates the identification of qualified applicants from disadvantaged backgrounds that were less likely to be admitted under a test-optional policy (Friedman et al., 2025a; Sacerdote et al., forthcoming).

The results from our study of a large public urban university system find the opposite: high school grades are a vastly superior predictor of student academic success than is the SAT. Our findings are not new, but they are important to re-establish given revisions to the SAT and the high-profile analyses from Opportunity Insights (Chetty et al., forthcoming; Friedman et al., 2025a). Moreover, our results are remarkably close to previous work across a range of public colleges, universities, cohorts and samples. But COVID, and now artificial intelligence, may change the preparation and learning of students going forward. Whether high school grades remain the most salient predictor of college graduation will need continued study.

References

Allensworth, E. M., & Clark, K. (2020). High school GPAs and ACT scores as predictors of college completion: Examining assumptions about consistency across high schools. *Educational Researcher*, 49(3), 198–211. https://doi.org/10.3102/0013189X20902110

Alon, S., & Tienda, M. (2007). Diversity, opportunity, and the shifting meritocracy in higher education. *American Sociological Review*, 72(4), 487–511.

Avery, C., Shi, L., & Magouirk, P. (2025). Test-optional college admissions: ACT and SAT scores, applications, and enrollment changes. *National Bureau of Economic Research*, Working Paper (34260).

Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2009). *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton University Press. http://www.jstor.org/stable/j.ctt7rp39

U.S. Census Bureau. (2017). *Decomposing the R-squared of a regression using the Shapley value in SAS*. Retrieved from https://www.census.gov/library/working-papers/2017/econ/2017-coleman.html

Cascio, E., Sacerdote, B., Staiger, D., & Tine, M. (2024). *Report from working group on the role of standardized test scores in undergraduate admissions*. Dartmouth College.

Chetty, R., Deming, D. J., & Friedman, J. N. (2023). Diversifying society's leaders? The determinants and causal effects of admission to highly selective private colleges. *National Bureau of Economic Research*, Working Paper 31492. July.

Chetty, R., Deming, D. J., & Friedman, J. N. (forthcoming). Diversifying society's leaders? The determinants and causal effects of admission to highly selective private colleges. *The Quarterly Journal of Economics*.

Deming, D. (2024, March 7). The worst way to do college admissions. *The Atlantic*.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Fletcher, J. M., & Tienda, M. (2010). Race and ethnic differences in college achievement: Does high school attended matter? *The Annals of the American Academy of Political and Social Science*, 627(1), 144–166. https://doi.org/10.1177/0002716209348749

Freedle, R. O. (2003). Correcting the SAT's ethnic and social-class bias: A method for reestimating SAT scores. *Harvard Educational Review*, 73(1), 1–43. https://doi.org/10.17763/haer.73.1.8465k88616hn4757

Friedman, J. N., Sacerdote, B., Staiger, D. O., & Tine, M. (2025a). Standardized test scores and academic performance at Ivy-Plus colleges. *National Bureau of Economic Research*, Working Paper 33570. https://doi.org/10.3386/w33570

Friedman, J. N., Sacerdote, B., Staiger, D. O., & Tine, M. (2025b). Standardized test scores and academic performance at Ivy-Plus colleges. *AEA Papers and Proceedings*, 115, 676–681. https://doi.org/10.1257/pandp.20251056

Galla, B. M., Shulman, E. P., Plummer, B. D., Gardner, M., Hutt, S. J., Goyer, J. P., D'Mello, S. K., Finn, A. S., & Duckworth, A. L. (2019). Why high school grades are better predictors of on-time college graduation than are admissions test scores: The roles of self-regulation and cognitive ability. *American Educational Research Journal*, 56(6), 2077–2115. https://doi.org/10.3102/0002831219843292

Gelbach, J. B. (2016). When do covariates matter? And which ones, and how much? *Journal of Labor Economics*, 34(2), 509–543. https://doi.org/10.1086/683668

Goodman, J., Gurantz, O., & Smith, J. (2020). Take two! SAT retaking and college enrollment gaps. *American Economic Journal: Economic Policy*, 12(2), 115–158. https://doi.org/10.1257/pol.20170503

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *Validity of the SAT for predicting first-year college grade point average*. College Board.

