Which Colleges Increase Voting Rates?

By D'Wayne Bell, John B. Holbein, Samuel J. Imlay, and Jonathan Smith*

Using millions of SAT takers linked to college-enrollment and voting records, we document substantial variation in voter participation across otherwise similar colleges. We show that part of this variation is attributable to the colleges themselves. Leveraging family fixed effects and controlling for students' college application portfolios and pre-college voting behavior, we find that 98 percent of students attend a college that has positive value-added on voting relative to not enrolling. The standard deviation of these value-added estimates ranges from 3 to 5 percentage points (6–10 percent). These effects appear shortly after initial enrollment—as colleges create new voters—and persist after college.

I. Introduction

Mass participation in elections is vital to the health of electoral democracy (e.g., Downs, 1957; Dahl, 1973; Wolfinger and Rosenstone, 1980; Verba et al., 1995; Galston, 2004; Cascio and Washington, 2014; Gentzkow, 2006; Washington, 2006), and institutions of higher education may play an important role in the democratic process by shaping the lifelong voting habits of college-educated adults. In the United States, college-educated adults are 50 percent more likely to vote than those with only a high school diploma (Ahearn et al., 2023; Pender et al., 2026), and some of this disparity may be causal (e.g., Dee, 2004; Doyle and Skinner, 2017; Jensen, 2025; Apfeld et al., 2024a,b). Indeed, creating civically engaged citizens is a core mission of higher education and a central aim of public education policy (e.g., Colby et al., 2010; Holbein and Hillygus, 2020).[1]

While the relationship between voting and college enrollment, years of schooling, and degree attainment has been extensively studied (e.g., Dee, 2004; Tenn, 2007; Kam and Palmer, 2008; Sondheimer and Green, 2010; Berinsky and Lenz, 2011; Henderson and Chatfield, 2011; Mayer, 2011; Mettler, 2005; Jensen, 2025;

* Bell, Harvard University; Holbein, University of Virginia; Imlay, College Board; Smith, Georgia State University and IZA. This paper reflects the opinions of the authors and not that of their corresponding institutions or the providers of the data. We are grateful for helpful comments and assistance from Michael Barber, Jing Feng, David Deming, Tali Mendelberg, Joshua Goodman, Kirsten Mumma, Sreyartha Mukherjee, and participants at conferences and seminars at the Association for Education Finance and Policy, Southern Economic Association, Georgia State University, Workshop on the Economics of Education (Valle Nevado), the Southeastern Conference for Public Administration, the Midwest Political Science Association, Brigham Young University, Princeton University, and Aarhus University.

[1] We agree that voter turnout captures only one dimension of civic engagement. Other important components include political knowledge, democratic norms, political attitudes, and non-electoral forms of participation. We focus on turnout because it is a clearly defined, behaviorally revealed measure that can be observed consistently at scale and reliably linked to educational records across states and cohorts.

Apfeld et al., 2024a,b), scholars know less about how the *type* of college one attends shapes the voting decision (Mendelberg et al., 2021; Thomas et al., 2021; Firoozi, 2022). This is unfortunate, as American colleges vary widely in attributes thought to underpin their voting effects, from student composition and instructional quality to labor-market returns (e.g., Clotfelter, 2017; Hoxby, 2009). We ask the new question—how does the college one attends affect one's chances of voting?

We study this question by assembling one of the largest and richest datasets to date, linking students through high school and college to administrative voting records. We connect the universe of PSAT, SAT, and AP takers in the high school graduating cohorts of 2004 to 2012 with college enrollment data. The testing data include students' high school demographics, SAT scores, a proxy for their college applications, and sibling identifiers. We further link this education data to validated voting records indicating whether students voted in each general election between 2004 and 2016.

The millions of students in our data allow us to unearth voting rates at thousands of colleges across the U.S. by cohort and election. We document that colleges' voting rates are highly stable across cohorts and elections and vary systematically by institutional characteristics; relatively selective colleges, including R1 universities and liberal arts colleges, have the highest voting rates, while two-year colleges have the lowest. These institutional characteristics do not explain all the variation in college voting rates, however. Thus, two students considering academically similar colleges might end up at colleges with significantly different voting rates.

We estimate over 2,600 colleges' value-added on students' probability of voting in their mid- to late twenties. Although we do not use a quasi-experiment, our selection-on-observables approach makes use of a uniquely rich set of controls, in addition to control variables commonly used in studies of college choice and voter turnout. First, our preferred specification uses family fixed effects, which control for potentially important household-specific determinants of voting and college choice, such as parental influence and income (Plutzer, 2002). Second, we control for students' college application portfolios, following Dale and Krueger (2002) and others (e.g., Ge et al., 2022; Mountjoy and Hickman, 2021). In doing so, we leverage comparisons among students who apply to similar colleges, as measured by the average SAT scores and voting rates of their students, but ultimately enroll in different institutions. Third, in alternative specifications, we control for whether students voted in their first eligible election before entering college (i.e., pre-treatment). We do so by limiting our analytic sample to students who were old enough to vote in high school during an election year—the oldest 25 percent of students in odd-year graduation cohorts. This powerful control allows us to leverage comparisons between individuals who have similar voting histories even before college entry. Later in the paper, we conduct several analyses to assess the threat of unobservables, and each one leads to the same conclusion: unobservables

are unlikely to be driving our value-added estimates.[2]

Our 2,600 college value-added estimates show that most colleges increase their enrollees' probability of voting in the 2016 election relative to non-enrollees, and there is a meaningful distribution of college impacts. We find that 80 percent of colleges, which enroll 98 percent of students in our sample, have positive value-added estimates. The student-weighted mean of these college value-added estimates is about 9 percentage points (20 percent), and the standard deviation is between 3 and 5 percentage points, depending on the sample used, controls included, and whether we compute a raw or bias-corrected standard deviation. This standard deviation is 6 to 10 percent of sample students' voting rate in the 2016 election (46 percent). The value-added estimates are positively correlated with colleges' selectivity and unconditional voting rates and, to a lesser extent, the fraction of political- and social science majors and measures of campuses' participation in voter-friendly initiatives.

Next, we explore mechanisms, alternative outcomes, and heterogeneous effects by focusing on the voting effects of attending colleges with higher average SAT scores and voting rates–two college attributes strongly associated with higher value-added estimates. We uncover several new facts that help explain our value-added results. First, selective and high voting-rate colleges create new voters: their voting effects are concentrated among students who did not vote at their first opportunity, and they substantially increase their students' probability of registering to vote.[3] In other words, these colleges increase turnout by activating new voters rather than re-mobilizing previous voters. Second, attending high voting-rate colleges, more so than selective colleges, impacts voting in both presidential and midterm elections. The impacts in presidential elections are roughly twice as large, but so is the fraction of the sample who votes, so the percent impacts are similar. Third, the impacts of attending high voting-rate and selective colleges arise soon after initial enrollment and do not require 2-4 years of on-campus exposure. Fourth, the impacts also persist through the fifth (and last observed) election after high school but slightly fade out. Fifth, there are some differences in estimates across demographics and academic measures, but enrolling in high voting-rate and selective colleges increases turnout for most groups.

Our findings make broad contributions to three literatures. First, we add to a longstanding literature in political science, education, and economics that considers the role of education in individuals' voting habits (e.g., Dee, 2004; Doyle and Skinner, 2017; Wolfinger and Rosenstone, 1980; Sondheimer and Green, 2010; Smets and Van Ham, 2013; Verba et al., 1995; Willeck and Mendelberg, 2022).

[2]The strongest analysis is a placebo test conducted on the subsample of students who were eligible to vote while in high school: it assesses whether students who enroll in colleges with higher voting rates and average SAT scores—two college attributes associated with higher value-added estimates—voted at higher rates while in high school (i.e. pre-treatment). Reassuringly, we find no statistically significant relationship between these college attributes and students' pre-college voting.

[3]We also show more generally that colleges' value-added on their entrants' probability of voting is strongly correlated with their value-added on entrants' probability of voter registration.

This research estimates the impacts of additional years of schooling or educational attainment on voting, leveraging matching methods, distance instruments, changes in state laws, and admission cutoffs for identification. These studies generally find that an additional year of late-secondary schooling increases individuals' voting probability by 7-9 percentage points (Dee, 2004), while attending college increases turnout by 12-18 percentage points (Dee, 2004; Ahearn et al., 2023; Heckman et al., 2018), with incremental effects over each year of college (Doyle and Skinner, 2017) and an additional boost for college completion (Heckman et al., 2018).[4] We add to this literature by considering how education's voting effects vary *within* levels of educational attainment, demonstrating in particular that the voting effects of college attendance vary widely by institution type. Although it is not the focus of our paper, our analyses also produce estimates of the voting effect of college attendance (versus non-attendance) that are consistent with Dee (2004), despite using a very different sample and method.

Second, we add to the vast literature on the consequences of college choice (e.g., Long, 2004; Hoxby, 2007, 2009; Smith, 2013; Mountjoy and Hickman, 2021; Lovenheim and Smith, 2022). This literature focuses almost exclusively on the impacts of college choice on degree completion and earnings (Lovenheim and Smith, 2022). Although many authors have discussed the potential effects of college choice on non-pecuniary outcomes, to date research in this area has been primarily descriptive (e.g., Pender et al., 2026). Drawing methodologically from college choice studies that identify colleges' returns by matching students on application and/or admission portfolios (Dale and Krueger, 2002, 2011; Smith, 2013; Ge et al., 2022; Mountjoy and Hickman, 2021; Chetty et al., 2023), we estimate college voting effects that vary significantly by institution, showing how college choice affects one foundational civic behavior. This enterprise has great value given the foundational role that schools can, should, and (sometimes) do play in ensuring active participation in democracies.

Finally, our findings add to a growing body of scholarship that specifically considers how campus environments shape young adults' political behavior (e.g., Firoozi, 2022; Mendelberg et al., 2017, 2021). Descriptive research in this vein shows that undergraduates' voting rates vary systematically by institution type (Thomas et al., 2021) [5] while other studies find that attending certain types of colleges influences students' policy preferences and voting behavior (Firoozi, 2022; Mendelberg et al., 2017, 2021). We extend both strands of literature by recovering college voting rates—and college value-added on voting—across a much wider swath of higher education and shedding new light on institutional correlates

[4] Although some studies estimate small or null effects (e.g., Tenn, 2007; Kam and Palmer, 2008; Berinsky and Lenz, 2011; Persson, 2015), the bulk of this literature finds that college attendance increases voter turnout. This is especially true when one pools study results together (Jensen, 2025).

[5] The Institute for Democracy & Higher Education calculates student voting rates by college and selected student- and institution characteristics and shares these statistics with participating campuses. These data are not typically available to researchers and contain far less granular information than our dataset does.

and possible mechanisms.

The remainder of the paper proceeds as follows. The next section introduces our data sources and describes how we constructed our sample datasets. Section III defines college voting rates and describes how they vary by student and college characteristics. Section IV discusses our conceptual framework and empirical strategy. We present our main results in Section V. We discuss some additional results in Section VI, including robustness tests that assess the threat of unobservables, alternative outcomes, and heterogeneous effects. Section VII concludes.

II. Data, Matching, and Samples

A. Data

This study primarily uses three datasets: (1) testing data covering the universe of PSAT, SAT, and Advanced Placement (AP) takers in the high school graduation cohorts of 2004-2012; (2) college enrollment data from the National Student Clearinghouse (NSC); and (3) national voting records. We describe each in turn.

B. Testing Data

The base data for this study are individual-level administrative records for all PSAT, SAT, and AP takers from the 2004-2012 high school graduating cohorts (21.3 million students). The analytic sample is limited to SAT takers. The SAT is one of two college entrance exams considered in admissions and program placement by thousands of colleges across the U.S. Approximately 1.3 million students per cohort take the exam. The SAT is scored between 400 and 1600—200 to 800 its math and verbal sections. Upon registration, students complete a questionnaire that captures their name, date of birth, and demographic characteristics such as gender, race/ethnicity, and parental income/education. SAT data also capture which high school students attend and their SAT scores on all attempts.

The SAT data also include a record of students' SAT score sends, which indicate the set of colleges students are considering applying to and serve as a proxy for their college applications (Pallais, 2015; Smith, 2018). Score sends are official documentation of a student's SAT score frequently used in college admissions. When registering for the SAT (and just prior to receiving a score), students receive four free score sends. After students receive their SAT scores, each additional score send costs between $10 and $15, depending on the year. We observe up to 30 score sends per student.

The PSAT is an exam taken prior to the SAT. It qualifies students for scholarships and college outreach and is the qualifying exam for the National Merit Scholarship. It has broader reach than the SAT and is usually taken during the sophomore and/or junior year of high school. Students take AP exams at the end of a school year, typically after taking a corresponding AP course. Performing well on the exams can earn students college credit while still in high school.

The testing data also include a sibling identifier, which has been used in several other papers (e.g., Goodman et al., 2015; Gurantz et al., 2020; Altmejd et al., 2021).[6]

C. College Enrollment Data

National Student Clearinghouse data are a near-census of college enrollment spells in the U.S. They include enrollment dates for each student in each college for approximately 98 percent of all enrollees in the U.S.[7] Our analyses focus on the first college in which students enrolled, but the data follow students if they attend multiple colleges. NSC and SAT data are matched six years after students' high school graduation (e.g., NSC data on the 2006 high school graduation cohort records students' college enrollment history through the 2011-12 academic year).[8]

NSC data are also matched to the U.S. Department of Education's Integrated Postsecondary Education Data System (IPEDS), which provides college-by-year level institutional characteristics for nearly all colleges in the U.S. IPEDS contains some key variables in our analysis, such as location, measures of institutional selectivity, size, tuition, and the share of Pell grant recipients.

D. Voting Records

In the United States, each state collects and reports its own voting data, but all states publicly detail whether (but not for whom) each registered voter votes. We obtained nationwide voting data from the Data Trust, LLC—one of the many vendors in this space. The Data Trust combines and standardizes data from each state's election governing body on the tens of millions of people who vote in biannual national elections from 2004 to 2016.

The data record each individual's name, date of birth (DOB), state of residence, and voter-turnout status in various primary and general elections. Thus, for someone who was 18 years of age by November 2004, we observe their voting history between 2004 and 2016.[9] We collapse the voting records into a single observation per person—as opposed to a single observation per person–state—based on name and DOB. This process is detailed in Appendix B.

Our voter data are drawn from a snapshot compiled in 2017—i.e., the "close of books" snapshot from after the 2016 election—which aggregates historical registration and turnout information as reported by states. As a result, turnout in earlier elections (e.g., 2008 and 2010) is measured using this later snapshot. Because we do not have access to earlier snapshots of the voter file, participation in more distant elections may be somewhat undercounted due to list maintenance,

[6] Generally speaking, siblings are defined as students with the same last name, and home address. See Goodman et al. (2015) for more details and limitations.

[7] NSC data's biggest deficiency is for-profit colleges, despite including some of the largest ones.

[8] NSC data for the 2004 cohort tracks students for eight years.

[9] Voters can be removed from a state's records. Removals vary over time and by state but do not impact most voter records.

migration, or administrative updates that remove individuals from the file over time. It should be noted, however, that under the National Voter Registration Act's list-maintenance procedures, states generally cannot remove voters solely for inactivity. Instead, if the state suspects that a voter has moved, it must first send a confirmation notice; only if the voter fails to respond to that notice and does not vote repeatedly in the following federal general elections may the state remove them from the voter file.

To assess the importance of voter-file vintage, we conduct two robustness checks. First, we re-estimate our main specifications focusing on elections at varying points in time. Our most recent outcome, voting in the 2016 election, is our main outcome throughout the paper, but analyses of earlier elections yield qualitatively similar results. Second, we restrict the sample to states in which aggregate turnout rates in the voter file more closely match official state-level turnout statistics. In both cases, the results are qualitatively unchanged.

E. Matching

We match the education records to the voting records using students' names and dates of birth (DOB). We limit the dataset to individuals who live in the 50 states or D.C., so we remove international test takers, students from U.S. territories, and the very small number of students missing a DOB.

We start by matching unique name and DOB combinations from each dataset and then employ a series of fuzzy matching methods, as described in Appendix B. To indicate potential false matches (and non-matches), we also construct a series of indicators—flags for common names, partially missing DOB in the voting data, and females (who are more likely to change their last name)—which we later use in robustness tests. In a series of validation exercises in Appendix B, we show that the patterns of voting rates in our matched dataset follow those of nationally available statistics by age, race/ethnicity, and election.

Individuals who are not found in the voter registration records are coded as not having registered and not having voted. This approach follows common practice in studies that link administrative records to voter files, where unmatched individuals are typically treated as nonparticipants in electoral activity. Because voter files are constructed from lists of registered voters, individuals who do not appear in the records are most plausibly those who were never registered or who did not vote in the relevant elections. Coding unmatched individuals as nonparticipants therefore provides a consistent way to measure turnout across the full sample while avoiding the potential selection bias that would arise from restricting the analysis only to individuals successfully matched to voter records.

F. Samples and Summary Statistics

Table 1 summarizes the demographic characteristics, SAT performance, score-send portfolios, and voting histories of students in the full sample, the sibling

subsample, the subsample of students who are eligible to vote while still in high school, and the intersection of the last two subsamples.

The first two columns describe students in the fully matched dataset, which includes all PSAT, SAT, and AP takers in the 2004-2012 high school graduation cohorts who attended high schools in the 50 states and Washington, D.C. (20.9 million students). Again, students who did not match to the voting data remain in the sample (and are coded as non-voters). This is the broadest sample that covers the most students across the most colleges. We use this sample to calculate college voting rates. Students in this sample are 52.7 percent female and have an average math and verbal SAT score of 513 and 503, respectively; 21.2 percent of these students do not attend college (often our reference group), and 33.9 percent vote in the 2016 election.

The next two columns describe our primary analytic sample, consisting of almost 2.1 million siblings from the full sample who were members of the 2005-2012 high school graduation cohorts, took the SAT, and have a score-send portfolio that includes at least one four-year college.[10]. Compared to students in the full sample, these students are more likely to be white and to have a college-educated parent; they also have higher SAT scores, voting rates, and college-going rates.

The third sample in Table 1 is the "eligible-in-high-school" sample, which includes about 850,000 students. These students are domestic SAT score-senders who were eligible to vote in a general election held during their senior year of high school. This age condition includes only the oldest 25 percent of students who graduated high school in the spring of 2005, 2007, 2009, and 2011, but it gives us valuable information about students' pre-college voting behavior, which we use in a placebo test and as a key control variable in some analyses.

The final two columns describe the sample of siblings who were eligible to vote in high school (34,034 students). We use this sample in a placebo test of our preferred empirical strategy. This sample is limited to sets of siblings in the eligible-in-high-school sample; it is small because inclusion in this sample requires *both* siblings to be born relatively late in the calendar year and *both* siblings to be members of odd-year high school graduation cohorts. Apart from their age, students in this sample have similar characteristics to those in the broader sibling sample.