Kurlaender, M., & Cohen, K. (2019). *Predicting college success: How do different high school assessments measure up?* Policy Analysis for California Education. https://edpolicyinca.org/sites/default/files/R_Kurlaender_Mar-2019.pdf

Kurlaender, M., Reber, S., & Rothstein, J. (2020). *Commentary on the UC Standardized Testing Task Force report*.

Lemann, N. (1999). *The Big Test: The Secret History of the American Meritocracy*. Farrar, Straus and Giroux, 19 Union Square West, New York, NY 10003.

Leonhardt, D. (2024, January 7). The misguided war on the SAT. *The New York Times*. https://www.nytimes.com/2024/01/07/briefing/the-misguided-war-on-the-sat.html

Mattern, K. D., & Patterson, B. F. (2014). *Synthesis of recent SAT validity findings: Trend data over time and cohorts*. College Board. https://eric.ed.gov/?id=ED556462

Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and practical use for clinicians. *Korean Journal of Anesthesiology*, 75(1), 25–36. https://doi.org/10.4097/kja.21209

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics*, 121(1), 297–317. https://doi.org/10.1016/j.jeconom.2003.10.003

Sacerdote, B., Staiger, D. O., & Tine, M. (2025). How test optional policies in college admissions disproportionately harm high achieving applicants from disadvantaged backgrounds. *National Bureau of Economic Research*, Working Paper 33389. https://doi.org/10.3386/w33389

Sacerdote, B., Staiger, D. O., & Tine, M. (forthcoming). How test optional policies in college admissions disproportionately harm high achieving applicants from disadvantaged backgrounds. *American Economic Review: Insights*.

Soares, J. A. (2011). *SAT Wars: The Case for Test-Optional College Admissions*. (New York: Teachers College Press.)

The Economist. (2025, April 16). America's progressives should love standardized tests. *The Economist*.

Westrick, P. A., Marini, J., Young, L., Ng, H., Shmueli, D., & Shaw, E. J. (2019). *Validity of the SAT for predicting first-year grades and retention to the second year*. College Board.

Tables and Figures

Table 1: *Student Characteristics and Outcomes: Pre- vs. Post-COVID*

Variable	Pre-COVID (2010–2019) Mean	Post-COVID (2020–2024) Mean
Student Demographics		
URM (Underrepresented Minority)	0.42	0.50
Academic Preparedness		
High School GPA	85.70	87.66
SAT Score	1112.48	–
Pell Status		
No Pell	0.41	0.34
Some Pell	0.19	0.23
Maximum Pell	0.40	0.44
Key System Outcomes		
First-Year GPA	2.86	2.82
Retained after Year 1	0.88	0.84
4-Year Graduation (bach4)	0.35	0.36
6-Year Graduation (bach6)	0.64	–

Notes: Pre-COVID $N = 134,360$. Post-COVID $N = 91,654$. Dashes (–) indicate data not available or not applicable for the recent cohorts (SAT was not required/available for many; 6-year graduation window has not yet closed for 2020+ entrants).

Table 2: *Academic Outcomes: SAT vs. ELA*

	(1) HSGPA Only	(2) Test Only	(3) HSGPA + SAT	(4) HSGPA + SAT	(5) HSGPA + ELA	(6) HSGPA + ELA
Panel A: First-Year GPA						
Z-HSGPA	0.380***		0.358***	0.409***	0.360***	0.405***
	(0.003)		(0.003)	(0.003)	(0.003)	(0.003)
Z-SAT Total		0.224***	0.142***	0.042***		
		(0.003)	(0.003)	(0.004)		
Z-ELA					0.086***	0.052***
					(0.003)	(0.003)
HS FE	No	No	No	Yes	No	Yes
Mean Y	2.86	2.86	2.86	2.86	2.84	2.84
R^2	0.253	0.146	0.263	0.308	0.261	0.314
N	122,862	121,696	118,582	118,127	112,462	112,337
Panel B: Retention to Second Year						
Z-HSGPA	0.053***		0.052***	0.060***	0.053***	0.059***
	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
Z-SAT Total		0.011***	-0.001	-0.016***		
		(0.001)	(0.001)	(0.001)		
Z-ELA					0.001	-0.004***
					(0.001)	(0.001)
HS FE	No	No	No	Yes	No	Yes
Mean Y	0.89	0.89	0.89	0.89	0.89	0.89
R^2	0.055	0.030	0.051	0.076	0.057	0.081
N	123,991	122,766	119,623	119,163	113,444	113,315

Notes: OLS (Panel A) and LPM (Panel B). Columns (1) HSGPA only; (2) SAT only; (3) HSGPA+SAT; (4) HSGPA+SAT with HS FE; (5) HSGPA+ELA; (6) HSGPA+ELA with HS FE. All columns include fall-year and college fixed effects. Controls include: male, Pell (<max and max), URM, and age bins. Predictors are standardized (z-scores) within the analysis window. Coefficients with standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3: *Graduation Outcomes: SAT vs. ELA*

	(1) HSGPA Only	(2) Test Only	(3) HSGPA + SAT	(4) HSGPA + SAT (+ FE)	(5) HSGPA + ELA	(6) HSGPA + ELA (+ FE)
Panel A: 4-Year Graduation						
Z-HSGPA	0.130***		0.120***	0.139***	0.123***	0.141***
	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)
Z-SAT Total		0.098***	0.068***	0.028***		
		(0.002)	(0.002)	(0.002)		
Z-ELA					0.030***	0.013***
					(0.002)	(0.002)
Mean Y	0.33	0.33	0.35	0.35	0.35	0.35
R^2	0.155	0.117	0.163	0.191	0.158	0.192
N	91,749	90,025	88,419	87,959	85,440	84,112
Panel B: 6-Year Graduation						
Z-HSGPA	0.128***		0.127***	0.148***	0.127***	0.143***
	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)
Z-SAT Total		0.043***	0.015***	-0.023***		
		(0.002)	(0.002)	(0.002)		
Z-ELA					0.009***	-0.001
					(0.002)	(0.002)
Mean Y	0.65	0.65	0.65	0.65	0.65	0.65
R^2	0.141	0.086	0.138	0.169	0.144	0.174
N	107,804	106,784	104,170	104,170	98,335	98,335

Notes: Linear probability model. Dependent variables are 4-year and 6-year graduation indicators. All columns include fall-year and college fixed effects. Columns (4) and (6) absorb high-school fixed effects. Controls: male, Pell, URM, and age. Honors, ACE, and SEEK are excluded. Predictors are standardized. 4-Year sample restricted to entrants $\geq$ 2013; 6-Year sample restricted to entrants $\leq$ 2018. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4: *Decomposition of Coefficient Changes (Gelbach) and Model Variance (Shapley)*

Variable	**Coefficient Stability (Gelbach)** (1) Unadjusted β	(2) Adjusted β	(3) Total Difference	**Explained By** (4) Due to Other Test	(5) Due to Controls	**Shapley** (6) Share of Total R^2
Panel A: First-Year GPA						
Main Var: Z-SAT	0.225***	0.042***	0.183	0.213	-0.030	21.8%
	(0.003)	(0.004)		(HSGPA)		
Main Var: Z-HSGPA	0.427***	0.409***	0.018	0.015	0.003	70.3%
	(0.002)	(0.003)		(SAT)		
Panel B: 6-Year Graduation						
Main Var: Z-SAT	0.061***	-0.023***	0.084	0.075	0.009	15.5%
	(0.002)	(0.002)		(HSGPA)		
Main Var: Z-HSGPA	0.157***	0.148***	0.010	-0.008	0.018	70.2%
	(0.001)	(0.002)		(SAT)		

Notes: **Cols 1-5 (Gelbach):** Estimates decomposition of the change in the main variable's coefficient when moving from a bivariate model (Col 1) to the full model (Col 2). "Due to Other Test" shows how much of the coefficient change is explained by adding the alternative metric (HSGPA for SAT rows, and vice-versa). "Due to Controls" includes student demographics, Pell, and FE. **Col 6 (Shapley):** Reports the percentage contribution of the variable to the total R^2. Total R^2 is 0.252 for FYGPA and 0.122 for Graduation. All models include college and cohort fixed effects and absorb high school fixed effects. Coefficients shown with standard errors in parentheses; *** $p < 0.001$.