III. College Voting Rates

A. Constructing Voting Rates

Using the full sample of 21 million PSAT, SAT, and AP takers, we construct "college voting rates" by calculating the voting rate of each college's voting-age entrants by cohort and election. For example, we calculate the fraction of College A's entrants from the 2004 high school graduation cohort who voted in the 2004,

[10]More precisely, students must send their SAT scores to at least one college that reports its first-year enrollees' 25th and 75th percentile SAT scores to IPEDS. These are generally four-year colleges.

2006,..., and 2016 elections. That means each college's 2004 cohort has voting rates across seven elections.

We do the same for subsequent cohorts, but only for elections held after high school graduation.[11] For example, we calculate the fraction of College A's entrants from the 2006 cohort who voted in the 2006, 2008,..., and 2016 elections. Those six elections yield one fewer statistic than for the 2004 cohort. For the 2012 cohort, the last cohort we focus on, we only construct college voting rates in the 2012, 2014, and 2016 elections. Our main analyses focus on the 2016 election, which occurred after all cohorts had time to complete four years of college.[12]

Table 2 describes the mean of colleges' voting rates by cohort and presidential election. Data in each cell summarizes voting rates for over 2,600 colleges. It shows that mean college voting rates increased across the 2004-2016 presidential elections, with smaller differences across cohorts within each election. Most of our analyzes focus on the 2016 election, which has a mean college voting rate of around 35 percent across all cohorts.

B. Variation in College Voting Rates

In this subsection, we show that college voting rates vary substantially across institutions, even within the same cohort and election. We also show that this variation is correlated with several institutional characteristics (e.g. Carnegie Classification) but there remains meaningful variation in voting rates within institution type and among institutions with similar characteristics.

We first demonstrate the variation in college voting rates in Table 2. For each cohort and presidential election, we calculate the mean and standard deviation of voting rates for colleges. The standard deviations for the 2016 election are between 9 and 12.2 percentage points, depending on the cohort. This is approximately 25 to 33 percent of the magnitude of the means. This suggests that there is wide variation in colleges' voting rates.

The variation in colleges' voting rates can also be seen in Figure 1, which plots the variation in college voting rates by cohort and election. The ridge plots show a wide distribution of college voting rates within every cohort and election, frequently including voting rates as low as 10 percent and as high as 60 percent. There is also variation across cohorts and elections.

Colleges' voting rates also vary by institutional characteristics. Figure 2 shows bivariate correlations between several college characteristics and the college voting rates of the 2004 cohort in the 2016 election, when most students were 30 or 31 years old. For example, the top left panel shows a correlation of 0.53 between colleges' average SAT score and voting rates. On the other hand, the bottom

[11] Throughout all analyses, we also suppress voting rates for college-cohorts with fewer than 100 voting-age student observations.

[12] Since voting rates vary more by election than by age or cohort, we generally prefer to estimate voting probability in particular elections (pooling data across different-age cohorts) rather than pool across elections to estimate students' voting probability at a particular age.

right panel shows a negative correlation of -0.37 between the share of Pell grant recipient enrollees (a measure of financial need) and voting rates.

Table 3 shows that the bivariate relationships depicted in Figure 2 hold for a broader set of cohorts and conditional on a host of other college characteristics. Specifically, we regress college-cohorts' voting rates in 2016 on numerous college characteristics, with fixed effects for cohort and college state. Consistent with the bivariate relationships, several measures of college selectivity, including average SAT of enrollees, is positively correlated with the voting rate. Additionally, the R^2 of the regressions are relatively high, around 75 percent. An important contributor to the explanatory power are the state fixed effects, since states' turnout rates differ in predictable ways. However, the 25 percent of unexplained variation in voting rates is something we exploit in our main analyses.[13] This implies that two students who attend similar colleges on many of the dimensions of selectivity and quality used in previous research may still attend colleges whose entrants vote at very different rates in early adulthood.

Lastly, in Appendix Tables A1 and A2 and Appendix Figure A2, we show that colleges' relative voting rates are fairly stable across cohorts and elections. High and low voting-rate colleges remain as such regardless of the cohort, which suggests our results will not be sensitive to our choice of which cohort's college voting rate to use. It also suggests that there is not meaningful variation in voting behavior across cohorts to exploit.

IV. Methodology - College Value-Added

A. Conceptual Framework

A primary goal of this paper is to estimate colleges' valued-added on their students' probability of voting in an election. That is, how much does attending a specific college influence a student to vote over and above the student's preexisting likelihood of voting (or relative to attending a different college)? Our analysis is in the spirit of Mountjoy and Hickman (2021) who focus on bachelor's degree completion and earnings in Texas, whereas we focus on students' probability of voting across the entire U.S.[14] In this section, we develop a framework for distinguishing colleges' value-added on students' voting probability from students' preexisting propensity to vote.

Formally, let $V_{ik}(t)$ represent the propensity to vote of person i, who attends college k, in year t, as described below[15]:

[13] Appendix Figure A1 displays this variation by plotting the distribution of residuals from column 5 in Table 3. It shows that college voting rates vary meaningfully for every cohort, even after controlling for differences in basic institution characteristics.

[14] Mountjoy and Hickman (2021) build on a larger value-added literature for colleges (e.g., Cunha and Miller, 2014), teachers (e.g., Chetty et al., 2014), firms (e.g., Abowd et al., 1999), and beyond.

[15] For simplicity, assume an election and year are the same. We also assume linearity and additive separability, which are common in the value-added framework.

$$V_{ik}(t) = X_i(t) + Z_k(t) \tag{1}$$

The first argument, X_i, is an individual voting propensity that is endowed to people prior to college entry. It may reflect or correlate with sex, race, parental circumstances, high school experiences, etc. It is also a function of t because people with different attributes or endowments can have different propensities to vote as they age and in different elections.

Z_k is the college value-added.[16] In this context, we focus on experiences at the first college attended, but Z_k could include workplace experiences for those who do not enroll in college.

When students have any pre-dispositions to vote prior to enrolling in college, equation 1 highlights that the college voting rates examined in the previous descriptive section are not the same as college value-added. Rather, they are the average voting propensities across all individuals i who enroll in college k, or $V_k = \sum_i V_{ik}/n$. Moreover, if students' pre-dispositions vary by the college they attend, the voting rates are subject to selection effects that obfuscate the college value-added. The next subsection discusses how we empirically address these selection issues to identify colleges' value-added.

B. Empirical Strategy

Our primary objective is to recover unbiased estimates of Z_k for all $k \in K$ colleges. Empirically, this could be accomplished with a complete set of individual fixed effects (X_i) and college fixed effects (Z_k). Unfortunately, our data do not permit the estimation of individual fixed effects for several reasons; most salient among these is that the treatment of interest does not vary substantially within individuals over time. Instead, we include an array of student characteristics that approximates the individual fixed effect, including family fixed effects, college application portfolio controls, and measures of students' pre-disposition to vote prior to college entry.

We estimate the following equation to identify each college's value-added on students' probability of voting:

$$v_{ijke} = Z_k'\alpha + \delta_j + AppPort_i + \beta PastVote_i + x_{ij}'\gamma + \epsilon_{ijke} \tag{2}$$

v_{ijke} indicates whether student i from family j who initially enrolls in college k votes in election e. The vector $\alpha = [\alpha_1, ...\alpha_K]$ are the coefficients on the complete set of college fixed effects Z_k.

[16] We assume a common college value-added across individuals. We later test and find support for this assumption regarding basic demographics.

The next four terms approximate an individual fixed effect and are meant to capture students' pre-college propensity to vote in election e. The first term, δ_j, is a family fixed effect. This captures the common family factors that influence voting across siblings, which may include things like financial circumstances, parental voting propensity, geography, and schools. This control is particularly valuable given the formative role parents play in their children's political socialization (e.g., Gidengil et al., 2016).

The next term is student i's college application portfolio, $AppPort_i$. This is a useful control because students' college application portfolios contain valuable information about their academic achievement, college ambitions, and lifestyle interests not captured in standard demographic and academic controls. It may also contain information about whether students are interested in attending colleges that encourage voting and civic engagement.

$PastVote_i$ indicates whether student i voted in the election held in November of their senior year of high school. This variable is only available for, roughly, the oldest quarter of students in odd year cohorts. This is typically the first election in which these students are eligible to vote and, for college goers, the last election before college entry. Students' pre-college voting behavior is an intuitively appealing proxy for their pre-college propensity to vote in election e, and it predicts their later voting quite well: in 2016, when 39 percent of sample students voted, the 22 percent of students who voted in their first eligible election were more than twice as likely to vote (68 percent) as the 78 percent of students who did not (31 percent).

The last term associated with an individual's pre-college propensity to vote is x_{ij}. It includes demographics (sex and cohort) and academics (SAT math and verbal scores). In specifications that exclude family fixed effects, it also includes race/ethnicity, parental income and education, and high school fixed effects.

Lastly, ϵ_{ijke} is the error term.

C. Implementation

We rely on two different samples (and corresponding controls) to implement the above equation - the sibling sample and the sample of students eligible to vote in high school.

The primary advantages of this sample are the relatively large number of individuals and the ability to include a family fixed effect δ_j.

The sample of eligible voters in high school is smaller than the sibling sample but includes a potentially important variable $PastVote_i$, which measures whether the student voted while still in high school. The sample is smaller because only about one-eighth of all students in the data are eligible to vote in high school—the oldest quarter of students in odd-year cohorts.

In a few analyses, we use the intersection of both samples—siblings who are both eligible to vote in an election while still in high school. This sample is 1.6 percent and 4 percent the size of the other two samples, respectively, and

the standard errors are an order of magnitude larger, so we lose precision at the expense of a more compelling set of controls.

We follow the literature (e.g., Dale and Krueger, 2002; Mountjoy and Hickman, 2021; Ge et al., 2022) to implement the application portfolio control ($AppPort_i$) in both samples. Our preferred specifications control for relevant characteristics of students' application portfolios—specifically, the number of colleges in the portfolio and the minimum, mean, and maximum of their average freshman SAT scores. This implementation of the "self-revelation" approach accounts for key differences in students' academic achievement and college aspirations while employing an analytic sample that is generally representative of college applicants.[17]

Unlike previous studies that uses an application portfolio control approach, we also control for the minimum, mean, and maximum of the voting rates of the colleges in each student's application portfolio. This allows us to control for self-selection into colleges on dimensions unrelated to selectivity but potentially related to voting behavior.

Lastly, our main analyses consider voting in the 2016 election—the latest election we observe—which occurred after all sample students had time to complete four years of college. We also test the robustness of our results in other elections. In some analyses, we define elections ordinally relative to students' high school graduation year and consider how colleges' voting effects evolve across students' first post-high school election, second post-high school election, and so on.

D. Identification

Our goal is to identify the effect (i.e., value-added) of attending each college on entrants' probability of voting in their mid- to late twenties. The above regression relies on a selection-on-observables approach. In our preferred specification, we rely on variation within siblings and within application portfolios, along with more standard control variables.

The identifying assumption is the conditional independence assumption (CIA) whereby students' potential voting outcomes are independent of the college in which they enrolled, conditional on a set of observed controls. In other words, the error term ϵ_{ijke} is assumed to be independent of Z_k. Less formally, we assume there are no additional unobservables that are correlated with both the college that students attend and their probability of voting.

The conditional independence assumption is largely untestable, but our approach has a uniquely rich set of control variables and a few available tests that support the assumption.

As to the rich set of controls, one lingering identification concern is that we do not control for college admissions portfolio, as in Mountjoy and Hickman (2021), but rather, the college application portfolio. This means there may be

[17] By contrast, implementing application-portfolio fixed effects would dramatically reduce the size and representativeness of the student sample because many students have unique application portfolios.

something unobservable that admissions committees observe that is correlated with the error term. Mountjoy and Hickman (2021) call this an "insufficient specification." However, Smith (2013) finds that controlling for application or admission portfolios yields similar results with a national sample, as we do in this paper. We also control for a series of academic controls, such as SAT scores and high school attended, that are generally the primary determinant of admissions outside the most selective colleges. Additionally, we control for the college voting rates in the portfolio, which helps account for students making decisions directly on those types of non-academic factors.

We also control for the family fixed effect, which is commonly unavailable in college value-added research. Parents, home life, and upbringing are demonstrably correlated with voting habits (and other outcomes in adulthood), and most researchers can only control for some basic parental information. Our identification compares two siblings who chose to enroll in different colleges despite having very similar upbringings. The refined threat to identification then becomes whether one *sibling* chose a college for reasons correlated with the probability of voting compared to her sibling, as opposed to one *student* compared to another student. This is still indeed a threat, but arguably less so. Fortunately, we still have a rich set of controls, including the application portfolio and past voting rates, to account for this potential threat. And this is the same control that others (somewhat convincingly) relied on without family fixed effects.

On the few available tests of the conditional independence assumption, we assess the threat of unobservables in three ways: (1) we implement a version of a covariate balance test that assesses whether students' pre-college predicted probability of voting is correlated with where they enroll, conditional on our "key" controls; (2) we assess the stability of the estimates as we add more controls, including using an Oster bounding exercise; and (3) using the subsample of students eligible to vote in high school, we implement a placebo test that assesses whether students' voting behavior in high school remains correlated with the college they attended conditional on our preferred set of controls. Collectively, the results of these tests suggest that unobservables are unlikely to be driving our estimates. We describe these tests and their results in further detail below.

Bayesian Approach to the College Value-Added Estimates

The value-added literature often adopts Bayesian techniques to improve the set of value-added estimates by "borrowing strength" from other similar colleges via the estimated prior (Morris, 1983). This may be particularly important in our context for colleges with relatively few students and consequently introducing measurement and sampling error. Numerous examples are described in Walters (2024), whose techniques we follow. By doing so, we shrink the noisy estimates $\hat{\alpha}_k$ towards the mean of those estimates in proportion to its signal-to-noise ratio.

To compute the bias-corrected variance, we take each of the college value added estimates, measured by $\hat{\alpha}_k$, calculate the variance of the estimates, and subtract

the average squared standard error (then take the square root). Mathematically, this can be expressed as:

$$\hat{\mu_\alpha} = \frac{1}{K}\sum_{k=1}^{K}\hat{\alpha}_k, \tag{3}$$

$$\hat{\sigma}_\alpha^2 = \frac{1}{K}\sum_{k=1}^{K}\left[(\hat{\alpha}_k - \hat{\mu}_\alpha)^2 - s_k^2\right]. \tag{4}$$

s_k^2 is the bias-correction that accounts for excess variability of $\hat{\alpha}_k$ and is calculated as the average of the squared standard errors associated with the college fixed effect coefficients. In practice, we deviate from the standard correction above by calculating a non-parametric bias-corrected standard deviation in case there are correlations in noise across colleges, especially when we include the high-dimensional family fixed effects. The details and equations are described in Appendix D.

V. College Value-Added Results

A. Assessing the Method - Covariate Balance

In this subsection, we assess whether there are potentially unobserved variables that could bias our college value-added estimates. (We further examine the potential for bias in our models in the "Testing for Unobservables" subsection below.) To do so, we estimate a modified balance test, following Mountjoy and Hickman (2021). Our test assesses whether an individual's predicted propensity to vote is correlated with where they enroll in college, accounting for some "key controls," but not all of our controls. We take the following steps:

1) We first predict each individual's probability of voting in the 2016 election, using only demographics and high school achievement.

2) We regress the predicted probability on the set of college fixed effects (Z_k), conditional on our "key controls" (similar to equation 2).

3) We estimate the standard deviation of the distribution of fixed effects, where each fixed effect is weighted by the number of students enrolled in the college.

A degenerate distribution implies that there is no evidence that students with a relatively high or low probability of voting sort into different colleges, conditional on our controls. This is, admittedly, a high bar to clear because even in the absence of sorting, we would expect some variation in the estimated fixed effects

due to sampling noise alone. As such, meaningfully decreasing the standard deviations observed across model specifications suggests that the potential for selection on unobservables is diminished.

Additionally and importantly, the controls included in this balance test are not the full set used in our main models because some variables are used to construct the predicted probability of voting. This implies that a non-degenerate distribution does not necessarily mean the estimates are biased; rather, this provides an upper bound on the potential bias arising from residual sorting on observables.

The first set estimates of the standard deviation of predicted voting rates across colleges is shown in Table 4. The first column shows the standard deviation of the college fixed effects estimate when there are no controls, but for cohort fixed effects. The 8.1 percentage point standard deviation is large and statistically different from zero. The bias-corrected standard deviation is similarly sized (7.7 percentage points) and is also statistically significant. This confirms that students with relatively low or high predicted probabilities of voting, based entirely on some of their high school attributes, sort into different colleges. This also suggests that we need a strong set of controls to remove any bias.

The second column adds family fixed effects, which are one of our primary controls, and the standard deviation drops to 2.5 percentage points in the first row and 1.6 percentage points in the second bias-corrected row. Adding portfolio summary statistics, as we do in column 3, reduces the standard deviation even further, to 2.2 percentage points in the first row and now only 1.2 percentage points in the bias-corrected row. Our primary controls attenuate the standard deviation by 73 to 84 percent in the two rows, respectively. This is not a perfectly degenerate distribution; however, the remaining dispersion is small in magnitude. Additionally, it is somewhat unreasonable to expect a perfectly degenerate distribution with over 2,600 college fixed effect estimates; some of these are likely to be statistically different than zero by chance. With that many colleges, we expect over 130 of the college fixed effects to differ from zero by chance, at the 5 percent level.

There may remain potential for bias in our design; however, the potential for bias is severely reduced by our preferred controls.[18] Moreover, the fact that this is a hard/conservative test—for the reasons we note above—provides some reassurance. This is an upper bound estimate of the bias in our design. Finally, we note that later—in the "Testing for Unobservables" section—we further assess the potential extent of bias from lingering unobservables through a bounding exercise and a placebo test. These provide us with additional assurance that bias is largely mitigated by our preferred controls.

[18] Appendix Table A3 shows the analogous results but without student-enrollment weights on the college value-added estimates. We find similar results—with some evidence that our standard deviations are, in some specifications, even smaller in magnitude relative to the unweighted results, suggesting that our earlier results may be somewhat conservative.

B. Distribution of Value-Added Across Colleges

The histogram in Figure 3 shows the distribution of the college value-added estimates from equation 2, using the sibling sample, weighting the estimates by student enrollment, and normalizing estimates relative to non-college goers. The histogram shows that the college value-added estimates are positively skewed and have a significant amount of variation. Quite a few colleges have negative estimates relative to non-college goers, while the mean is closer to 10 percentage points. There are also many colleges with value-added estimates of more than 20 percentage points.

The solid line in Figure 3 shows the deconvolved distribution, layered on top of the original histogram of the value-added estimates. Visually, the deconvolved distribution has a lower mean, a smaller standard deviation, and support largely between -5 and 20 percentage points.

Next, we formally assess whether the distribution in Figure 3 has, as it appears, a distribution that is economically and statistically meaningful. Formally, we estimate the standard deviation of the distribution of college value-added estimates. If all colleges have the same impact on voting (or no impact), then we have a degenerate distribution with a standard deviation that is not statistically different from zero. This outcome is plausible, even after seeing Figure 3, if each college's estimate is particularly noisy.