Table 5: *Out-of-Sample Prediction Accuracy and Bias for First-Year GPA*

Target Year	Training Years	Model	Mean Y	MAE	MAE / Mean Y	% Under	Sign Rank z
Panel A: Pre-COVID (2019)							
2019	2017–2018	HSGPA + SAT	2.92	0.611	0.209	65.7%	28.18
		HSGPA Only	2.92	0.618	0.212	65.4%	27.84
		SAT Only	2.92	0.694	0.238	67.7%	31.80
Panel B: During COVID (2020)							
2020	2017–2019	HSGPA + SAT	2.90	0.677	0.233	61.6%	12.57
		HSGPA Only	2.90	0.679	0.234	61.9%	13.91
		SAT Only	2.90	0.773	0.267	65.1%	21.93
Panel C: Post-COVID (2021–2023)							
2021	2020	HSGPA Only	2.77	0.653	0.236	50.4%	-12.83
2022	2020–2021	HSGPA Only	2.77	0.636	0.230	53.3%	-2.59
2023	2020–2022	HSGPA Only	2.81	0.631	0.225	57.0%	7.84

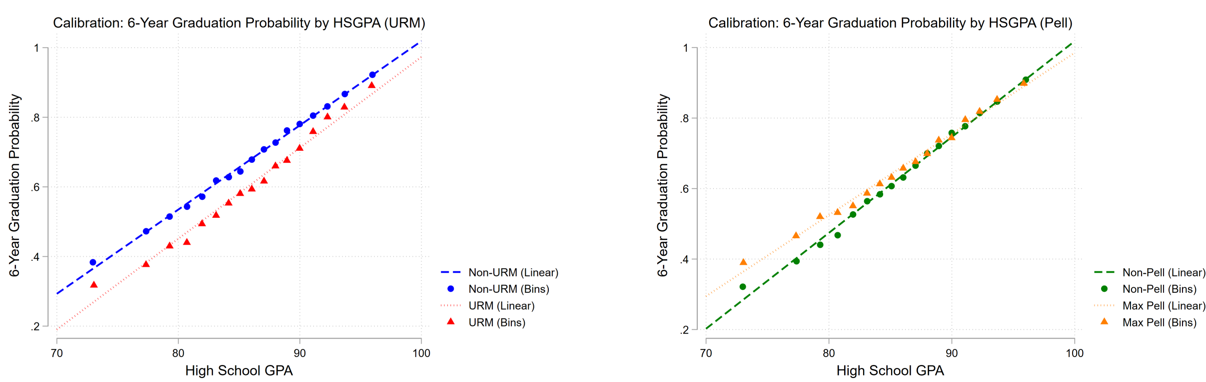
Notes: **MAE** is the Mean Absolute Error of the prediction. **Mean** Y is the observed average First-Year GPA for the target cohort. **% Under** represents the percentage of students whose actual GPA was higher than their predicted GPA (prediction error > 0). A value greater than 50% indicates systematic underprediction; a value near or below 50% with a negative z-score suggests overprediction or symmetry. **Sign Rank** z is the z-statistic from a Wilcoxon signed-rank test testing the null hypothesis that the prediction errors are symmetric around zero. A large positive z indicates systematic underprediction (Actual > Predicted); a negative z indicates overprediction. All reported z-statistics are statistically significant at $p < 0.01$. In Panel C (Post-COVID), models rely solely on HSGPA and controls because SAT scores were not required. All prediction models include college and cohort fixed effects and absorb high school fixed effects.