The first row of Table 5 shows the estimates of the standard deviation of the college value-added distribution. The first column shows the standard deviation of the over 2,600 college value-added estimates from the sibling sample when we include no controls, but for a cohort fixed effect. The standard deviation is 9.3, which indicates that the unconditional standard deviation of college voting rates is 9.3 percentage points. The second column shows the distribution of estimates after adding one of our primary controls: family fixed effects. The standard deviation of the distribution drops precipitously to 5 percentage points. Adding additional student controls in column 3, including application portfolio summary statistics, further reduces the standard deviation to 4.2 percentage points.

The standard deviation is statistically different from zero and meaningful in size. Assuming a normal distribution, the estimate implies that attending a college at the 10th versus 90th percentiles of the distribution of value-added would increase a student's probability of voting by 10.8 percentage points ($2 * 1.28 * \hat{\theta}$), which is about a 30 percent increase relative to the sample's voting rate.

Columns 4 and 5 show the analogous estimates of the standard deviation of the college value added distribution using the eligible voters in the high school sample. The unconditional standard deviation in column three is about 8.9 percentage points, which is similar to the sibling sample. However, the model with a full set of controls, including whether they voted while in high school, yields a 4.5 percentage point standard deviation. The estimate implies that attending a college at the 10th versus 90th percentiles of the distribution of value-added would increase a student's probability of voting by 11.5 percentage points, which is also about a

30 percent increase relative to the sample's voting rate.

The main estimates of the deconvolved standard deviations are 2.7 percentage points (column 3) and 3.7 percentage points (column 5) in Table 5's second row, which correspond to the siblings and eligible-to-vote in high school samples, respectively. These estimates are attenuated relative to the uncorrected estimates but reflect distributions that are statistically different from zero and not degenerate.

Appendix Figure A3 and Appendix Table A4 shows the standard deviation of the distribution of college value-added estimates without weighting by student enrollment, graphically and in table form, respectively. Graphically and in the table, the estimated standard deviation is larger than the unweighted estimates but is qualitatively unchanged.

Finally, we re-estimate these distributions of college value-added with an alternative outcome - whether the individual has registered to vote. Similarly, we find evidence that colleges increase the probability of registering to vote and there is a non-degenerate and meaningful distribution of college vale-added estimates. The histogram of value-added estimates and the deconvolved distribution can be seen in Appendix Figure A4 (and weighted by college-enrollment in Appendix Figure A5). The corresponding estimates of the standard deviations of the distribution of college value-added estimates, which is 3.4-4.4 percentage points, can be found in Appendix Table A5. These estimates are slightly smaller than the estimates of the distribution when voting is the outcome.

C. Correlates with the Value-Added Estimates

Table 6 shows the correlates with the college value-added estimates. The simple Pearson correlations, weighted by student enrollment, descriptively inform us of which characteristics of a college are associated with relatively large value-added estimates.

Among the Carnegie Classifications, research universities (R1s) have the strongest correlation, followed by liberal arts colleges. Associate's institutions (or two-year colleges) have a negative correlation with the value-added estimates.

The next three correlations, which are all positive, are markers of the campus environment for voters. NSLVE are campuses that participate in a national study on voting and civic engagement, and ALL IN and "Voter-Friendly Campus" are independent initiatives that track and certify efforts to create a voting campus.

Next, we find that the correlation with the college value-added estimates of registering to vote is 0.74. The high correlation is also displayed in a scatter plot in Appendix Figure A6. Although this is only suggestive, the strong correlation is consistent with colleges influencing students to register to vote (i.e., creating new potential voters) as a key mechanism. We later show additional evidence to support this mechanism.

Our measure of whether the campus has a polling station comes from the 2012 election, but we see no relationship with value-added estimates. Conversely, the

fraction of degrees in social science and political science is strongly correlated with value-added estimates.

Two types of variables are highly correlated with college value-added: college voting rates and the colleges' selectivity. These two variables are also of interest to the college choice literature. Colleges' average SAT is a common measure used by researchers to estimate the return to selectivity (e.g., Dale and Krueger, 2002; Smith, 2013). It is also a metric that students use to choose colleges. On the opposite extreme, college voting rates are a metric we introduce into the college choice literature that has a direct connection to college value-added in relation to the probability of voting.

In the next section, we let these two variables serve as the treatment variables in equation 2, in exchange for the high-dimensional college fixed effects. Moving away from the value-added framework allows us to do several things. First, we can test for several mechanisms, some of which rely on additional outcomes. Second, we can reassess our identification strategy and consider the lingering threat of unobservables. Third, we can easily conduct some heterogeneous effect analyses to examine where the results are strongest and some potential mechanisms.

VI. Additional Results

A. College Selectivity or College Voting Rates?

In this section, we re-estimate equation 2 but use college voting rates and the average SAT of the college as the two treatment variables instead of the set of college fixed effects. These are two variables we found to be strongly correlated with our college value-added estimates and are directly connected to the extant literature and research questions in this paper. Table 7 presents the results.

The first column shows the relationship between college voting rates and the probability of voting in 2016, the college average SAT, and the probability of voting in 2016, with no additional controls in the sibling sample. Attending a college with a 10 percentage point higher voting rate than the alternative is associated with a 10 percentage point increase in the probability of voting. This is largely a tautology, but for the fact that we also control for college average SAT. Attending a college where students average SAT scores are 100 points higher than the alternative is associated with a 2.1 percentage point increase in the probability of voting.

Column 2 shows that both college attributes are independently positively predictive of a student's probability of voting, even after adding a full set of controls, including family fixed effects. The coefficients are roughly one-quarter the size of the previous column's estimates but are still statistically different from zero. It also suggests that the college voting rate is a relatively strong predictor of the probability of voting, over and above the returns to college selectivity. The latter has been the focus of the college choice literature, but we show that the returns may differ when we consider outcomes other than earnings.

Estimates are remarkably similar in columns three and four, which are analogous to the two previous columns but use the eligible to vote in high school sample and controlling for whether they voted in high school. The last two columns are also analogous but use the intersection of the two samples—siblings who were all eligible to vote in high school. The coefficient on college voting rate increases in magnitude, while the coefficient on average SAT decreases in magnitude and loses statistical significance. This may be in part due to the smaller sample and larger standard errors.

Testing for Unobservables

We run three tests that assess the extent of unobservables: (1) a placebo/balance test, (2) a coefficient stability test, and (3) a bounding exercise.

We first run a placebo/balance test that has reassuring results in Table 8. We use our small subsample of siblings who were both eligible to vote in high school and test whether high school voting (the outcome) is correlated with college voting rates and college selectivity, controlling for all other variables (other than voting in high school). The first column reproduces the last column of Table 7, but without the control of whether they voted in high school. The results are very similar to before—the college voting rate is positively related to the probability of voting in 2016, while the college average SAT is statistically zero.

Column 2 removes the controls and changes the outcomes to the first eligible election, which, for this sample, is while they are still in high school (i.e., pre-treatment). In this specification, we find a positive relationship between college voting rates and the probability of voting in the first eligible election. There is also a positive relationship between the average SAT score of the college that the student will eventually attend and the probability of voting.

However, column 3—the strongest placebo test—shows no statistical evidence of these relationships after we control for family fixed effects and application portfolios. This provides evidence that sorting into college types is unrelated to one's propensity to vote, as measured by voting in high school. This reduces the concern that lingering unobservables are driving the results in Table 7 and, by extension, the estimates of the distribution of college value-added estimates in Table 5.

Second, in Appendix Tables A6 and A7, we demonstrate the stability of our estimates as we added increasingly more controls and as the R-squared increases dramatically. This provides evidence that we have likely controlled for many of the unobservable factors correlated with these two college attributes that may lead to omitted variable bias. Table A6 uses the sibling sample and starts with the unconditional regression from Table 7 (column 1). We then add in family fixed effects - our primary control - and the coefficients on college voting rates drops dramatically and, to a lesser extent, so does the coefficient on college selectivity. The R-squared jumps from 0.025 to 0.556. The next two columns add college application portfolio controls and student characteristics, respectively, and the co-

efficient on college voting rate is relatively stable. A similar exercise is completed with the eligible-to-vote in high school sample, contained in Appendix Table A7.

Third, the Oster Bounding test in Appendix Table A8 suggests that if the remaining unobservables are roughly as large as our observables ($\delta = 1$), we would still have positive but attenuated coefficients on college voting rates. But it is highly unlikely, if not impossible, that remaining unobservables are as large as the observables, given our set of controls that includes family fixed effects. Prior work finds that parents/home circumstances and geography, which are both captured in those fixed effects, are among the largest determinants of whether people vote. This is clear in our R-squared of 0.557. Although we cannot entirely rule out remaining unobservables that systematically influence voting patterns, they are almost certainly dwarfed by family fixed effects in addition to our other control variables.

Robustness Tests

Next, we show that our results are robust to alternative samples and specifications. In Appendix Table A9, we confirm that our result holds among college goers only and in subpopulations that may experience different rates of error in matching to the voting data, such as exact name/date of birth matches and high coverage states. Similarly, we show that results are similar across sex even though women are more likely to change their last name and make matching more difficult.

In Appendix Table A10 we address concerns about selection on unobserved correlates of voting propensity by implementing our main specification in subsamples with different college-going probabilities and destinations. We find similar results even among students with low probabilities of attending a four-year college and among those who did not attend a selective four-year college—students whose college choices are generally not driven by admissions processes that might select students on unobserved correlates of voting propensity.

B. Registration and Election Type, Year, and Order

In this subsection, we consider three questions: (1) does attending a high voting-rate college or selective college impact the probability of registering to vote; (2) does the election year and type matter?; and (3) does the impact persist, fade, or increase over elections?

The first column of Table 9 shows that attending high voting-rate colleges and more selective colleges increases the probability of registering to vote - a precondition to casting a vote. Attending a college with a 10 percentage point higher voting rate increases the probability of registering to vote by 2.7 percentage points, which is notably similar to the coefficient on the probability of voting in previous results (2.6 percentage points). Similarly, attending a college with relatively high average SAT scores increases the probability of registering to vote

and voting in 2016 by similar magnitudes. These results suggest that colleges are not creating registered voters who do not vote, rather, they are creating new voters. This is reaffirmed in the next subsection.

The next four columns of Table 9 show the results for four different elections—2010, 2012, 2014, and 2016. 2012 and 2016 are presidential elections during which aggregate turnout is higher, while 2010 and 2014 are non-presidential elections. Our estimates show that attending a college with a 10 percentage point higher college voting rate than the alternative leads to a 4 percentage point increase in the probability of voting in a presidential election. This represents more than a 10 percent increase over the mean voting rate. Alternatively, the college voting rate estimate for non-presidential elections is roughly half the size of that for presidential elections, as is the mean voting rate, leading to a similarly sized effect in percentage terms. Across each column, we see small in magnitude estimates on college average SATs scores that are often not statistically different from zero.

Next, Table 10 shows that the impacts of attending a college with a high voting-rate arise soon after enrolling on campus and persist for several elections. This is accomplished by changing the outcomes to be students' first through fifth elections after high school. Which election this relates to depends on the students' cohorts. We rely only on the 2005-2008 sample of siblings so we can keep a consistent sample that has the opportunity to vote at least five times. All coefficients on college voting rates are positive and statistically different from zero. The first election is the largest in magnitude (2.7 percentage points), and the fifth election is the smallest in magnitude (1.5 percentage points). These suggest that colleges with high voting-rates have a lasting impact on the probability of voting—it fades to some extent but does not fully go away within the timeline we can examine. Conversely, college selectivity has no consistent impact on voting across elections.

C. New Voters and Heterogeneous Effects

We begin our analysis of heterogeneous effects by asking whether our impacts are driven by new or persistent voters. That is, does attending a high voting-rate college activate new voters or retain voters who already voted in high school?

Table 11 shows that attending colleges with high voting-rates and more selective colleges creates new voters and does not change the likelihood of voting for existing voters. In the first column, which uses the sibling sample, the coefficient on college voting rate is 2.9 percentage points, and the interaction term with whether they voted in high school is -4.2 percentage points. The sum of the two coefficients, which measure the impact of attending a high voting-rate college for those who voted while in high school, is not statistically different from zero. This is affirmed in the next two columns, which use the eligible to vote in high school sample, and reaffirmed in the last column, which is the intersection of the two samples. Similarly, but smaller in magnitude, any relationship between college average SAT and the probability of voting is counteracted for pre-existing voters.

Overall, these results indicate that the impacts of attending a college with a high voting-rate are driven by activating new voters. This aligns with prior research showing that voting behavior is persistent (Coppock and Green, 2016), so we might not expect people who voted at the first opportunity to change their habits much.

Next, we estimate heterogeneous effects by student characteristics. The results are in Appendix Table A11, where we focus on attributes in which there are known voting disparities (e.g., demographics and SAT scores), but also on some post-matriculation behaviors (e.g., out-of-state status and graduation status). There are some differences in impacts by subgroup, but most students are positively impacted by attending high voting-rate and selective colleges, some more so than others. Most notably, we find that females are impacted more than males by attending a high voting-rate college, which reinforces or contributes to relatively higher voting rates among females. Similarly, Black students are less impacted than White students by attending a high voting-rate college, which reinforces or contributes to overall lower voting rates among Black students. We also find that college graduates have somewhat larger estimates than non-graduates, but non-graduates are still impacted by colleges with high voting-rates and selectivity. This suggests that the longer-term voting outcomes are likely not entirely driven by changes to earnings because these non-graduates have fewer opportunities. Rather, the college itself is impacting students. Other heterogeneous effect estimates are either small in magnitude or zero.

VII. Conclusion

In this paper, we have studied the voting effects of American colleges using administrative data covering nearly 21 million students. Across more than 2,600 institutions, we have shown that college goers' voting rates vary widely by college, and some of this is attributable to the colleges themselves. In fact, most colleges have a positive value-added to students' probability of voting in their mid- to late twenties relative to non-enrollees. Moreover, attending relatively higher-voting-rate colleges and more selective colleges meaningfully increases the chances of creating new voters with persistent effects.

These findings have a handful of important implications. First, our results support the interpretation that colleges influence not only labor-market outcomes but also civic behaviors such as voter turnout. Drawing on a much larger sample of validated student and voter records than has been used in prior work, our estimates provide strong evidence that colleges play a role in making young voters. This is consequential because scholars and democratic theorists have long argued that schools can—and should—help cultivate active participation in democratic life.

Second, beyond simply whether students attend college, the particular college they attend appears to matter. Although most colleges increase students' probability of voting relative to non-enrollees, there is substantial heterogeneity in the

value-added that colleges provide with respect to turnout. The distribution of college value-added estimates is sizable, ranging from roughly 6–10 percent of the mean voting rate in the sample.

Third, our finding that colleges' voting effects emerge early in students' time on campus and are present even among non-graduates suggests that these turnout effects arise at least in part from students' experiences during college itself, rather than solely from colleges' effects on post-college outcomes (e.g., labor-market returns). What specific features of the college experience drive these increases in voting—particularly in the early years of enrollment—remain an important question for future research.

Fourth, differences in colleges' voting effects could matter politically because colleges' effects on voting are not uniformly distributed throughout the college-going population. In particular, we show that the college choice process sorts higher-SAT students into colleges with larger voting effects, modestly amplifying initial disparities in voting propensities based on academic achievement.

Sixth, our results underscore the importance of institutional environments in shaping political participation. Much of the research on turnout focuses on individual-level resources and motivations, yet our findings suggest that organizational contexts—here, colleges—can meaningfully shape civic behavior. This suggests that the environments young adults inhabit during formative years can have lasting consequences for democratic participation. Further, the persistence of these effects into individuals' mid to late twenties suggests that college environments may contribute to the formation of durable voting habits. If colleges activate students who might otherwise miss their first opportunity to vote, they may help set individuals onto a trajectory of sustained political participation.

Seventh, our findings suggest that conventional metrics used to evaluate colleges—such as test scores or institutional selectivity—capture only part of what makes a college impactful. Although more selective colleges tend to produce somewhat higher voting rates, we show that campus voting environments independently predict students' likelihood of voting, even after accounting for selectivity and rich controls, including family fixed effects. This implies that aspects of colleges not captured by traditional academic metrics—such as institutional norms, civic programming, administrative practices, and/or opportunities for political engagement—may meaningfully shape students' civic participation. As a result, evaluations of college quality that focus narrowly on academic credentials or labor-market returns may overlook an important dimension of colleges' societal contributions: their role in cultivating democratic participation.

Finally, our findings have policy implications for those seeking to increase young adults' turnout. Our results suggest that colleges are creating new voters, but some are doing so more than others; this matters given the large baseline gaps in participation rates. However, our paper cannot speak to why or how this happens. Determining which practices and policies work for colleges may benefit other colleges that have not instituted those ideas, but some practices may translate to

settings outside of campuses. More broadly, our results show that colleges play a formative role in their students' voting habits. This suggests that institutions of higher education, which enroll nearly 70 percent of 18-19-year-olds, can play an important part in efforts to boost young adults' turnout.

REFERENCES

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Ahearn, C. E., Brand, J. E., and Zhou, X. (2023). How, and for whom, does higher education increase voting? *Research in Higher Education*, 64(4):574–597.

Altmejd, A., Barrios-Fernández, A., Drlje, M., Goodman, J., Hurwitz, M., Kovac, D., Mulhern, C., Neilson, C., and Smith, J. (2021). O brother, where start thou? sibling spillovers on college and major choice in four countries. *The Quarterly Journal of Economics*, 136(3):1831–1886.

Apfeld, B., Coman, E., Gerring, J., and Jessee, S. (2024a). The impact of university attendance on partisanship. *Political Science Research and Methods*, 12(1):45–58.

Apfeld, B., Coman, E., Gerring, J., and Jessee, S. (2024b). *A liberal education: The social and political impact of the modern university*. Cambridge University Press.

Bell, D., Feng, J., Holbein, J. B., and Smith, J. (2024). Do stem students vote? 12. *American Educational Research Journal*, 61(1):48–73.

Berinsky, A. J. and Lenz, G. S. (2011). Education and political participation: Exploring the causal link. *Political Behavior*, 33:357–373.

Cascio, E. U. and Washington, E. (2014). Valuing the vote: The redistribution of voting rights and state funds following the voting rights act of 1965. *The Quarterly Journal of Economics*, 129(1):379–433.

Chetty, R., Deming, D. J., and Friedman, J. N. (2023). Diversifying society's leaders? the causal effects of admission to highly selective private colleges. Technical report, National Bureau of Economic Research.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American economic review*, 104(9):2633–2679.

Chetty, R., Friedman, J. N., Saez, E., Turner, N., and Yagan, D. (2017). Mobility report cards: The role of colleges in intergenerational mobility. Technical report, national bureau of economic research.

Clotfelter, C. T. (2017). *Unequal colleges in the age of disparity.* Harvard University Press.

Colby, A., Beaumont, E., Ehrlich, T., and Corngold, J. (2010). *Educating for democracy: Preparing undergraduates for responsible political engagement*, volume 19. John Wiley & Sons.