Table 6: *Out-of-Sample Classification Accuracy for Year-1 Retention*

Target	Training Years	Model	AUC	Opt. Cut	Sens.	Spec.
Panel A: Pre-COVID (2019)						
2019	2017–2018	HSGPA + SAT	0.702	0.885	0.67	0.62
2019	2017–2018	HSGPA only	0.702	0.879	0.67	0.64
2019	2017–2018	SAT only	0.640	0.894	0.60	0.58
Panel B: During COVID (2020)						
2020	2017–2019	HSGPA + SAT	0.707	0.896	0.66	0.66
		HSGPA Only	0.705	0.889	0.66	0.66
		SAT Only	0.653	0.887	0.67	0.55
Panel C: Post-COVID (2021–2023)						
2021	2020	HSGPA only	0.640	0.867	0.60	0.67
2022	2020–2021	HSGPA only	0.650	0.853	0.66	0.65
2023	2020–2022	HSGPA only	0.640	0.861	0.62	0.66

Notes: AUC represents the Area Under the Receiver Operating Characteristic Curve. Optimal cutpoints for classification are determined using the Liu method. Sensitivity (Sens.) refers to the true positive rate, while Specificity (Spec.) refers to the true negative rate. Post-COVID models use HSGPA only due to test-optional policies.

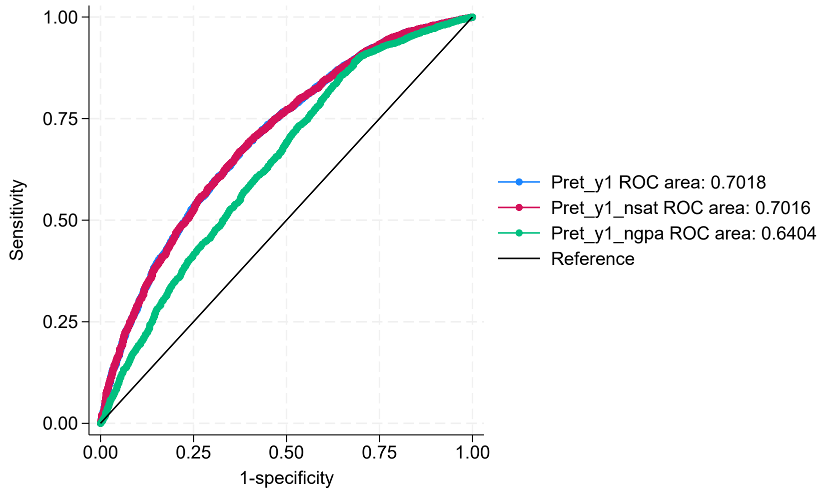
Figure 1: *Calibration of 6-Year Graduation Probability by High School GPA, stratified by URM and Pell*



(a) *URM vs. Non-URM* **(b)** *Pell vs. Non-Pell*

Notes: Binned scatter plots of 6-year graduation probability against High School GPA. The dashed lines represent linear fits from a regression of graduation on HSGPA, SAT scores, and demographic controls (age, gender, cohort, college), absorbing high school fixed effects. The dots represent predictive margins for 17 quantiles of HSGPA, calculated while holding other covariates at their means.

Figure 2: *ROC Curves for First-Year Retention Prediction (Pre-COVID 2019)*



Notes: The figure compares the predictive power of models using HSGPA only (Blue), SAT only (Red), and both (Green) for the 2019 cohort. The diagonal line represents random guessing (AUC = 0.50).