Coppock, A. and Green, D. P. (2016). Is voting habit forming? new evidence from experiments and regression discontinuities. *American Journal of Political Science*, 60(4):1044–1062.

Cunha, J. M. and Miller, T. (2014). Measuring value-added in higher education: Possibilities and limitations in the use of administrative data. *Economics of Education Review*, 42:64–77.

Dahl, R. A. (1973). *Polyarchy: Participation and Opposition.* Yale University Press.

Dale, S. and Krueger, A. B. (2011). Estimating the return to college selectivity over the career using administrative earnings data. Technical report, National Bureau of Economic Research.

Dale, S. B. and Krueger, A. B. (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *The Quarterly Journal of Economics*, 117(4):1491–1527.

Dee, T. S. (2004). Are there civic returns to education? *Journal of public economics*, 88(9-10):1697–1720.

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of political economy*, 65(2):135–150.

Doyle, W. R. and Skinner, B. T. (2017). Does postsecondary education result in civic benefits? *The Journal of Higher Education*, 88(6):863–893.

Firoozi, D. (2022). The effect of research universities on student partisanship and turnout.

Galston, W. A. (2004). Civic education and political participation. *PS: Political Science and Politics*, 37(2):263–266.

Ge, S., Isaac, E., and Miller, A. (2022). Elite schools and opting in: Effects of college selectivity on career and family outcomes. *Journal of Labor Economics*, 40(S1):S383–S427.

Gentzkow, M. (2006). Television and voter turnout. *The Quarterly Journal of Economics*, 121(3):931–972.

Gidengil, E., Wass, H., and Valaste, M. (2016). Political socialization and voting: The parent–child link in turnout. *Political Research Quarterly*, 69(2):373–383.

Goodman, J., Hurwitz, M., Smith, J., and Fox, J. (2015). The relationship between siblings' college choices: Evidence from one million sat-taking families. *Economics of Education Review*, 48:75–85.

Gurantz, O., Hurwitz, M., and Smith, J. (2020). Sibling effects on high school exam taking and performance. *Journal of Economic Behavior & Organization*, 178:534–549.

Heckman, J. J., Humphries, J. E., and Veramendi, G. (2018). The nonmarket benefits of education and ability. *Journal of human capital*, 12(2):282–304.

Henderson, J. and Chatfield, S. (2011). Who matches? propensity scores and bias in the causal effects of education on participation. *The Journal of Politics*, 73(3):646–658.

Holbein, J. B. and Hillygus, D. S. (2020). *Making Young Voters: Converting Civic Attitudes Into Civic Action*. Cambridge University Press.

Hoxby, C. M. (2007). *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*. University of Chicago Press.

Hoxby, C. M. (2009). The changing selectivity of american colleges. *Journal of Economic perspectives*, 23(4):95–118.

Jensen, A. V. (2025). Educating for democracy? going to college increases political participation. *British Journal of Political Science*, 55:e1.

Kam, C. D. and Palmer, C. L. (2008). Reconsidering the effects of education on political participation. *The Journal of Politics*, 70(3):612–631.

Long, B. T. (2004). How have college decisions changed over time? an application of the conditional logistic choice model. *Journal of econometrics*, 121(1-2):271–296.

Lovenheim, M. F. and Smith, J. (2022). Returns to different postsecondary investments: Institution type, academic programs, and credentials. Technical report, National Bureau of Economic Research.

Mayer, A. K. (2011). Does education increase political participation? *The Journal of Politics*, 73(3):633–645.

Mendelberg, T., McCabe, K. T., and Thal, A. (2017). College socialization and the economic views of affluent americans. *American Journal of Political Science*, 61(3):606–623.

Mendelberg, T., Mérola, V., Raychaudhuri, T., and Thal, A. (2021). When poor students attend rich schools: Do affluent social environments increase or decrease participation? *Perspectives on Politics*, 19(3):807–823.

Mettler, S. (2005). *Soldiers to citizens: The GI Bill and the making of the greatest generation*. Oxford University Press on Demand.

Morris, C. N. (1983). Parametric empirical bayes inference: theory and applications. *Journal of the American statistical Association*, 78(381):47–55.

Mountjoy, J. and Hickman, B. R. (2021). The returns to college (s): Relative value-added and match effects in higher education. Technical report, National Bureau of Economic Research.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Pallais, A. (2015). Small differences that matter: Mistakes in applying to college. *Journal of Labor Economics*, 33(2):493–520.

Pender, M., Ma, J., Hu, X., and Edwards, A. (2026). Education pays 2026: The benefits of higher education for individuals and society. trends in higher education series. *College Board*.

Persson, M. (2015). Education and political participation. *British Journal of Political Science*, 45(3):689–703.

Plutzer, E. (2002). Becoming a habitual voter: Inertia, resources, and growth in young adulthood. *American political science review*, 96(1):41–56.

Rowell, J. and Kulkarni, S. (2015). Beyond college rankings: A value-added approach to assessing two-and four-year schools. *Metropolitan Policy Program at Brookings Institute*.

Smets, K. and Van Ham, C. (2013). The embarrassment of riches? a meta-analysis of individual-level research on voter turnout. *Electoral studies*, 32(2):344–359.

Smith, J. (2013). Ova and out: Using twins to estimate the educational returns to attending a selective college. *Economics of Education Review*, 36:166–180.

Smith, J. (2018). The sequential college application process. *Education Finance and Policy*, 13(4):545–575.

Sondheimer, R. M. and Green, D. P. (2010). Using experiments to estimate the effects of education on voter turnout. *American Journal of Political Science*, 54(1):174–189.

Tenn, S. (2007). The effect of education on voter turnout. *Political Analysis*, 15(4):446–464.

Thomas, N., Gismondi, A., Gautam, P., and Brinker, D. (2021). Democracy counts 2020: Record breaking turnout and student resiliency. *Institute for Democracy & Higher Education. Retrieved November*, 15:2021.

Verba, S., Schlozman, K. L., and Brady, H. E. (1995). *Voice and equality: Civic voluntarism in American politics*. Harvard University Press.

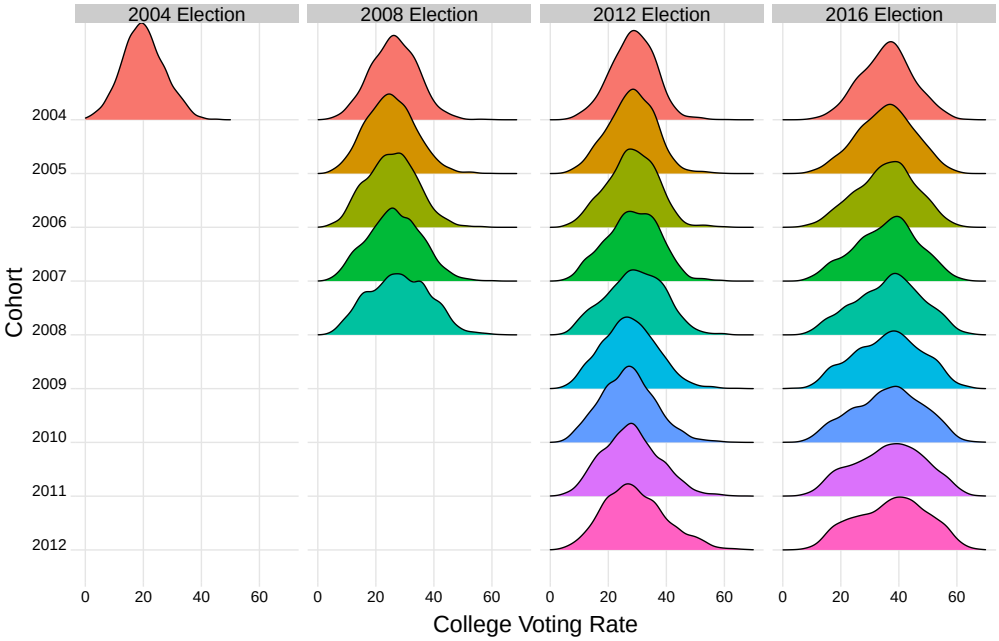
Walters, C. (2024). Empirical bayes methods in labor economics. In *Handbook of Labor Economics*, volume 5, pages 183–260. Elsevier.

Washington, E. (2006). How black candidates affect voter turnout. *The Quarterly Journal of Economics*, 121(3):973–998.

Willeck, C. and Mendelberg, T. (2022). Education and political participation. *Annual Review of Political Science*, 25:89–110.

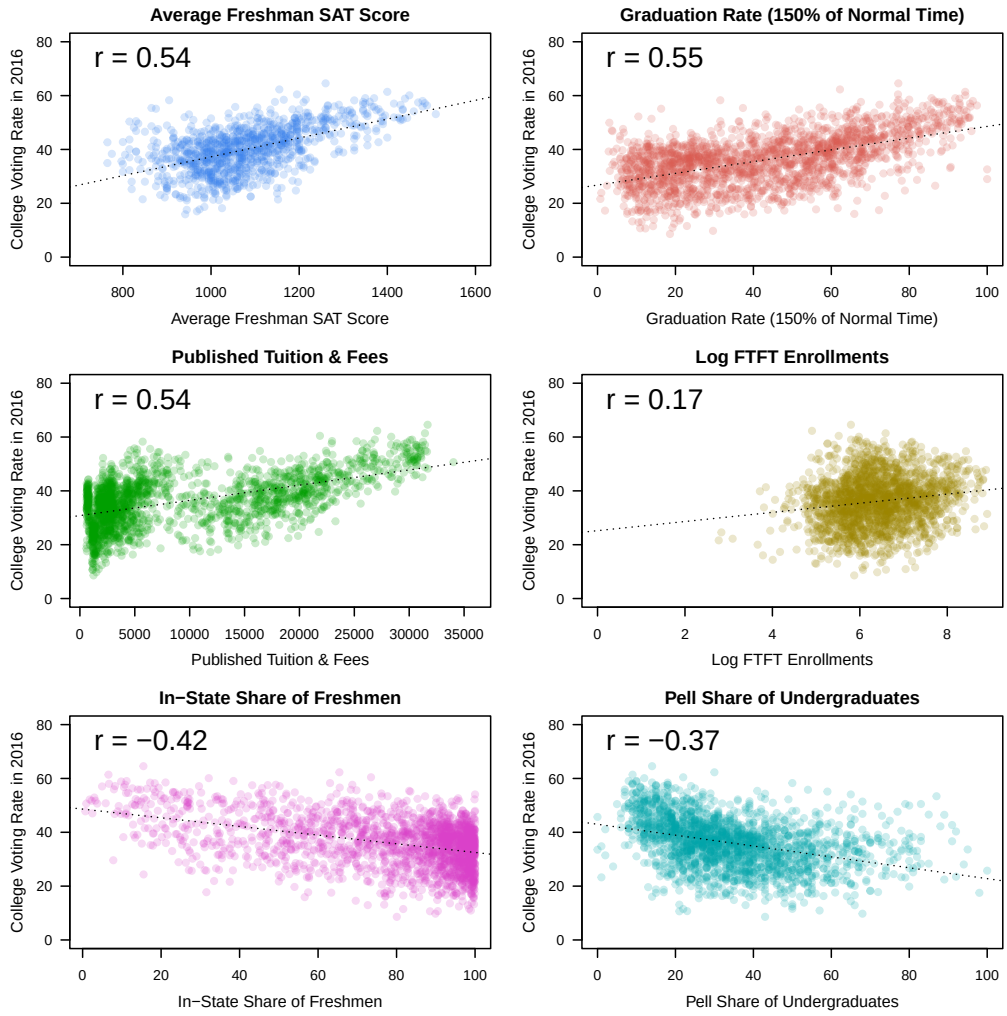
Wolfinger, R. E. and Rosenstone, S. J. (1980). *Who votes?* Yale University Press.

Figure 1. Distribution of College Voting Rates by Election and Cohort.



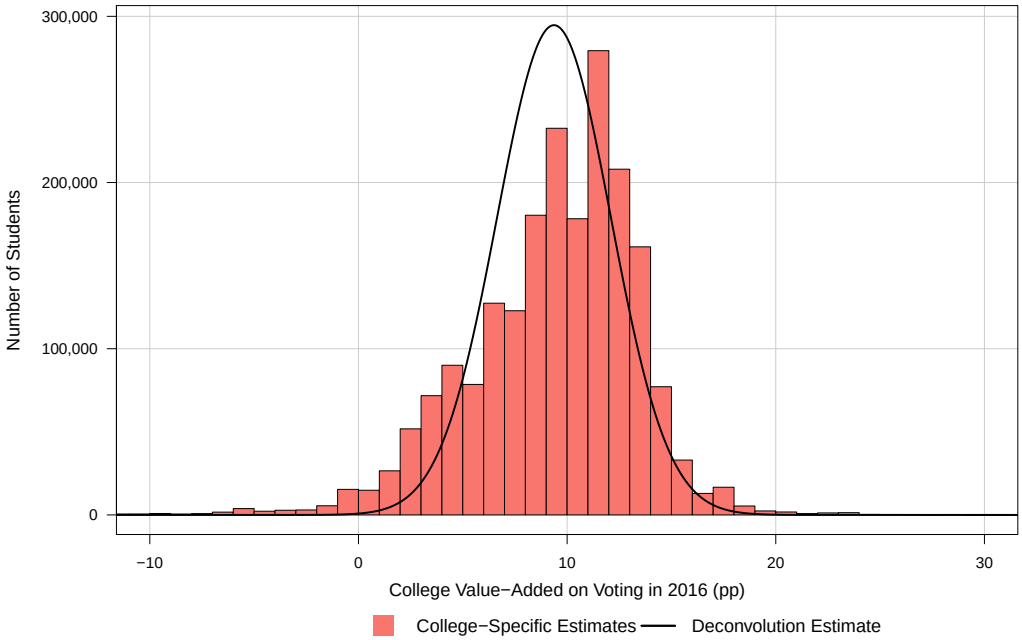
Note: Each plot shows the distribution of voting rates of over 2,000 colleges by cohort and presidential election. Voting rates are calculated as a fraction of colleges' first-time enrollees and are based on college-goers in the full sample. Voting rates based on fewer than 100 voting-age student observations are suppressed.

Figure 2. College Voting Rates in the 2016 Election by College Characteristics - 2004 Cohort.



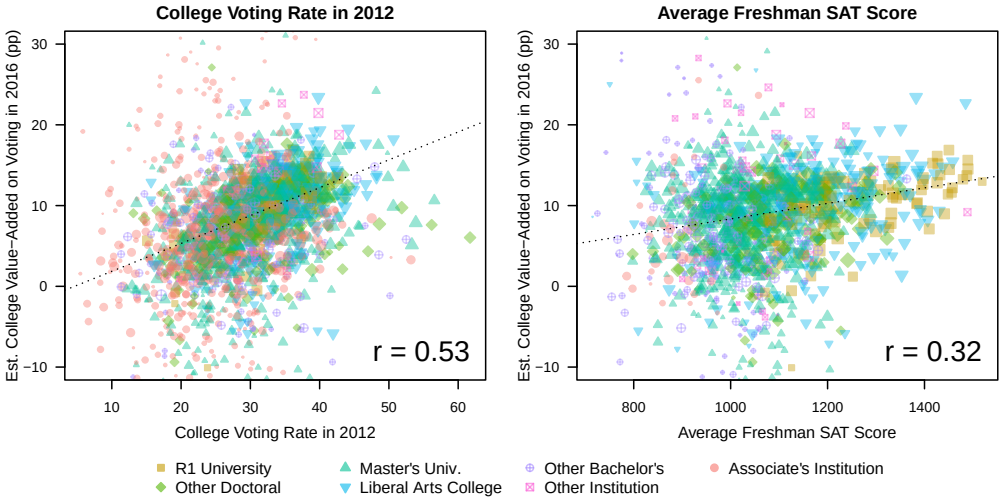
Note: Each observation represents a college's entrants from the 2004 high school graduation cohort. College characteristics are constructed from IPEDS data for Fall 2004 entrants. Voting rates are based on college-goers in the full sample; voting rates based on fewer than 100 voting-age student observations are suppressed.

Figure 3. Student-Weighted Distribution of Estimated College Value-Added on Voting in the 2016 Election



Note: In this figure, the histogram shows the student-weighted distribution of estimates of colleges' value-added on students' probability of voting in the 2016 election for 2,626 colleges in the siblings sample. Estimates of colleges' value-added are recovered from the specification described in Table 5, column 3. The density curve represents a student-weighted deconvolution estimate of the distribution of college value-added, obtained by applying the deconvolution procedure described in the text. Colleges' value-added on students' voting probability is expressed in percentage points and indexed relative to the estimate for non-college goers.

Figure 4. Estimated College Value-Added on Voting in the 2016 Election by College Characteristics



Note: This figure arrays estimates of colleges' value-added on students' probability voting in the 2016 election by two college characteristics of interest: college voting rates and colleges' average freshman SAT scores. Each observation represents a college. Estimates of colleges' value-added are recovered from the specification described in Table 5, column 3. College voting rates are the college-average voting rates of entrants from the 2004 cohort in the 2012 election. Colleges' average freshman SAT scores are calculated from IPEDS data. Pearson correlations are student-weighted.

Table 1—Sample Summary Statistics.

	Full sample		Siblings		Eligible-in-HS sample		Eligible-in-HS siblings	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Demographics								
Female	0.527	0.499	0.523	0.499	0.456	0.498	0.428	0.495
Male	0.473	0.499	0.477	0.499	0.544	0.498	0.572	0.495
Asian	0.067	0.250	0.097	0.296	0.087	0.283	0.063	0.244
Black	0.139	0.346	0.067	0.250	0.107	0.310	0.032	0.176
Hispanic	0.153	0.360	0.088	0.283	0.103	0.304	0.059	0.236
White	0.572	0.495	0.704	0.456	0.638	0.481	0.808	0.394
Other race	0.041	0.198	0.035	0.185	0.035	0.184	0.026	0.158
Missing race	0.027	0.163	0.008	0.091	0.029	0.167	0.011	0.106
Parental Educ - No college	0.113	0.316	0.106	0.308	0.178	0.383	0.092	0.289
Parental Educ - College, no BA	0.130	0.337	0.162	0.369	0.193	0.395	0.140	0.347
Parental Educ - BA or higher	0.286	0.452	0.649	0.477	0.534	0.499	0.698	0.459
Parental Educ - Missing	0.471	0.499	0.083	0.276	0.095	0.293	0.070	0.256
SATs								
Math score	512.9	113.5	557.0	105.8	518.5	116.0	553.3	106.1
Verbal score	503.3	111.2	539.7	104.9	505.4	114.5	536.5	106.9
Number of score sends	4.015	3.143	4.819	3.289	4.403	3.139	4.507	3.211
Mean SAT of score send colleges	1,124.0	118.1	1,151.0	114.1	1,127.4	116.1	1,147.6	111.2
Voting								
First eligible election	0.112	0.315	0.127	0.333	0.223	0.416	0.285	0.452
2016 election	0.339	0.473	0.462	0.499	0.393	0.488	0.440	0.496
Age on Election Day 2016	26.523	2.618	26.149	2.216	27.256	2.285	27.132	2.088
College characteristics								
Non-college goer	0.212	0.409	0.039	0.193	0.081	0.273	0.050	0.218
College avg. freshman SAT	1,107.5	130.1	1,135.3	130.6	1,111.0	132.5	1,132.9	126.0
College graduation rate	47.997	25.757	62.588	22.316	55.353	24.589	62.271	22.243
College voting rate in 2012	29.882	6.971	31.819	6.163	30.964	6.589	31.341	5.983
N Students	20,902,663		2,099,366		850,721		34,034	

Note: The full sample includes all PSAT, SAT, and AP takers in the 2004-2012 high school graduation cohorts living in one of the 50 states or Washington, D.C who were of voting age by the 2016 election. The analytic sample is limited to siblings in the 2005-2012 high school graduation cohorts who took the SAT and had a qualifying score-send portfolio. The eligible-in-HS sample includes domestic SAT takers in the 2005, 2007, 2009, and 2011 high school graduation cohorts who were of voting age by the general election held in November of their senior year of high school and had a qualifying score-send portfolio. The eligible-in-HS siblings subsample is limited to sets of siblings in the eligible-in-HS sample.

Table 2—College-Cohort Voting Rates by Presidential Election, 2004-2016.

	College-cohort voting rates by election							
	2004		2008		2012		2016	
Cohort	Mean	SD	Mean	SD	Mean	SD	Mean	SD
2004	20.0%	(7.2)	26.4%	(8.0)	28.6%	(7.6)	35.9%	(9.0)
2005			25.2%	(8.4)	28.0%	(8.1)	35.6%	(9.7)
2006			25.5%	(8.7)	28.1%	(8.6)	35.6%	(10.3)
2007			26.5%	(9.3)	28.4%	(9.1)	35.7%	(10.8)
2008			28.0%	(10.3)	29.1%	(9.9)	36.4%	(11.4)
2009					27.3%	(9.5)	36.2%	(11.5)
2010					26.9%	(9.4)	36.1%	(11.8)
2011					27.9%	(9.9)	36.6%	(12.1)
2012					29.2%	(10.8)	37.4%	(12.2)

Note: Each statistic is calculated from aggregated data that capture the voting rates of entrants to each college by high school graduation cohort, based on college-goers in the full sample. Voting rates based on fewer than 100 voting-age student observations are excluded from the analysis.

Table 3—Institutional correlates of college-cohort voting rates in the 2016 election.

	Dependent variable:				
	College-cohort voting rate in 2016				
	(1)	(2)	(3)	(4)	(5)
Carnegie Classification *(ref.=Associate's Inst.)*					
R1 University	14.50***	7.16***	6.22***	4.63***	1.71***
	(0.46)	(0.58)	(0.60)	(0.64)	(0.66)
Other Doctoral	10.35***	6.68***	4.92***	4.42***	2.07***
	(0.38)	(0.47)	(0.47)	(0.53)	(0.49)
Master's Univ.	7.58***	5.15***	3.15***	3.02***	1.82***
	(0.27)	(0.39)	(0.37)	(0.46)	(0.43)
Liberal Arts College	13.90***	8.97***	7.24***	6.48***	5.06***
	(0.49)	(0.58)	(0.57)	(0.61)	(0.54)
Other Bachelor's	4.96***	3.30***	1.52***	1.50***	0.93**
	(0.36)	(0.43)	(0.40)	(0.47)	(0.41)
Other Institution	9.24***	6.80***	4.30***	4.44***	4.10***
	(1.01)	(0.95)	(0.92)	(0.94)	(0.89)
Avg. freshman SAT *(1sd=134)*		2.80***		1.59***	1.33***
		(0.18)		(0.22)	(0.20)
150% graduation rate *(1sd=23pp)*			3.45***	2.18***	2.12***
			(0.20)	(0.24)	(0.23)
Log FTFT enrollment *(1sd=0.89)*					0.77***
					(0.11)
Percent in-state *(1sd=24pp)*					−0.32*
					(0.17)
Pell share of students *(1sd=18pp)*					−2.02***
					(0.14)
Percent Black *(1sd=18pp)*					2.22***
					(0.13)
Cohort FEs	Y	Y	Y	Y	Y
College-state FEs	Y	Y	Y	Y	Y
Observations	17,763	17,763	17,763	17,763	17,763
R^2	0.72	0.75	0.76	0.76	0.79

Note: Data are college-by-cohort observations based on college-goers in the full sample. The sample is limited to college-cohorts with at least 100 voting-age student observations. All continuous predictors are standardized. College-clustered standard errors in parentheses (*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

Table 4—Residual Variation in Estimated College Value-Added on Entrants' Covariate-Predicted Probability of Voting in the 2016 Election

	Dependent variable:				
	Covariate-predicted probability of voting in 2016				
	(1)	(2)	(3)	(4)	(5)
SD($\hat{\theta}$)	8.071***	2.484***	2.191***	7.767***	4.284***
	(0.111)	(0.034)	(0.030)	(0.107)	(0.059)
Bias-corrected SD(θ)	7.662***	1.641***	1.241***	7.165***	3.400***
	(0.123)	(0.244)	(0.325)	(0.128)	(0.115)
College + cohort FEs	Y	Y	Y	Y	Y
Family FEs		Y	Y		
Portfolio summary statistics			Y		Y
HS state FEs					Y
SAT math score					Y
Voted in HS					Y
N colleges	2,626	2,626	2,626	2,623	2,623
N students	2,099,366	2,099,366	2,099,366	850,721	850,721

Note: This table reports student-weighted estimates of the standard deviation of colleges' value-added on students' covariate-predicted probability of voting in the 2016 election across five empirical specifications. In each specification, we recover college-specific value-added estimates as college fixed effects from OLS regressions of students' covariate-predicted probability of voting in the 2016 election on the indicated control variables. For each specification, the first row of estimates reports the standard deviation of the college fixed effect point estimates. The second row reports a bias-corrected estimate of the standard deviation of colleges' value-added that accounts for noise in the college value-added estimates while allowing for non-normal errors in the college value-added estimates and correlated noise across college value-added estimates. This variance estimate draws upon the full variance-covariance matrix of the college value-added estimates, which we recover via bootstrap estimation with 100 bootstrap samples per specification. Score-send portfolios are the set of colleges to which students sent their SAT score. In columns 1-3, students' probability of voting in the 2016 election is predicted from an OLS regression of voting in the 2016 election on categorical indicators of gender, race/ethnicity, parental education, and parental income; linear terms in SAT math and verbal scores; and high school fixed effects. The regression sample is limited to domestic SAT-taking siblings in the 2005-2012 high school graduation cohorts who had qualifying score-send portfolios. In columns 4-5, students' probability of voting in the 2016 election is predicted from an OLS regression of voting in the 2016 election on categorical indicators of gender, race/ethnicity, parental education, and parental income; a linear term in SAT verbal score; and high school fixed effects. Regression samples are limited to domestic SAT takers in the 2005, 2007, 2009, and 2011 high school graduation cohorts who were of voting age by the general election held in November of their senior year of high school and had a qualifying score-send portfolio. Standard errors in parentheses (*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

Table 5—Variation in Estimated College Value-Added on Entrants' Probability of Voting in the 2016 Election

	Dependent variable: Voted in 2016				
	(1)	(2)	(3)	(4)	(5)
SD($\hat{\theta}$)	9.257***	5.026***	4.155***	8.909***	4.544***
	(0.128)	(0.069)	(0.057)	(0.123)	(0.063)
Bias-corrected SD(θ)	9.101***	3.922***	2.732***	8.503***	3.686***
	(0.132)	(0.224)	(0.308)	(0.135)	(0.131)
College + cohort FEs	Y	Y	Y	Y	Y
Family FEs		Y	Y		
Student characteristics			Y		Y
Portfolio summary statistics			Y		Y
High school FEs					Y
Voted in HS					Y
N colleges	2,626	2,626	2,626	2,623	2,623
N students	2,099,366	2,099,366	2,099,366	850,721	850,721

Note: This table reports student-weighted estimates of the standard deviation of colleges' value-added on students' probability of voting in the 2016 election across five empirical specifications. In each specification, we recover college-specific value-added estimates as college fixed effects from OLS regressions of student voting in the 2016 election on the indicated control variables. For each specification, the first row of estimates reports the standard deviation of the college fixed effect point estimates. The second row reports a bias-corrected estimate of the standard deviation of colleges' value-added that accounts for noise in the college value-added estimates while allowing for non-normal errors in the college value-added estimates and correlated noise across college value-added estimates. This variance estimate draws upon the full variance-covariance matrix of the college value-added estimates, which we recover via bootstrap estimation with 100 bootstrap samples per specification. Student characteristics include: categorical indicators of gender, race/ethnicity, parental education, and parental income; and linear terms in SAT math and verbal scores. Score-send portfolios are the set of colleges to which students sent their SAT score. In columns 1-3, the regression sample is limited to domestic SAT-taking siblings in the 2005-2012 high school graduation cohorts who had qualifying score-send portfolios. In columns 4-5, regression samples are limited to domestic SAT takers in the 2005, 2007, 2009, and 2011 high school graduation cohorts who were of voting age by the general election held in November of their senior year of high school and had a qualifying score-send portfolio. Standard errors in parentheses (*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

Table 6—Institutional correlates of estimated college value-added on voting in the 2016 election

	Correlation with est. college voting effect in 2016			
College characteristic	Correlation	Std. Error	t-value	p-value
R1 University	0.16	(0.02)	8.39	0.00
Other Doctoral	-0.07	(0.02)	-3.50	0.00
Master's University	0.02	(0.02)	0.81	0.42
Liberal Arts College	0.10	(0.02)	5.38	0.00
Other Bachelor's	-0.11	(0.02)	-5.76	0.00
Other Institution	0.05	(0.02)	2.79	0.01
Associate's Institution	-0.20	(0.02)	-10.20	0.00
NSLVE campus	0.21	(0.02)	10.76	0.00
ALL IN campus	0.16	(0.02)	8.47	0.00
Voter Friendly Campus	0.07	(0.02)	3.68	0.00
On-campus polling place in 2012	0.03	(0.02)	1.43	0.15
Est. value-added on voter registration in 2018	0.74	(0.01)	55.58	0.00
College voting rate in 2012	0.53	(0.02)	27.85	0.00
College peers' pre-college voting rate	0.31	(0.02)	13.30	0.00
College avg. freshman SAT score	0.32	(0.03)	12.63	0.00
College graduation rate (150% of normal time)	0.38	(0.02)	20.63	0.00
Faculty:student ratio	0.26	(0.02)	11.07	0.00
Log per pupil educational expenditure	0.35	(0.02)	15.09	0.00
Percent of students with top 20% personal incomes	0.46	(0.02)	19.83	0.00
Percent of students with top 20% parent incomes	0.28	(0.02)	11.05	0.00
Est. value-added on mid-career earnings	0.30	(0.03)	10.41	0.00
Percent of degrees in social sciences	0.28	(0.03)	11.30	0.00
Percent of degrees in political science	0.22	(0.03)	7.41	0.00

Note: This table presents student-weighted Pearson correlations between college characteristics and colleges' estimated voting effects in the 2016 election. Colleges are weighted by their number of students in the analytic sample. Institutions' Carnegie Classification, average freshman SAT score, 150%-time graduation rate, faculty-to-student ratio, log educational expenditure per student, percent of degrees granted in social sciences, and percent of degrees granted in political science are calculated from IPEDS data. Campus participation in NSLVE, campus participation in the ALL IN Campus Democracy Challenge, and the presence of an on-campus polling place in 2012 come from the data repository at https://www.github.com/nickeubank/mtv_viacom_capstone. Voter Friendly Campus designation comes from https://www.voterfriendlycampus.org/. Estimates of colleges' value-added on their students' voter registration in 2018 are posterior estimates (equivalently, Bayesian shrinkage estimates) calculated from the regression specification described in Table A5, column 2. College voting rates are the college-average voting rates of entrants from the 2004 cohort in the 2012 election. College peers' pre-college voting rates are an average of the pre-college voting rates of entrants from the 2005 and 2009 cohorts in the 2004 and 2008 elections. The percent of students with top 20% parent incomes and the percent of students with top 20% personal incomes come from Chetty et al. (2017). Estimates of colleges' value-added on their students' mid-career earnings come from Rowell and Kulkarni (2015).

Table 7—College Effects on Entrants' Probability of Voting in the 2016 Election, by College Voting Rate and Selectivity

	Dependent variable: Voted in 2016					
	(1)	(2)	(3)	(4)	(5)	(6)
College voting rate in 2012 (10pp)	0.102***	0.026***	0.086***	0.028***	0.082***	0.043***
	(0.003)	(0.001)	(0.003)	(0.002)	(0.006)	(0.009)
College avg. freshman SAT score (100s)	0.021***	0.005***	0.021***	0.007***	0.018***	0.002
	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)	(0.005)
Voted in HS				0.351***		0.292***
				(0.004)		(0.010)
Non-college goer	-0.144***	-0.092***	-0.206***	-0.131***	-0.153***	-0.082***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.004)	(0.011)
Observations	2,099,366	2,099,366	850,721	850,721	34,034	34,034
R^2	0.025	0.557	0.031	0.164	0.019	0.594
Cohort FEs	Y	Y	Y	Y	Y	Y
Student characteristics		Y		Y		Y
Portfolio summary statistics		Y		Y		Y
Family FEs		Y				Y
High school FEs				Y		

Note: OLS regressions. College voting rates are the college-average voting rates of entrants from the 2004 cohort in the 2012 election. Student characteristics include: linear terms in SAT math and verbal scores and categorical indicators of gender, race/ethnicity, parental education, and parental income. Score-send portfolios are the set of colleges to which students sent their SAT score. In columns 1 and 2, regression samples are limited to domestic SAT-taking siblings in the 2005-2012 high school graduation cohorts who had qualifying score-send portfolios. In columns 3 and 4, regression samples are limited to domestic SAT takers in the 2005, 2007, 2009, and 2011 high school graduation cohorts who were of voting age by the general election held in November of their senior year of high school and qualifying score-send portfolios. In columns 5 and 6, regression samples are limited to domestic SAT-taking siblings in the 2005, 2007, 2009, and 2011 high school graduation cohorts who were of voting age by the general election held in November of their senior year of high school and had qualifying score-send portfolios. College-clustered standard errors in parentheses (*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

Table 8—Placebo Test: College Effects on Entrants' Probability of Voting, by College Voting Rate and Selectivity

	Dependent variable: Voted in ...		
	2016 election (Age 24-30)	1st eligible election (Age 18)	1st eligible election (Age 18)
	(1)	(2)	(3)
College voting rate in 2012 (10pp)	0.046***	0.043***	0.011
	(0.009)	(0.007)	(0.009)
College avg. freshman SAT score (100s)	0.003	0.013***	0.004
	(0.005)	(0.003)	(0.004)
Non-college goer	-0.094***	-0.084***	-0.039***
	(0.011)	(0.005)	(0.010)
Cohort FEs	Y	Y	Y
Student characteristics	Y		Y
Portfolio summary statistics	Y		Y
Family FEs	Y		Y
Sample turnout	44.0%	28.5%	28.5%
Observations	34,034	34,034	34,034
R^2	0.566	0.076	0.602

Note: OLS regressions. College voting rates are the college-average voting rates of entrants from the 2004 cohort in the 2012 election. Student characteristics include: categorical indicators of gender, race/ethnicity, parental education, and parental income; and linear terms in SAT math and verbal scores. Score-send portfolios are the set of colleges to which students sent their SAT score. Regression samples are limited to domestic SAT-taking siblings in the 2005, 2007, 2009, and 2011 high school graduation cohorts who were of voting age by the general election held in November of their senior year of high school and had qualifying score-send portfolios. College-clustered standard errors in parentheses (*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

Table 9—Alternative Outcomes: College Effects on Voter Registration and Voting in Different Elections, by College Voting Rate and Selectivity

	Dependent variable:				
	Reg'd. in 2018 (Age 24-32)	Voted 2010 (Age 22-24)	Voted 2012 (Age 22-24)	Voted 2014 (Age 22-24)	Voted 2016 (Age 22-24)
	(1)	(2)	(3)	(4)	(5)
College voting rate in 2012 (10pp)	0.026***	0.025***	0.041***	0.016***	0.040***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
College avg. freshman SAT score (100s)	0.006***	-0.002**	-0.005***	-0.002***	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Non-college goer	-0.116***	-0.037***	-0.075***	-0.030***	-0.130***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.004)
Family + cohort FEs	Y	Y	Y	Y	Y
Student characteristics	Y	Y	Y	Y	Y
Portfolio summary statistics	Y	Y	Y	Y	Y
Focal cohorts	2005-2012	2005-2006	2007-2008	2009-2010	2011-2012
Focal cohorts' turnout	73.5%	14.9%	35.1%	14.0%	47.2%
Focal cohort observations	2,099,366	415,631	593,403	586,886	457,483
Observations	2,099,366	1,344,216	2,085,677	2,099,269	2,099,366
R^2	0.573	0.578	0.571	0.567	0.558

Note: OLS regressions. In these regressions, the outcomes are binary indicators of being registered to vote on October 17, 2018 and voting in the 2010 midterm election, 2012 presidential election, 2014 midterm election, and 2016 presidential election. College voting rates are the college-average voting rates of entrants from the 2004 cohort in the 2012 election. Student characteristics include: linear terms in SAT math and verbal scores and categorical indicators of gender, race/ethnicity, parental education, and parental income. Score-send portfolios are the set of colleges to which students sent their SAT score. Regression samples are limited to domestic SAT-taking siblings in the 2005-2012 high school graduation cohorts who had qualifying score-send portfolios. College-clustered standard errors in parentheses (*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

Table 10—Time Path of College Effects on Voting, by College Voting Rate and Selectivity

	Dependent variable: Voted in...				
	1st post-HS Election (Age 18-20)	2nd post-HS Election (Age 20-22)	3rd post-HS Election (Age 22-24)	4th post-HS Election (Age 24-26)	5th post-HS Election (Age 26-28)
	(1)	(2)	(3)	(4)	(5)
College voting rate in 2012 (10pp)	0.027***	0.024***	0.021***	0.016***	0.015***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
College avg. freshman SAT score (100s)	-0.003**	-0.006***	-0.002*	-0.000	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Non-college goer	-0.024***	-0.035***	-0.046***	-0.057***	-0.055***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Family + cohort FEs	Y	Y	Y	Y	Y
Student characteristics	Y	Y	Y	Y	Y
Portfolio summary statistics	Y	Y	Y	Y	Y
Sample cohorts	2005-2008	2005-2008	2005-2008	2005-2008	2005-2008
Sample turnout	19.2%	19.9%	23.8%	23.1%	29.7%
Observations	545,188	545,161	545,159	545,137	533,312
R^2	0.587	0.575	0.581	0.576	0.590