Appendix

Table A1: *First-Year Outcomes by URM Status*

	(1) HSGPA Only	(2) Test Only	(3) HSGPA + SAT	(4) HSGPA + SAT	(5) HSGPA + ELA	(6) HSGPA + ELA
Panel A: First-Year GPA						
URM Students						
Z-HSGPA	0.390***		0.376***	0.422***	0.374***	0.417***
	(0.004)		(0.004)	(0.005)	(0.005)	(0.005)
Z-SAT Total		0.212***	0.122***	0.016*		
		(0.006)	(0.006)	(0.006)		
Z-ELA					0.083***	0.038***
					(0.005)	(0.005)
HS FE	No	No	No	Yes	No	Yes
Mean Y	2.64	2.65	2.63	2.65	2.63	2.65
R^2	0.217	0.111	0.222	0.283	0.224	0.288
N	51,029	59,652	49,315	49,134	48,206	48,111
Non-URM Students						
Z-HSGPA	0.375***		0.347***	0.401***	0.351***	0.399***
	(0.003)		(0.003)	(0.003)	(0.003)	(0.003)
Z-SAT Total		0.241***	0.154***	0.059***		
		(0.003)	(0.004)	(0.004)		
Z-ELA					0.103***	0.068***
					(0.004)	(0.004)
HS FE	No	No	No	Yes	No	Yes
Mean Y	3.01	3.03	3.00	3.00	2.99	2.99
R^2	0.227	0.122	0.248	0.295	0.242	0.300
N	71,833	82,885	69,267	68,993	64,256	64,226
Panel B: Retention to Second Year						
URM Students						
Z-HSGPA	0.055***		0.052***	0.060***	0.053***	0.059***
	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
Z-SAT Total		0.011***	0.004***	-0.016***		
		(0.001)	(0.001)	(0.001)		
Z-ELA					0.002	-0.004***
					(0.001)	(0.001)
HS FE	No	No	No	Yes	No	Yes
Mean Y	0.86	0.87	0.86	0.86	0.84	0.86
R^2	0.055	0.030	0.051	0.076	0.037	0.081
N	123,991	122,766	119,623	119,163	115,999	113,315
Non-URM Students						
Z-HSGPA	0.054***		0.054***	0.061***	0.056***	0.061***
	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
Z-SAT Total		0.021***	0.004***	-0.015***		
		(0.001)	(0.001)	(0.001)		
Z-ELA					0.002	-0.004***
					(0.001)	(0.001)
HS FE	No	No	No	Yes	No	Yes
Mean Y	0.89	0.89	0.89	0.89	0.89	0.89
R^2	0.047	0.105	0.047	0.075	0.053	0.080
N	119,623	144,106	119,596	118,200	113,444	112,415

Notes: OLS (Panel A) and LPM (Panel B). All models include college and cohort fixed effects. Columns (4) and (6) absorb high-school fixed effects. Controls include: male, Pell (<max and max), and age bins. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table A2: *Graduation Outcomes by URM Status*

	(1) HSGPA Only	(2) Test Only	(3) HSGPA + SAT	(4) HSGPA + SAT	(5) HSGPA + ELA	(6) HSGPA + ELA
Panel A: 4-Year Graduation						
URM Students						
Z-HSGPA	0.107***		0.104***	0.117***	0.102***	0.114***
	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)
Z-SAT Total		0.067***	0.044***	0.010**		
		(0.003)	(0.003)	(0.003)		
Z-ELA					0.022***	0.008***
					(0.002)	(0.002)
HS FE	No	No	No	Yes	No	Yes
Mean Y	0.25	0.26	0.25	0.25	0.25	0.25
R^2	0.116	0.078	0.119	0.155	0.117	0.154
N	51,567	58,780	49,804	49,618	48,699	48,600
Non-URM Students						
Z-HSGPA	0.140***		0.126***	0.144***	0.132***	0.151***
	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)
Z-SAT Total		0.113***	0.081***	0.044***		
		(0.002)	(0.002)	(0.003)		
Z-ELA					0.039***	0.019***
					(0.002)	(0.002)
HS FE	No	No	No	Yes	No	Yes
Mean Y	0.39	0.39	0.39	0.39	0.38	0.38
R^2	0.150	0.113	0.164	0.193	0.157	0.195
N	72,424	80,425	69,819	69,545	64,745	64,715
Panel B: 6-Year Graduation						
URM Students						
Z-HSGPA	0.137***		0.138***	0.157***	0.136***	0.152***
	(0.003)		(0.003)	(0.003)	(0.003)	(0.003)
Z-SAT Total		0.032***	0.006	-0.039***		
		(0.003)	(0.003)	(0.004)		
Z-ELA					0.008**	-0.008**
					(0.003)	(0.003)
HS FE	No	No	No	Yes	No	Yes
Mean Y	0.55	0.55	0.55	0.55	0.55	0.55
R^2	0.118	0.059	0.115	0.160	0.119	0.161
N	43,748	43,394	42,375	42,375	41,264	41,264
Non-URM Students						
Z-HSGPA	0.119***		0.122***	0.141***	0.120***	0.136***
	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)
Z-SAT Total		0.049***	0.020***	-0.013***		
		(0.002)	(0.002)	(0.003)		
Z-ELA					0.013***	0.003
					(0.002)	(0.002)
HS FE	No	No	No	Yes	No	Yes
Mean Y	0.72	0.72	0.72	0.72	0.72	0.72
R^2	0.114	0.063	0.114	0.149	0.117	0.152
N	64,056	63,390	61,795	61,795	57,071	57,071