Note: OLS regressions. In these regressions, the outcomes are binary indicators of whether students voted in their 1st, 2nd, 3rd, 4th, and 5th eligible elections after on-time high school graduation. College voting rates are the college-average voting rates of entrants from the 2004 cohort in the 2012 election. Student characteristics include: linear terms in SAT math and verbal scores and categorical indicators of gender, race/ethnicity, parental education, and parental income. Score-send portfolios are the set of colleges to which students sent their SAT score. Regression samples are limited to domestic SAT-taking siblings in the indicated high school graduation cohorts who had qualifying score-send portfolios. College-clustered standard errors in parentheses (*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

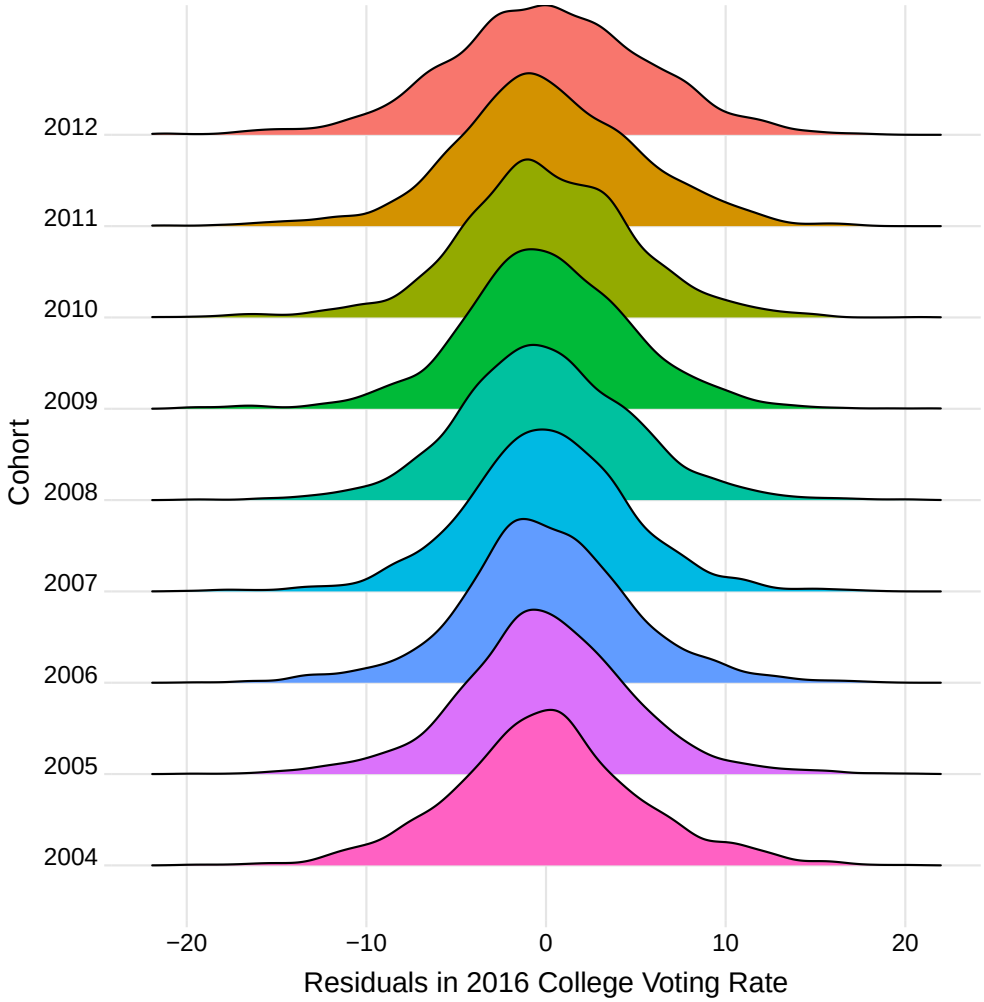
Table 11—Heterogeneous Effects of College Voting Rate and Selectivity, by High School Voting History.

	Dependent variable: Voted in 2016		
	(1)	(2)	(3)
College voting rate in 2012 (10pp)	0.029***	0.037***	0.055***
	(0.005)	(0.002)	(0.011)
College voting rate × Voted in HS	-0.042***	-0.039***	-0.045**
	(0.009)	(0.004)	(0.018)
College avg. freshman SAT score (100s)	0.012***	0.012***	0.007
	(0.002)	(0.002)	(0.005)
College avg. SAT × Voted in HS	-0.022***	-0.022***	-0.018**
	(0.004)	(0.002)	(0.009)
Voted in HS	0.273***	0.353***	0.297***
(ref. = eligible non-voters)	(0.005)	(0.003)	(0.010)
Non-college goer	-0.085***	-0.132***	-0.083***
	(0.004)	(0.001)	(0.011)
Cohort FEs	Y	Y	Y
Student characteristics	Y	Y	Y
Portfolio summary statistics	Y	Y	Y
Family FEs	Y		Y
High school FEs		Y	
Observations	2,099,366	850,721	34,034
R^2	0.560	0.166	0.595

Note: OLS regressions. College voting rates are the college-average voting rates of entrants from the 2004 cohort in the 2012 election. Student characteristics include: linear terms in SAT math and verbal scores and categorical indicators of gender, race/ethnicity, parental education, and parental income. Score-send portfolios are the set of colleges to which students sent their SAT score. In column 1, the regression sample is limited to domestic SAT-taking siblings in the 2005-2012 high school graduation cohorts who had qualifying score-send portfolios. In column 2, the regression sample is limited to domestic SAT takers in the 2005, 2007, 2009, and 2011 high school graduation cohorts who were of voting age by the general election held in November of their senior year of high school and had qualifying score-send portfolios. In column 3, the regression sample is limited to domestic SAT-taking siblings in the 2005, 2007, 2009, and 2011 high school graduation cohorts who were of voting age by the general election held in November of their senior year of high school and had qualifying score-send portfolios. The regression specification in column 1 differs from those in columns 2 and 3. Because most students in the analytic sample were not eligible to vote in high school, the specification reported in column 1 interacts all regression terms except family ID with an indicator of whether the student was eligible to vote in high school; coefficients reported in column 1 pertain to the subset of students in the analytic sample who were eligible to vote in high school. College-clustered standard errors in parentheses (*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

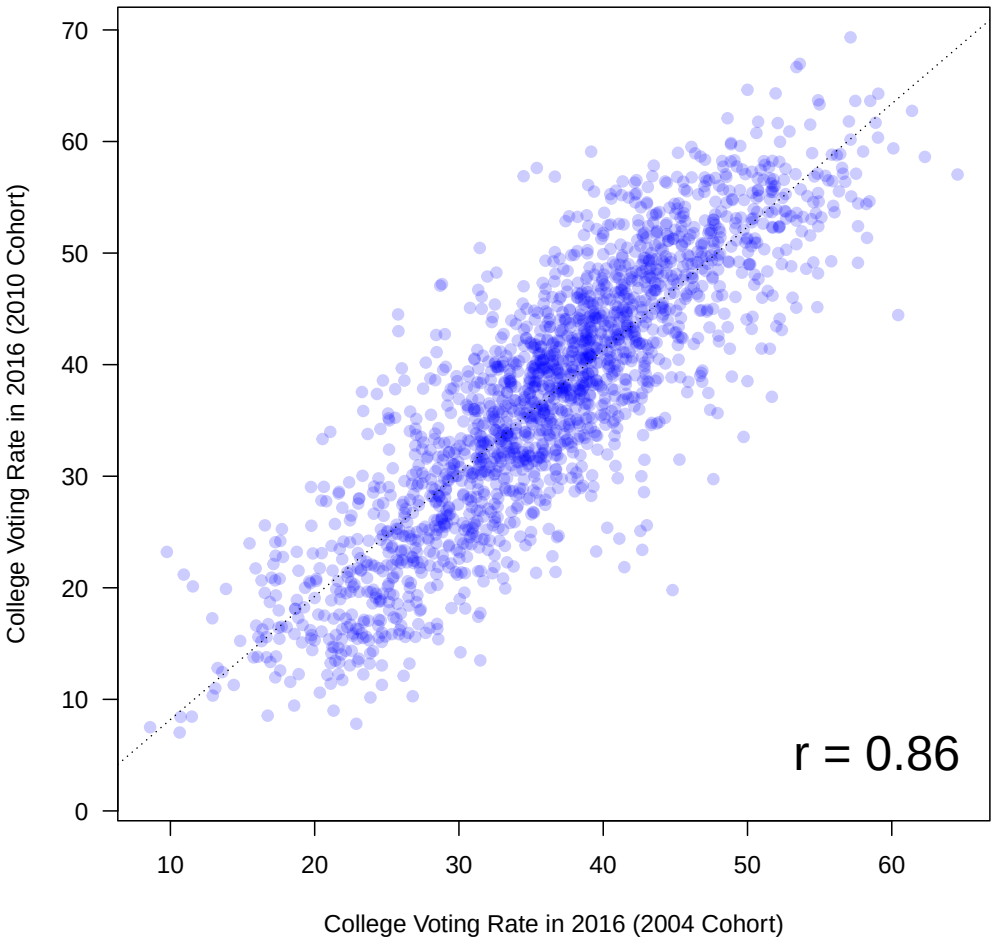
Appendix A - Additional Results

Figure A1. Unexplained Variation in College Voting Rates in the 2016 Election, by Cohort.



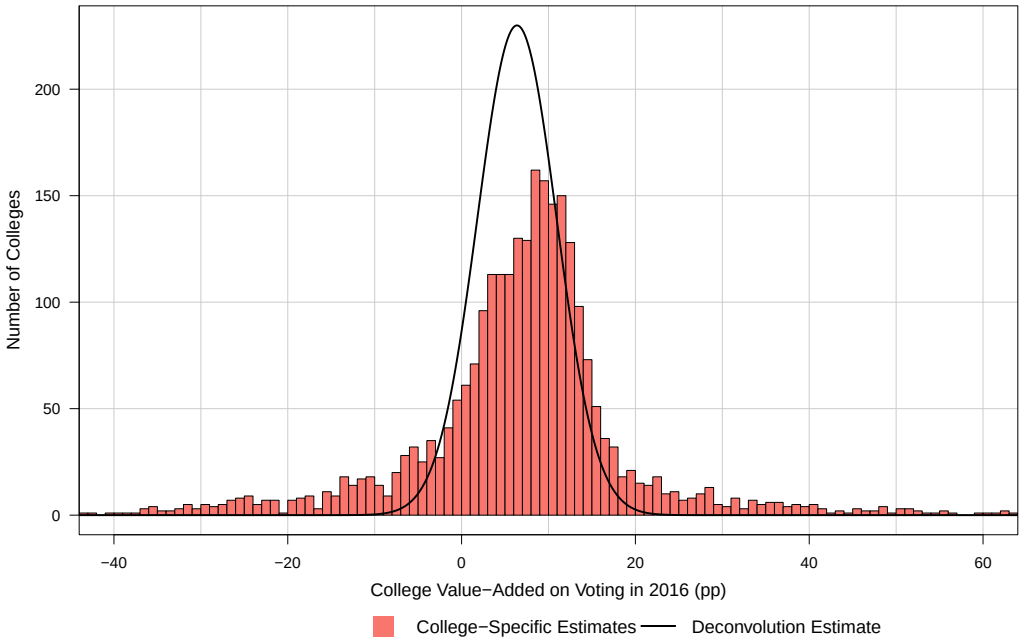
Note: Each plot represents the distribution of each cohort's residuals from the regression in Table 3, Column 5. The residuals represent the unexplained variation in voting rates by college in the 2016 presidential election. Voting rates calculated as fraction of first-time enrollees at a college, with at least 100 students.

Figure A2. College Voting Rates in the 2016 Election – 2004 vs. 2010 Cohorts.



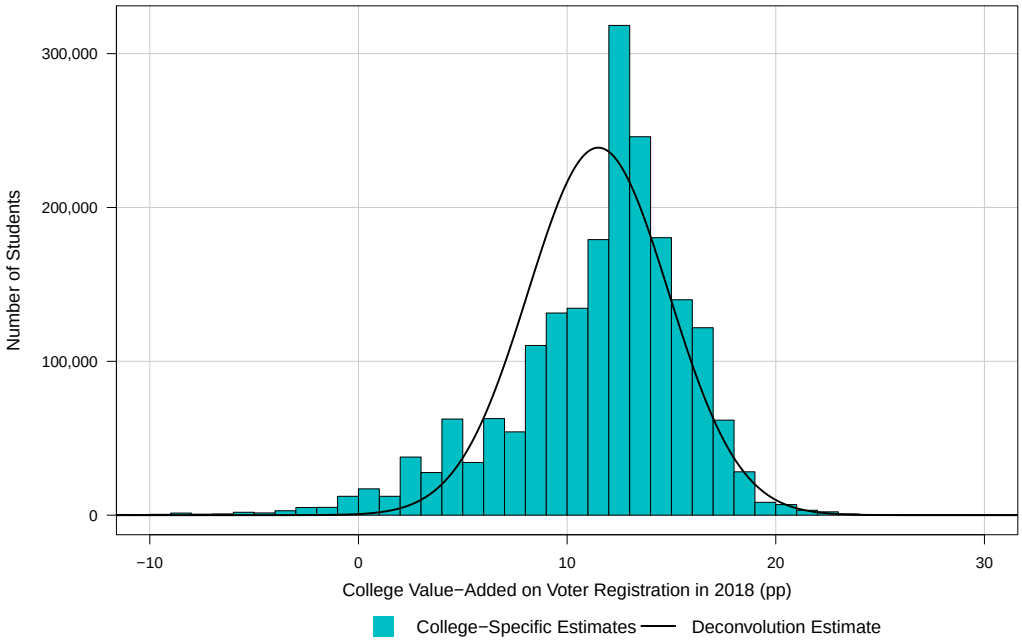
Note: Each observation represents a college. Voting rates are based on college-goers in the full sample; voting rates based on fewer than 100 voting-age student observations are suppressed.

Figure A3. College-Weighted Distribution of Estimated College Value-Added on Entrants' Probability of Voting in the 2016 Election



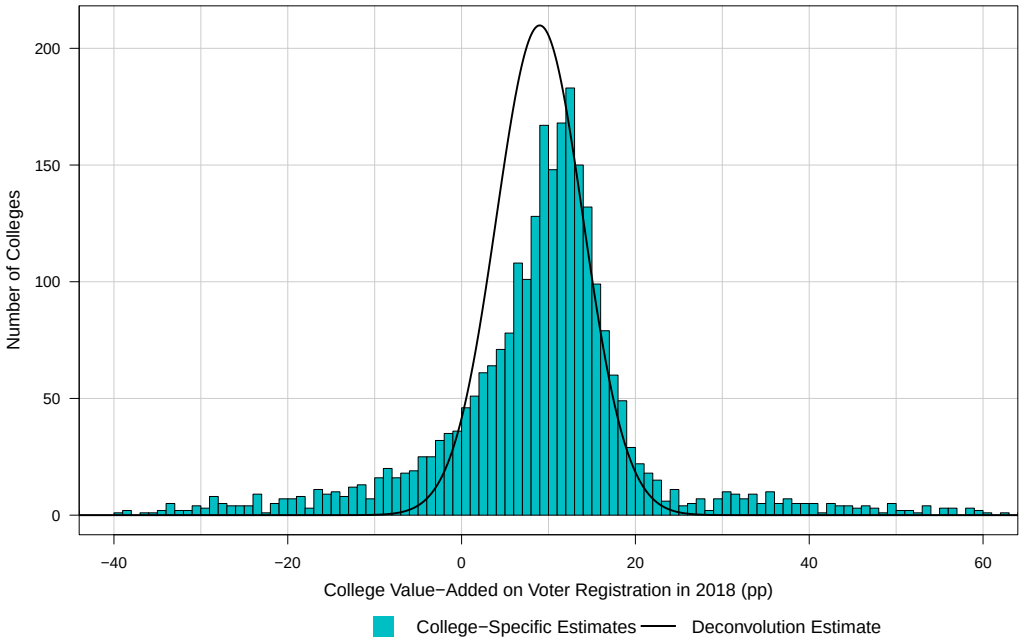
Note: In this figure, the histogram shows the distribution of estimates of colleges' value-added on students' probability of voting in the 2016 election for 2,626 colleges in the siblings sample. Estimates of colleges' value-added are recovered from the specification described in Table A4, column 3. The density curve represents a deconvolution estimate of the distribution of college value-added, obtained by applying the deconvolution procedure described in the text. Colleges' value-added on students' voting probability is expressed in percentage points and indexed relative to the estimate for non-college goers.

Figure A4. Student-Weighted Distribution of Estimated College Value-Added on Entrants' Probability of Voter Registration in 2018



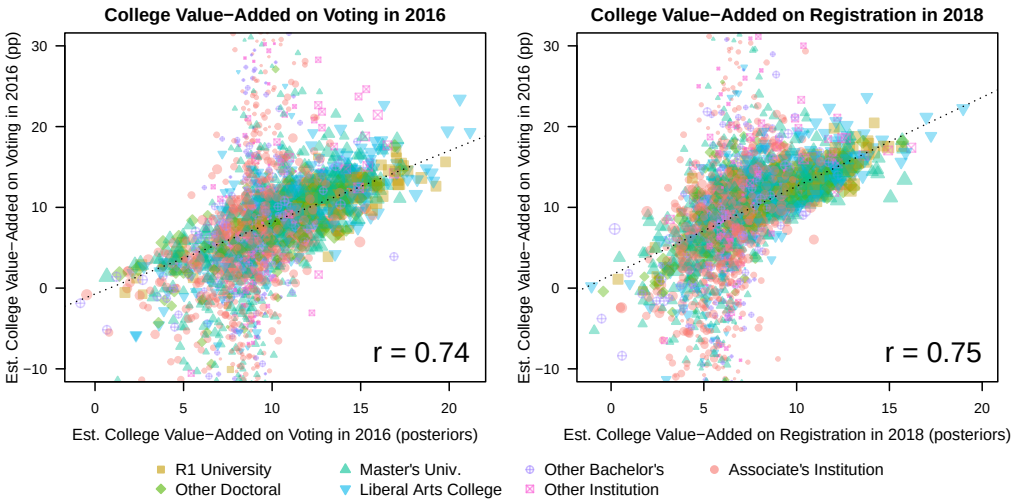
Note: In this figure, the histogram shows the student-weighted distribution of estimates of colleges' value-added on students' probability of voter registration in 2018 for 2,626 colleges in the siblings sample. Estimates of colleges' value-added are recovered from the specification described in Table A5, column 2. The density curve represents a student-weighted deconvolution estimate of the distribution of college value-added, obtained by applying the deconvolution procedure described in the text. Colleges' value-added on students' voting probability is expressed in percentage points and indexed relative to the estimate for non-college goers.

Figure A5. College-Weighted Distribution of Estimated College Value-Added on Entrants' Probability of Voter Registration in 2018



Note: In this figure, the histogram shows the distribution of estimates of colleges' value-added on students' probability of voter registration in 2018 for 2,626 colleges in the siblings sample. Estimates of colleges' value-added are recovered from the specification described in Table A5, column 4. The density curve represents a deconvolution estimate of the distribution of college value-added, obtained by applying the deconvolution procedure described in the text. Colleges' value-added on students' voting probability is expressed in percentage points and indexed relative to the estimate for non-college goers.