Notes: Linear probability model. Dependent variables are 4-year and 6-year graduation. All models include college and cohort fixed effects. Columns (4) and (6) absorb high-school fixed effects. Controls include: male, Pell (<max and max), and age bins. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table A3: *First-Year Outcomes by Pell Status*

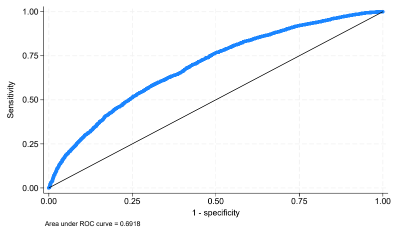
	(1) HSGPA Only	(2) Test Only	(3) HSGPA + SAT	(4) HSGPA + SAT	(5) HSGPA + ELA	(6) HSGPA + ELA
Panel A: First-Year GPA						
Max Pell Students						
Z-HSGPA	0.326***		0.315***	0.362***	0.313***	0.356***
	(0.004)		(0.004)	(0.004)	(0.004)	(0.004)
Z-SAT Total		0.177***	0.116***	0.014*		
		(0.005)	(0.005)	(0.006)		
Z-ELA					0.083***	0.045***
					(0.004)	(0.004)
HS FE	No	No	No	Yes	No	Yes
Mean Y	2.81	2.83	2.81	2.81	2.80	2.80
R^2	0.222	0.134	0.231	0.288	0.233	0.293
N	51,335	58,845	49,953	49,822	48,807	48,755
Non-Max Pell Students						
Z-HSGPA	0.422***		0.391***	0.446***	0.397***	0.445***
	(0.003)		(0.004)	(0.004)	(0.004)	(0.004)
Z-SAT Total		0.262***	0.154***	0.055***		
		(0.004)	(0.004)	(0.005)		
Z-ELA					0.101***	0.063***
					(0.004)	(0.004)
HS FE	No	No	No	Yes	No	Yes
Mean Y	2.89	2.90	2.89	2.89	2.87	2.87
R^2	0.275	0.163	0.285	0.335	0.284	0.342
N	71,527	83,692	68,629	68,305	63,655	63,582
Panel B: Retention to Second Year						
Max Pell Students						
Z-HSGPA	0.046***		0.046***	0.054***	0.046***	0.051***
	(0.001)		(0.002)	(0.002)	(0.002)	(0.002)
Z-SAT Total		0.003	-0.007***	-0.023***		
		(0.002)	(0.002)	(0.002)		
Z-ELA					0.000	-0.005**
					(0.002)	(0.002)
HS FE	No	No	No	Yes	No	Yes
Mean Y	0.89	0.88	0.90	0.90	0.89	0.89
R^2	0.048	0.130	0.046	0.083	0.050	0.084
N	51,520	59,117	50,127	49,995	48,973	48,921
Non-Max Pell Students						
Z-HSGPA	0.057***		0.056***	0.065***	0.059***	0.066***
	(0.001)		(0.001)	(0.002)	(0.001)	(0.002)
Z-SAT Total		0.018***	0.001	-0.012***		
		(0.001)	(0.002)	(0.002)		
Z-ELA					0.001	-0.005**
					(0.002)	(0.002)
HS FE	No	No	No	Yes	No	Yes
Mean Y	0.88	0.87	0.88	0.88	0.88	0.88
R^2	0.060	0.098	0.055	0.089	0.063	0.097
N	72,471	84,989	69,496	69,168	64,471	64,394

Notes: OLS (Panel A) and LPM (Panel B). All models include college and cohort fixed effects. Columns (4) and (6) absorb high-school fixed effects. Controls include: male, URM, and age bins. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

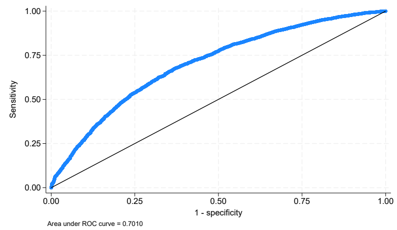
Table A4: *Graduation Outcomes by Pell Status*

	(1) HSGPA Only	(2) Test Only	(3) HSGPA + SAT	(4) HSGPA + SAT	(5) HSGPA + ELA	(6) HSGPA + ELA
Panel A: 4-Year Graduation						
Max Pell Students						
Z-HSGPA	0.105***		0.102***	0.118***	0.102***	0.118***
	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)
Z-SAT Total		0.075***	0.057***	0.022***		
		(0.003)	(0.003)	(0.003)		
Z-ELA					0.028***	0.012***
					(0.002)	(0.002)
HS FE	No	No	No	Yes	No	Yes
Mean Y	0.29	0.30	0.30	0.30	0.29	0.29
R^2	0.136	0.107	0.144	0.176	0.140	0.175
N	51,520	56,988	50,127	49,995	48,973	48,921
Non-Max Pell Students						
Z-HSGPA	0.142***		0.128***	0.145***	0.134***	0.150***
	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)
Z-SAT Total		0.112***	0.077***	0.039***		
		(0.002)	(0.002)	(0.003)		
Z-ELA					0.036***	0.018***
					(0.002)	(0.002)
HS FE	No	No	No	Yes	No	Yes
Mean Y	0.36	0.36	0.36	0.36	0.35	0.35
R^2	0.160	0.118	0.169	0.200	0.164	0.203
N	72,471	82,217	69,496	69,168	64,471	64,394
Panel B: 6-Year Graduation						
Max Pell Students						
Z-HSGPA	0.108***		0.109***	0.129***	0.108***	0.123***
	(0.002)		(0.002)	(0.003)	(0.003)	(0.003)
Z-SAT Total		0.025***	0.008*	-0.032***		
		(0.003)	(0.003)	(0.004)		
Z-ELA					0.008**	-0.003
					(0.003)	(0.003)
HS FE	No	No	No	Yes	No	Yes
Mean Y	0.64	0.64	0.65	0.65	0.64	0.64
R^2	0.130	0.088	0.128	0.170	0.133	0.172
N	44,257	43,245	43,126	43,126	42,002	42,002
Non-Max Pell Students						
Z-HSGPA	0.142***		0.140***	0.162***	0.141***	0.158***
	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)
Z-SAT Total		0.054***	0.017***	-0.019***		
		(0.002)	(0.002)	(0.003)		
Z-ELA					0.011***	-0.002
					(0.002)	(0.002)
Mean Y	0.65	0.65	0.65	0.65	0.65	0.65
R^2	0.151	0.087	0.147	0.184	0.154	0.190
N	63,547	63,539	61,044	61,044	56,333	56,333

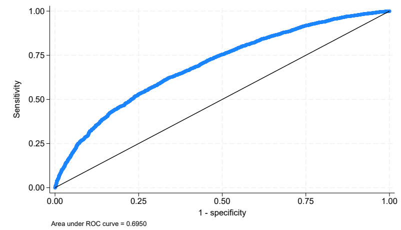
Notes: Linear probability model. Dependent variables are 4-year and 6-year graduation. All models include college and cohort fixed effects. Columns (4) and (6) absorb high-school fixed effects. Controls include: male, URM, and age bins. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.



(a) *Target: 2021 (Train: 2020)* ***AUC = 0.640***



(b) *Target: 2022 (Train: 20-21)* ***AUC = 0.650***



(c) *Target: 2023 (Train: 20-22)* ***AUC = 0.640***

Figure A1: *Out-of-Sample Prediction: First-Year Retention (Post-COVID)*

Note: The predictive power of the HSGPA-only model dropped from its pre-COVID baseline (0.702) to ≈0.640 in the test-optional era and has not recovered. All models predict retention to the second year using probit regressions trained on the preceding cohort(s).