Figure A6. Correlations Between Colleges' Value-Added on Entrants' Probability of Voting in the 2016 Election and Colleges' Value-Added on Entrants' Probability of Voter Registration in 2018



Note: This figure arrays estimates of colleges' value-added on students' probability voting in the 2016 election by estimates of colleges' value-added on students' probability of voter registration in 2018, and vice versa. Each observation represents a college. In the first panel, estimates of colleges' value-added on students' probability voting in the 2016 election are arrayed by posterior estimates of colleges' value-added on students' probability of voter registration in 2018. In the second panel, estimates of colleges' value-added on students' probability of voter registration in 2018 are arrayed by posterior estimates of colleges' value-added on students' probability voting in the 2016 election. (Employing posterior estimates of college value-added on the x-axis reduces attenuation bias in the slope of the regression line.) Standard and posterior estimates of colleges' value-added on voting are calculated from the specification described in Table 5, column 3; standard and posterior estimates of colleges' value-added on voter registration are calculated from the specification described in Table A5, column 2. Pearson correlations are student-weighted.

Table A1—Cross-cohort correlations in 2016 college voting rates, 2004-2012 Cohorts

	Cohort								
	2004	2005	2006	2007	2008	2009	2010	2011	2012
2004	1.00								
2005	0.91	1.00							
2006	0.91	0.93	1.00						
2007	0.90	0.93	0.94	1.00					
2008	0.89	0.92	0.93	0.95	1.00				
2009	0.87	0.90	0.92	0.94	0.95	1.00			
2010	0.85	0.89	0.91	0.93	0.94	0.95	1.00		
2011	0.84	0.88	0.90	0.92	0.93	0.95	0.96	1.00	
2012	0.81	0.84	0.87	0.89	0.91	0.92	0.94	0.95	1.00

Note: These Pearson correlation coefficients are calculated from aggregated data that capture the voting rates of entrants to each college by high school graduation cohort, based on college-goers in the full sample. Each coefficient gives the correlation between the college-average voting rates of entrants from the indicated high school graduation cohorts in the 2016 election. Voting rates based on fewer than 100 voting-age student observations are excluded from the analysis.

Table A2—Cross-election correlations in college voting rates, 2004 Cohort

	Election						
	2004	2006	2008	2010	2012	2014	2016
2004	1.00						
2006	0.66	1.00					
2008	0.80	0.55	1.00				
2010	0.50	0.57	0.68	1.00			
2012	0.66	0.47	0.90	0.77	1.00		
2014	0.30	0.38	0.46	0.69	0.65	1.00	
2016	0.53	0.38	0.74	0.73	0.86	0.67	1.00

Note: These Pearson correlation coefficients are calculated from aggregated data that capture the voting rates of entrants to each college by high school graduation cohort, based on college-goers in the full sample. Each coefficient gives the correlation between the college-average voting rates of entrants from the 2004 high school graduation cohorts in the indicated elections. Voting rates based on fewer than 100 voting-age student observations are excluded from the analysis.

Table A3—College-Weighted Residual Variation in Estimated College Value-Added on Entrants' Covariate-Predicted Probability of Voting in the 2016 Election

	Dependent variable:				
	Covariate-predicted probability of voting in 2016				
	(1)	(2)	(3)	(4)	(5)
SD($\hat{\theta}$)	7.277***	2.045***	1.301***	6.961***	2.794***
	(0.100)	(0.028)	(0.018)	(0.096)	(0.039)
Bias-corrected SD(θ)	7.267***	2.012***	1.250***	6.923***	2.717***
	(0.101)	(0.032)	(0.030)	(0.097)	(0.042)
College + cohort FEs	Y	Y	Y	Y	Y
Family FEs		Y	Y		
Portfolio summary statistics			Y		Y
HS state FEs					Y
SAT math score					Y
Voted in HS					Y
N colleges	2,626	2,626	2,626	2,623	2,623
N students	2,099,366	2,099,366	2,099,366	850,721	850,721

Note: This table reports college-weighted estimates of the standard deviation of colleges' value-added on students' covariate-predicted probability of voting in the 2016 election across five empirical specifications. In each specification, we recover college-specific value-added estimates as college fixed effects from OLS regressions of students' covariate-predicted probability of voting in the 2016 election on the indicated control variables. For each specification, the first row of estimates reports the standard deviation of the college fixed effect point estimates. The second row reports a bias-corrected estimate of the standard deviation of colleges' value-added that accounts for noise in the college value-added estimates while allowing for non-normal errors in the college value-added estimates and correlated noise across college value-added estimates. This variance estimate draws upon the full variance-covariance matrix of the college value-added estimates, which we recover via bootstrap estimation with 100 bootstrap samples per specification. Score-send portfolios are the set of colleges to which students sent their SAT score. In columns 1-3, students' probability of voting in the 2016 election is predicted from an OLS regression of voting in the 2016 election on categorical indicators of gender, race/ethnicity, parental education, and parental income; linear terms in SAT math and verbal scores; and high school fixed effects. The regression sample is limited to domestic SAT-taking siblings in the 2005-2012 high school graduation cohorts who had qualifying score-send portfolios. In columns 4-5, students' probability of voting in the 2016 election is predicted from an OLS regression of voting in the 2016 election on categorical indicators of gender, race/ethnicity, parental education, and parental income; a linear term in SAT verbal score; and high school fixed effects. Regression samples are limited to domestic SAT takers in the 2005, 2007, 2009, and 2011 high school graduation cohorts who were of voting age by the general election held in November of their senior year of high school and had a qualifying score-send portfolio. Standard errors in parentheses (*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

Table A4—College-weighted Variation in Estimated College Value-Added on Entrants' Probability of Voting in the 2016 Election

	Dependent variable: Voted in 2016				
	(1)	(2)	(3)	(4)	(5)
SD($\hat{\theta}$)	14.820***	14.728***	14.434***	14.574***	11.936***
	(0.204)	(0.203)	(0.199)	(0.201)	(0.165)
Bias-corrected SD(θ)	11.122***	5.504***	4.556**	11.011***	7.045***
	(0.385)	(1.876)	(2.242)	(0.368)	(0.557)
College + cohort FEs	Y	Y	Y	Y	Y
Family FEs		Y	Y		
Student characteristics			Y		Y
Portfolio summary statistics			Y		Y
High school FEs					Y
Voted in HS					Y
N colleges	2,626	2,626	2,626	2,623	2,623
N students	2,099,366	2,099,366	2,099,366	850,721	850,721

Note: This table reports college-weighted estimates of the standard deviation of colleges' value-added on students' probability of voting in the 2016 election across five empirical specifications. In each specification, we recover college-specific value-added estimates as college fixed effects from OLS regressions of student voting in the 2016 election on the indicated control variables. For each specification, the first row of estimates reports the standard deviation of the college fixed effect point estimates. The second row reports a bias-corrected estimate of the standard deviation of colleges' value-added that accounts for noise in the college value-added estimates while allowing for non-normal errors in the college value-added estimates and correlated noise across college value-added estimates. This variance estimate draws upon the full variance-covariance matrix of the college value-added estimates, which we recover via bootstrap estimation with 100 bootstrap samples per specification. Student characteristics include: categorical indicators of gender, race/ethnicity, parental education, and parental income; and linear terms in SAT math and verbal scores. Score-send portfolios are the set of colleges to which students sent their SAT score. In columns 1-3, the regression sample is limited to domestic SAT-taking siblings in the 2005-2012 high school graduation cohorts who had qualifying score-send portfolios. In columns 4-5, regression samples are limited to domestic SAT takers in the 2005, 2007, 2009, and 2011 high school graduation cohorts who were of voting age by the general election held in November of their senior year of high school and had a qualifying score-send portfolio. Standard errors in parentheses (*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

Table A5—Variation in Estimated College Value-Added on Entrants' Probability of Voter Registration in 2018

	Dependent variable: Registered to Vote in 2018			
	Student-weighted		College-weighted	
	(1)	(2)	(3)	(4)
SD($\hat{\theta}$)	11.295***	4.444***	16.266***	13.971***
	(0.156)	(0.061)	(0.224)	(0.193)
Bias-corrected SD(θ)	11.183***	3.370***	13.101***	4.993***
	(0.159)	(0.235)	(0.362)	(1.856)
College + cohort FEs	Y	Y	Y	Y
Family FEs		Y		Y
Student characteristics		Y		Y
Portfolio summary statistics		Y		Y
N colleges	2,626	2,626	2,626	2,626
N students	2,099,366	2,099,366	2,099,366	2,099,366

Note: This table reports student-weighted estimates of the standard deviation of colleges' value-added on students' probability of being registered to vote on October 17, 2018 across two empirical specifications. In each specification, we recover college-specific value-added estimates as college fixed effects from OLS regressions of students' voter registration status on the indicated control variables. For each specification, the first row of estimates reports the standard deviation of the college fixed effect point estimates. The second row reports a bias-corrected estimate of the standard deviation of colleges' value-added that accounts for noise in the college value-added estimates while allowing for non-normal errors in the college value-added estimates and correlated noise across college value-added estimates. This variance estimate draws upon the full variance-covariance matrix of the college value-added estimates, which we recover via bootstrap estimation with 100 bootstrap samples per specification. Student characteristics include: categorical indicators of gender, race/ethnicity, parental education, and parental income; and linear terms in SAT math and verbal scores. Score-send portfolios are the set of colleges to which students sent their SAT score. The regression sample is limited to domestic SAT-taking siblings in the 2005-2012 high school graduation cohorts who had qualifying score-send portfolios. Standard errors in parentheses (*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

Table A6—Intermediate Specifications: College Effects on Entrants' Probability of Voting in the 2016 Election, by College Voting Rate and Selectivity

	Dependent variable: Voted in 2016			
	(1)	(2)	(3)	(4)
College voting rate in 2012 (10pp)	0.102***	0.032***	0.027***	0.026***
	(0.003)	(0.001)	(0.001)	(0.001)
College avg. freshman SAT score (100s)	0.021***	0.016***	0.009***	0.005***
	(0.002)	(0.001)	(0.001)	(0.001)
Non-college goer	-0.153***	-0.100***	-0.097***	-0.095***
	(0.003)	(0.002)	(0.002)	(0.002)
Observations	2,099,366	2,099,366	2,099,366	2,099,366
R^2	0.025	0.556	0.557	0.557
Cohort FEs	Y	Y	Y	Y
Family FEs		Y	Y	Y
Portfolio summary statistics			Y	Y
Student characteristics				Y

Note: OLS regressions. College voting rates are the college-average voting rates of entrants from the 2004 cohort in the 2012 election. Student characteristics include: linear terms in SAT math and verbal scores and categorical indicators of gender, race/ethnicity, parental education, and parental income. Score-send portfolios are the set of colleges to which students sent their SAT score. The regression sample is limited to domestic SAT-taking siblings in the 2005-2012 high school graduation cohorts who had qualifying score-send portfolios. College-clustered standard errors in parentheses (*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

Table A7—Intermediate Specifications: College Effects on Entrants' Probability of Voting in the 2016 Election, by College Voting Rate and Selectivity

	Dependent variable: Voted in 2016				
	(1)	(2)	(3)	(4)	(5)
College voting rate in 2012 (10pp)	0.086***	0.044***	0.038***	0.029***	0.028***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
College avg. freshman SAT score (100s)	0.021***	0.018***	0.014***	0.011***	0.007***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
Voted in HS			0.367***	0.366***	0.351***
			(0.004)	(0.004)	(0.004)
Non-college goer	-0.213***	-0.187***	-0.154***	-0.151***	-0.134***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)
Observations	850,721	850,721	850,721	850,721	850,721
R^2	0.031	0.070	0.155	0.156	0.164
Cohort FEs	Y	Y	Y	Y	Y
High school FEs		Y	Y	Y	Y
Portfolio summary statistics				Y	Y
Student characteristics					Y

Note: OLS regressions. College voting rates are the college-average voting rates of entrants from the 2004 cohort in the 2012 election. Student characteristics include: linear terms in SAT math and verbal scores and categorical indicators of gender, race/ethnicity, parental education, and parental income. Score-send portfolios are the set of colleges to which students sent their SAT score. The regression sample is limited to domestic SAT takers in the 2005, 2007, 2009, and 2011 high school graduation cohorts who were of voting age by the general election held in November of their senior year of high school and qualifying score-send portfolios. College-clustered standard errors in parentheses (*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

Table A8—Coefficient Sensitivity Test (Oster, 2019)

	Coefficient sensitivity test for Table 7, column 2				
	Uncontrolled	Controlled	β^* if $\delta = 1$ and $R_{\max} = \ldots$		
	$\dot{\beta}$	$\tilde{\beta}$	$1.1 \cdot \tilde{R}^2$	$1.2 \cdot \tilde{R}^2$	$1.3 \cdot \tilde{R}^2$
	(1)	(2)	(3)	(4)	(5)
College voting rate in 2012 (10pp)	0.102	0.026	0.018	0.010	0.002
College avg. freshman SAT score (100s)	0.021	0.005	0.003	0.002	0.000
Non-college goer	-0.153	-0.095	-0.088	-0.082	-0.076
R^2	$\dot{R}^2 = 0.025$	$\tilde{R}^2 = 0.557$	$R_{\max} = 0.613$	$R_{\max} = 0.669$	$R_{\max} = 0.725$

Note: Following Oster (2019), we assess the sensitivity of the coefficients reported in Table 2, column 2 to omitted covariates by estimating bias-adjusted coefficients using the results of the uncontrolled and controlled regression specifications reported in Table 2, columns 1-2 (reproduced in columns 1-2 here). We report bias-adjusted coefficients in columns 3-5. Each bias-adjusted coefficient is calculated as $\beta^* = \tilde{\beta} - \delta \cdot (\dot{\beta} - \tilde{\beta}) \cdot \frac{R_{\max} - \tilde{R}^2}{\tilde{R}^2 - \dot{R}^2}$, where $\tilde{\beta}$ is the coefficient estimate from the controlled regression in column 2, $\dot{\beta}$ is the coefficient estimate from the uncontrolled regression in column 1, $R_{\max}$ denotes the maximum possible R^2 of a hypothetical regression specification that includes all observed and unobserved covariates predictive of voting in 2016, $\tilde{R}^2$ is the R^2 of the controlled regression in column 2, and $\dot{R}^2$ is the R^2 of the uncontrolled regression in column 1. In columns 3-5, we assume $\delta = 1$ and compute Oster's β^* in the case that $R_{\max}$ is 1.1, 1.2, and 1.3 times $\tilde{R}^2$, as recommended in Oster (2019).

Table A9—Alternative Samples: College Effects on Entrants' Probability of Voting in the 2016 Election by College Voting Rate and Selectivity

	Dependent variable: Voted in 2016						
	College goers	Exact matches	Females	Males	High-coverage colleges	High-coverage states	Matched-turnout states
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
College voting rate in 2012 (10pp)	0.026***	0.030***	0.021***	0.022***	0.026***	0.025***	0.019***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
College avg. freshman SAT score (100s)	0.006***	0.005***	0.004***	0.004***	0.006***	0.006***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Non-college goer		-0.091***	-0.081***	-0.102***		-0.092***	
		(0.002)	(0.003)	(0.003)		(0.002)	
Family + cohort FEs	Y	Y	Y	Y	Y	Y	Y
Student characteristics	Y	Y	Y	Y	Y	Y	Y
Portfolio summary statistics	Y	Y	Y	Y	Y	Y	Y
Observations	1,954,412	1,512,384	616,688	519,669	1,951,316	1,528,138	874,315
R^2	0.556	0.569	0.583	0.579	0.556	0.559	0.564

Note: OLS regressions. College voting rates are the college-average voting rates of entrants from the 2004 cohort in the 2012 election. Student characteristics include: linear terms in SAT math and verbal scores and categorical indicators of gender, race/ethnicity, parental education, and parental income. Score-send portfolios are the set of colleges to which students sent their SAT score. Regression samples are limited to domestic SAT-taking siblings in the 2005-2012 high school graduation cohorts who had qualifying score-send portfolios. In columns 1-7, respectively, the samples are further limited to: college goers; exact matches to the voting data; females; males; entrants to colleges where at least 70 percent of first-time undergraduates are represented in the education data; students from states where at least 70 percent of first-time undergraduates are represented in the education data; and students from states where the 2016 turnout rate in the matched data falls within 5 percentage points of the 2016 turnout rate among 18-24 year-olds in Census data. College-clustered standard errors in parentheses (*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

Table A10—Alternative Samples, Cont'd.: College Effects on Entrants' Probability of Voting in the 2016 Election by College Voting Rate

	Dependent variable: Voted in 2016						
	<25% prob. of 4yr college	No sends to sel. 4yr colleges	<25% prob. of sel. 4yr college	2-year colleges	4-year colleges	Non-sel. 4-year colleges	Selective 4-year colleges
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
College voting rate	0.013**	0.019***	0.018***	0.020*	0.028***	0.017***	0.040***
in 2012 (10pp)	(0.006)	(0.005)	(0.002)	(0.010)	(0.002)	(0.003)	(0.003)
College avg. freshman	-0.009	0.009***	0.006***	0.012	0.006***	0.005*	0.009***
SAT score (100s)	(0.016)	(0.004)	(0.002)	(0.017)	(0.001)	(0.002)	(0.002)
Non-college goer	-0.044	-0.106***	-0.101***				
	(0.044)	(0.008)	(0.004)				
High school + cohort FEs	Y	Y	Y	Y	Y	Y	Y
Student controls	Y	Y	Y	Y	Y	Y	Y
Portfolio summary statistics	Y	Y	Y	Y	Y	Y	Y
Observations	102,571	137,060	704,875	79,626	1,567,194	409,317	691,468
R^2	0.597	0.581	0.575	0.591	0.553	0.571	0.546

Note: OLS regressions. College voting rates are the college-average voting rates of entrants from the 2004 cohort in the 2012 election. Student characteristics include: linear terms in SAT math and verbal scores and categorical indicators of gender, race/ethnicity, parental education, and parental income. Score-send portfolios are the set of colleges to which students sent their SAT score. Regression samples are limited to domestic SAT-taking siblings in the 2005-2012 high school graduation cohorts who had qualifying score-send portfolios. In columns 1-7, respectively, the samples are further limited to: students with less than a 25% estimated probability of attending a 4-year college; students who did not send their scores to any selective 4-year college; students with less than a 25% estimated probability of attending a selective 4-year college; and students who attended 2-year colleges, 4-year colleges, non-selective 4-year colleges, and selective 4-year colleges. Selective colleges are defined as 4-year colleges with an average freshman SAT score of 1100 or greater. Attendance probabilities are estimated by regressing college-type indicators on the set of controls used in our preferred specification.

Table A11—Heterogeneous Effects of College Entry by College Voting Rate and Average SAT Score

	Dependent variable: Voted in 2016							
	Female (1)	Asian (2)	Black (3)	Hispanic (4)	Parent ed. BA+ (5)	SAT score (100s) (6)	Out-of-state (7)	Graduated college (8)
Reference group:								
College voting rate in 2012 (10pp)	0.017***	0.026***	0.029***	0.026***	0.022***	0.025***	0.023***	0.017***
	(0.003)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
College avg. freshman SAT score (100s)	0.004***	0.005***	0.004***	0.005***	0.007***	0.005***	0.006***	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Focal group × ...								
College voting rate in 2012	0.017***	-0.001	-0.021***	0.000	0.006***	0.001**	0.004	0.011***
	(0.005)	(0.004)	(0.003)	(0.004)	(0.002)	(0.001)	(0.003)	(0.003)
Avg. freshman SAT score	0.002	0.005***	0.002	0.001	-0.003***	-0.000	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.000)	(0.001)	(0.002)
Non-college goer	-0.093***	-0.092***	-0.093***	-0.092***	-0.093***	-0.092***	-0.098***	-0.053***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Family + cohort FEs	Y	Y	Y	Y	Y	Y	Y	Y
Student characteristics	Y	Y	Y	Y	Y	Y	Y	Y
Portfolio summary statistics	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2,099,366	2,099,366	2,099,366	2,099,366	2,099,366	2,099,366	2,099,366	2,099,366
R^2	0.557	0.557	0.557	0.557	0.557	0.557	0.558	0.558

Note: OLS regressions. Student characteristics include: linear terms in SAT math and verbal scores and categorical indicators of gender, race/ethnicity, parental education, and parental income. Score-send portfolios are the set of colleges to which students sent their SAT score. Regression samples are limited to domestic SAT-taking siblings in the 2005-2012 high school graduation cohorts who had qualifying score-send portfolios. College-clustered standard errors in parentheses (*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$).

Appendix B - Matching Process

In this appendix, we discuss the process of matching the education records with the voting records, and then we assess match quality by comparing the matched data against a nationally representative dataset on voting rates. Since this study employs the same base data as Bell et al. (2024), material in this appendix is largely reproduced from Appendices A and B of Bell et al. (2024) with the authors' permission.

B1. Preparing the Education Data

The College Board data initially consists of 39.6 million observations across cohorts graduating high school between 2004 and 2018. We then restrict the sample to students who attended high school in one of the 50 states or Washington, DC, and those who have complete information on birth date (i.e. month, day, and year). This leaves over 37.6 million observations that serve as the base sample. Our analyses focus on the 21.3 million observations in the 2004 to 2012 cohorts.

B2. Preparing the Voting Data

We begin with a dataset that keeps voters with a birth year after 1984 to be consistent with the education data. We also keep voters where their year of birth is missing, leaving us with 52,921,628 observations.

With this initial voting dataset, the goal is to get one observation per person. Getting one observation per person presents two challenges relative to the education data. First, the DOB may be partially incomplete and in rare cases, entirely missing.[19] In the education data, we removed those observations from the base sample. We cannot drop those voting records because someone with missing DOB may be a match to the education data. The supposed match in the education data will look as if they never voted if we remove those voting observations. Second, the same person can show up more than once across states if they move (and potentially within the same state if a state's records are inexact and/or missing a piece of the birth date).

To address the above issues, we first go from one observation per person per state to one observation per person across states. This involves finding exact matches and then fuzzy matches using name and DOB, all while making sure there is no overlap in the same election, which is indicative of this being more than one person. We also create a series of flags, which we use in robustness tests, when there are common names or missing information on the DOB.

[19] 4.5 percent had a missing birth year, 14.6 percent had a missing birth month, and 21 percent had a missing birth day.

B3. Identifying Voters That Are Not Duplicates

We first identify all voter observations across all states that do not need to be deduplicated and set them aside to what will be the final dataset of deduplicated voters. To do so, we start with all observations with a unique first and last name combination across all states. At the end of this step, 34.4 percent of the voter data are considered unique observations.

With the set of voters with non-unique first and last name combinations, we identify all observations that were unique by first name, last name, and birth year. After this step an additional 21.8 percent of the voter data are considered unique observations. We then identify all observations that were unique by first name, last name, birth year and birth month. At the end of this step, an additional 20.4 percent of the voter data are considered unique observations. Then we do the same by adding birth day, producing an additional 10.9 percent of the voter data considered unique. We do the same thing again with the addition of middle initial. After this step an additional 3.8 percent of the voter data are considered unique. Along each step, we remove the "unique" voter but generate a flag that describes how the uniqueness was determined.

At the end of this process, 91.3 percent of the voter data are considered unique observations.

B4. Identifying Duplicate Voters

With the remaining 8.7 percent of voting data, we identify the observations with the same DOB information, the same exact first and last name, but differing voting history (e.g. one observation voted in an election in 2010 but did not vote in 2012 and the observation with the same name and DOB information had the opposite voting pattern). The vast majority of these observations were collapsed into a single observation and we retained whether we collapsed two or three (Less than 1.1 percent of these observations were unable to be collapsed into a single observation). Almost all were initially two observations that we collapsed into one.

This deduplication process accounted for roughly 7.3 percent of the original voting observations, leaving just 1.5 percent of original voting observations. We take these observations and place them with the uniquely identified and deduplicated observations and treat them as unique.

From here, we created our final dataset to use in the matching to the College Board data. It contains each dataset that from each round of deduplication to identify unique observations and the remaining observations that were not uniquely identified. In all, we created a dataset with 50,975,728 observations of voting history.

After each de-duplication round, in which observations were uniquely identified, we created a flag to identify all the observations that we determined were uniquely identified for that round to use in a series of robustness tests.

B5. Exact Matching

We start by using exact matching of the education data to the voting data. An exact match required perfect agreement between the first name, last name, middle initial, and date of birth (DOB), none of which could be incomplete in either dataset for this step. We identified 14,175,803 unique matches.[20] This implies that 35.5 percent of the education sample was found in the voting records through exact matching.

B6. Fuzzy Matching

After removing any observation that was exactly matched from both the education and voting datasets, we implement a fuzzy matching algorithm, in three broad steps.

First, we use College Board's 23 step fuzzy matching algorithm - a process they use in other applications. The algorithm starts by very slowly loosening the exact matching criteria. Specifically, it starts with exact matches on first name, last name, and DOB but one or both of the datasets are missing a middle initial, everything matches exactly but for one edit to one name, and everything matches but the first and last name are swapped. The most relaxed criteria - the 23rd step - matches exactly on first name, DOB, and gender but one of the two last names from the two datasets is a suffix of the other and only one of the middle initials is missing. After each step, the matched observations are not replaced for additional matches. We also retain each step in which the observations are matched for robustness tests. This process generates 6,547,716 additional matches, which is an additional 16.4 percent of the education observations.

Second, we used the data linkage method (also known as the editing distance method) employed by Dusetzina, Tyree, Meyer, Meyer, Green, & Carpenter (2014). The method calculates a probability that two strings are a match with the following formula:

$$\text{(B1)} \qquad \sum_{i=1}^{2}[1-(length(name_i) * spedis(name_i, name_{-i}))/2400)]/2$$

The two names from each dataset, indexed by i, are compared in both character length and also "spelling distance" (i.e. *spedis*). Spelling distance is a common function in statistical software that compares the letters in the name.[21]

We consider any value greater than 0.95 a high enough probability to be a match. We first do this for first names, maintaining an exact match on last name

[20] 83,652 students were matched to multiple voting records (usually two), creating 193,986 observations. In our main analyses, we randomly choose one of these matches but our results are entirely insensitive to alternatives analyses, largely because this impacted such a small fraction of the sample.

[21] Documentation for *spedis* in SAS, the statistical software we used, can be found here: https://support.sas.com/resources/papers/proceedings/proceedings/sugi25/25/cc/25p086.pdf.

and DOB, and then again for last name. These generate an additional 27,557 matches (one percent of the education data).

Third and finally, we slightly loosen the criteria on birth date by using exact matches on everything previously described but the voting data has a missing birth day or birth month, but not both and not missing year, so there is no conflicting information. This generates 635,344 additional matches (nearly two percent of the education data).

In total, we matched approximately 27.7 million observations, accounting for almost 53.6 percent of the education sample.

Assessing Match Quality

Next, we assess the quality of our match. To do so, we compare voting rates in our matched data for different groups and different elections to reported voting rates in the Voting and Registration dataset from the U.S. Census Bureau.[22] These data come from a supplement to the Current Population Survey. We focus on the voting rates among 18-24 year olds in the 2012 and 2016 elections, a time period that our matched data covers well.

In Appendix Tables B1, B2, and B3, we compare voting rates in the two datasets by election, age, race, sex, state, and sometimes combinations of those variables. Generally speaking, we find lower voting rates in our matched data than in the Census data, but the patterns across subgroups follow one another. The lower voting rates are expected because our matching process is imperfect and somewhat conservative. Moreover, the U.S. Census Bureau's measure of voter turnout is self-reported and, thus, contains some people who reported voting when they actually did not.

Overall, it is reassuring that the relative voting rates across subgroups generally match. Appendix Table B1 shows voting rates in the two datasets by age and sex. The table shows that the voting increases with age (in both elections), according to the Census. Our matched data show a similar pattern, although at lower rates.

Appendix Table B1 also highlights the differences between the two samples by sex. Similar to the Census data, older males vote more than younger males in the matched data. However, this is not true for females, especially in the 2012 election. This is likely because we had difficulty matching to women who change their last name. This also explains why females vote at higher rates than males in both datasets for the youngest cohorts, before women typically get married and change their last name, but not so for older cohorts. For older cohorts, Census data suggest females are more likely to vote than males, but the matched data does not. These facts motivate some robustness tests that focus on males and recent cohorts.

Appendix Table B2 shows voting rates in the two datasets by race and sex. The table shows that Black people are the most likely to vote in 2012 in both elections

[22] https://www.census.gov/topics/public-sector/voting.html

and Asian people are least likely. Similarly, white people are most likely to vote in 2016 in both datasets and Asian people are the least likely. These similar patterns between the datasets is comforting. Finally, voting rates in the two datasets are compared by state. It is immediately clear that our matched sample reflects the voting rates of the Census in some states better than others. This is partially because College Board has relatively low coverage in some states, like Mississippi. However, in states where College Board has substantial coverage, such as Virginia, the two voting rates are well aligned. This motivates a few additional robustness tests, including only using states where College Board has substantial coverage and only using states where the voting rates between these two datasets are well aligned (see Table A.6).

Table B1—Comparison of voting rates in Census data and matched data in 2012 and 2016 elections, by age and sex.

		Voted in 2012 (Percent)			Voted in 2016 (Percent)		
	Census Age (in years)	High School Cohort	Census	Matched Data	High School Cohort	Census	Matched Data
Overall	18	2012	31.3	25.2	2016	32.3	30.0
	19	2011	33.9	25.2	2015	36.9	30.9
	20	2010	39.7	24.7	2014	40.1	31.0
	21	2009	37.5	24.8	2013	38.8	31.5
	22	2008	41.0	26.9	2012	41.4	33.0
	23	2007	40.0	26.7	2011	41.6	33.1
	24	2006	41.5	26.6	2010	43.8	33.5
Male	18	2012	27.8	23.2	2016	30.5	27.1
	19	2011	30.4	22.8	2015	34.4	27.2
	20	2010	34.4	22.7	2014	37.4	27.3
	21	2009	34.3	23.3	2013	35.8	27.9
	22	2008	38.3	26.3	2012	36.3	29.7
	23	2007	39.3	27.1	2011	38.5	30.4
	24	2006	37.1	28.2	2010	41.9	31.6
Female	18	2012	34.8	27.2	2016	34.3	32.8
	19	2011	37.5	27.5	2015	39.5	34.4
	20	2010	44.7	26.5	2014	42.7	34.5
	21	2009	40.9	26.2	2013	42.0	35.0
	22	2008	43.7	27.5	2012	46.1	36.3
	23	2007	40.8	26.4	2011	44.8	35.8
	24	2006	46.0	25.1	2010	45.8	35.2

Note: The U.S. Census data are a nationally representative sample, accessed online here: https://www.census.gov/topics/public-sector/voting.html. The matched data includes College Board test-taker data linked to Data Trust, LLC's national voter records. High school cohort is the year of graduation, comes from College Board data, and is an approximation of age.

Table B2—Comparison of voting rates in Census data and matched data in 2012 and 2016 elections, by race and sex.

	Voted in 2012 (Percent)		Voted in 2016 (Percent)	
	Census	Matched Data	Census	Matched Data
Total	38.0	25.7	39.4	31.8
Male	34.7	24.7	36.5	28.7
Female	41.3	26.7	42.4	34.9
Asian	20.1	17.2	25.2	25.9
Black	45.9	33.5	40.2	29.1
Hispanic	26.7	20.1	27.2	26.8
White	37.8	26.8	41.1	36.4
Male + Asian	18.1	15.3	21.2	21.8
Male + Black	41.4	26.5	36.5	22.1
Male + Hispanic	24.0	17.7	21.8	22.5
Male + White	34.6	28.0	38.3	34.9
Female + Asian	22.1	19.1	29.5	30.0
Female + Black	50.2	39.6	43.7	35.4
Female + Hispanic	29.7	22.3	32.7	30.7
Female + White	41.0	25.7	43.9	37.8

Note: Census voting rates are for 18-24 year-olds; 2012 voting rates in matched data are for the 2006-2012 high school graduation cohorts; 2016 voting rates in matched data are for the 2010-2016 high school graduation cohorts. The U.S. Census data are a nationally representative sample, accessed online here: https://www.census.gov/topics/public-sector/voting.html. The matched data includes College Board test-taker data linked to Data Trust, LLC's national voter records. High school cohort is the year of graduation, comes from College Board data, and is an approximation of age.

Appendix C - Non-Parametric Bias-Corrected Variance

We follow the non-parametric variance component framework in Section 3.1 of Walters (2024). This allows for an arbitrary covariance structure of the estimation errors in the college-value-added estimates.

Let $\hat{\boldsymbol{\alpha}}$ denote the $K \times 1$ vector of college value-added estimates and let V denote the sampling variance-covariance matrix of $\hat{\boldsymbol{\alpha}}$, where the diagonal elements correspond to the squared standard errors of the individual college value-added estimates and the off-diagonal elements may non-zero, allowing for correlated errors across the college value-added estimates. The requires:

$$E[\hat{\boldsymbol{\alpha}} \mid \boldsymbol{\alpha}, V] = \boldsymbol{\alpha}, \tag{D1}$$

and

$$E\left[(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})' \mid \boldsymbol{\alpha}, V\right] = V. \tag{D2}$$

Define the quadratic form

$$Q(\boldsymbol{\alpha}, A) = \boldsymbol{\alpha}' A \boldsymbol{\alpha}, \tag{D3}$$

where the weighting matrix

$$A_0 = \frac{1}{K-1}\left(I_K - \frac{1}{K}\mathbf{1}_K \mathbf{1}_K'\right) \tag{D4}$$

produces the sample variance of the latent college effects. Under this setup, the bias-corrected variance estimator is

$$\hat{Q}_{BC} = Q(\hat{\boldsymbol{\alpha}}, A_0) - \mathrm{tr}(A_0 \hat{V}), \tag{D5}$$

where $\hat{V}$ is an unbiased estimator of the variance-covariance matrix of the estimated college effects and $\mathrm{tr}(\cdot)$ denotes the trace operator. This estimator is unbiased even when estimation errors are heteroskedastic and correlated across colleges.

The corresponding unbiased estimator of the variance of the college effects is:

$$\hat{\sigma}_\alpha^2 = \hat{\boldsymbol{\alpha}}' A_0 \hat{\boldsymbol{\alpha}} - \mathrm{tr}(A_0 \hat{V}). \tag{D6}$$

.

Table B3—Comparison of voting rates in Census data and matched data in 2012 and 2016 elections, by state.

	Voted in 2012 (Percent)		Voted in 2016 (Percent)	
State	Census	Matched Data	Census	Matched Data
Mississippi	62.4	17.3	46.1	18.1
Minnesota	57.0	21.4	49.6	22.5
Wisconsin	53.3	24.7	45.6	21.2
Colorado	52.5	23.0	43.1	25.8
South Carolina	51.3	28.4	42.7	33.9
New Hampshire	50.0	17.8	*	21.1
Oregon	47.6	25.7	45.2	26.4
Iowa	46.6	41.1	35.5	52.9
Massachusetts	45.6	33.5	39.9	42.8
Rhode Island	45.2	27.2	*	35.9
North Carolina	45.1	16.8	44.8	18.5
Ohio	44.7	38.1	39.6	43.1
Michigan	43.5	17.1	36.1	22.3
Maine	42.8	13.7	48.8	14.3
Missouri	42.5	33.2	45.9	43.5
Maryland	42.1	33.3	48.0	40.1
Virginia	42.0	37.1	54.6	42.9
Montana	40.7	33.3	*	40.2
Delaware	40.5	21.0	*	22.0
Louisiana	40.4	11.6	49.2	14.9
Arizona	40.0	12.3	35.8	17.8
Pennsylvania	39.9	27.1	48.7	40.5
Nevada	38.2	27.3	37.6	32.3
Washington	38.2	33.7	42.2	38.7
Georgia	37.9	17.1	40.5	15.6
Kentucky	37.1	28.7	51.1	40.3
New Mexico	37.1	11.3	37.8	12.3
Connecticut	36.9	27.9	37.0	37.7
Florida	36.8	32.0	33.1	36.9
California	36.5	23.0	37.5	33.9
Nebraska	36.2	30.7	50.1	45.2
Indiana	35.7	24.8	40.8	31.9
Alabama	35.3	34.0	41.4	40.0
New York	35.0	27.1	34.6	33.2
New Jersey	34.9	32.4	35.9	38.4
Utah	34.8	28.2	41.6	38.3
Tennessee	34.0	26.9	29.9	32.7
South Dakota	32.7	12.3	*	16.6
Illinois	32.2	29.1	45.3	41.6
Kansas	30.1	26.5	33.8	36.1
Idaho	29.8	9.2	40.7	13.3
Oklahoma	27.2	21.9	32.4	34.2
Arkansas	24.3	26.0	33.1	30.9
West Virginia	22.6	28.1	32.2	38.0
Texas	22.5	19.1	27.3	24.1
Hawaii	22.1	8.0	20.4	10.1
District of Columbia	*	19.9	*	20.9
North Dakota	*	15.0	*	19.2
Alaska	*	12.4	*	15.1
Vermont	*	11.3	*	14.9
Wyoming	*	9.2	*	16.5

Note: Census voting rates are for 18-24 year-olds; voting rates in matched data are for the 2006-2012 high school graduation cohorts. The U.S. Census data are a nationally representative sample, accessed online here: https://www.census.gov/topics/public-sector/voting.html. The matched data includes College Board test-taker data linked to Data Trust, LLC's national voter records. High school cohort is the year of graduation, comes from College Board data, and is an approximation of age. *indicates that Census data are not available.

To implement this bias correction with more than 2,600 colleges, we use bootstrap estimation (with 100 bootstrap samples per regression specification) to recover unbiased estimates of colleges' value-added ($\hat{\boldsymbol{\alpha}}$) and an unbiased estimate of the variance-covariance matrix of the college value-added estimates ($\hat{V}$).