Do Schools Matter? Measuring the Impact of California High Schools on Test Scores and Postsecondary Enrollment*

Scott Carrell[1], Michal Kurlaender[2], Paco Martorell[2], Matthew Naven[3], and Christina Sun[2]

[1]University of Texas at Austin & NBER

[2]University of California, Davis

[3]U.S. Government Accountability Office

October 27, 2025

Abstract

We estimate high school impacts on test score performance, post-secondary enrollment, and the relationship between the two using administrative data on California high school students. We find that models using only standard controls for prior test scores and student demographics are biased (especially for college enrollment), but adding rich controls for peer, neighborhood, and family quality eliminates most of this bias. Our results suggest that there is substantial variation in quality across schools in both test scores and college enrollment. In our preferred (fully saturated) specifications, a one-standard deviation increase in our base school quality measure is associated with a 0.10 standard deviation increase in standardized test scores and a 4.8 percentage point increase in four-year college enrollment. Higher test score value-added schools increase college enrollment across multiple margins – lower-ability students move from no college to two-year colleges while higher ability students move from two-year to four-year colleges. Notably, most of the variation in college enrollment value-added is not explained by test score value-added, suggesting policies that focus solely on test scores miss important components of school effectiveness. Finally, using school-level survey results, we show that value-added models of school quality are highly correlated with measures of school climate, teacher and staff quality, and counseling support within the school.

*This study was conducted in collaboration with the California Department of Education (CDE) with the support of U.S. Department of Education's Institute of Education Sciences Research Grant R305E150006. The findings and views contained in this paper do not necessarily reflect those of the CDE or any other entity providing data for this study. The findings and conclusions in this article are those of the authors and are not purported to represent the views of the U.S. Government Accountability Office.

1 Introduction

The notion that "schools matter" motivates many types of education policies. For instance, policies that expand school choice and that establish school accountability systems emerge from the idea that access to high-quality schools, and the incentives faced by schools, have important consequences for students' educational outcomes. Moreover, the effectiveness of some policies and programs (e.g., school funding, or the choice of curricular materials) likely depends on school quality. The widespread perception of the importance of schools has been the catalyst behind controversial efforts to redraw school attendance zone boundaries and change priorities for assigning students to oversubscribed schools (Goldstein, 2018; Veiga, 2021).

In light of the importance of school quality, a growing literature attempts to estimate effects of schools on student outcomes. A number of recent studies use quasi-experimental variation in cities that use centralized school assignments and lotteries for oversubscribed schools.[1] These studies have made important methodological advances and produced valuable insights on school effects. However, these methods cannot be used to estimate the effectiveness of schools attended by the vast majority of students in the United States since centralized school assignment is rare. While observational 'value-added' models that control for prior test scores have been shown to perform fairly well for estimating effects on test scores (Deming, 2014; Angrist et al., 2017), it has been difficult to credibly estimate school impacts on longer-run outcomes such as college enrollment for which there are no lagged outcomes. Altogether, it remains unclear how school quality varies across the schools most students attend, especially for longer-run outcomes, and how this variation is related to school characteristics.

To fill these gaps in the literature, this paper examines the effects high schools have on test scores and postsecondary outcomes using statewide data from California. We make four primary contributions to the understanding of school quality. First, we assess the validity of value-added estimates of school impacts on college enrollment and show that models using standard controls have substantial bias, but using rich covariates for neighborhood, peer and family environments largely mitigates this bias. Second, we examine the degree to which impacts on test scores "pass-through" to postsecondary enrollment (i.e., the degree to which schools with high test score value-added translate those gains into increased college enrollment) and document important heterogeneity in this pass-through by student ability. Third, we investigate mechanisms for school quality by examining the association between our value-added estimates and student, parent and teacher perceptions from school climate surveys. Finally, by using data on over 1,400 high schools in the largest state in the country, we are able to examine how school quality varies with school characteristics and

[1] Examples include Angrist et al. (2017, 2021); Abdulkadiroğlu et al. (2017, 2022); see Angrist, Hull and Walters (2022) for a review of recent advances in this literature.

across school districts.

In the first part of the paper, we estimate school impacts on test scores and enrollment in either a two- or four-year college by adapting methods developed by Chetty, Friedman and Rockoff (2014*a*) for estimating teacher effects.[2] We begin by estimating a "base" model that controls for pre-high school test scores and student demographics — the standard covariates that have been used to estimate school value-added in other studies (Willms and Raudenbush, 1989; Meyer, 1997; Everson, 2017; Kane and Staiger, 2008; Chetty, Friedman and Rockoff, 2014*a*; Deming, 2014; Angrist et al., 2017). Estimates from this base model serve as a useful benchmark for comparability to other estimates in the literature as well as estimates from our models that control for an increasingly richer set of covariates. Using student name and home address, we construct measures of neighborhood quality, proximity to colleges, and sibling characteristics. We also include school-level averages of these covariates to account for peer characteristics (Altonji and Mansfield, 2018).

Results from our value-added estimates suggest schools make important contributions to both test scores and college going, but the magnitude of these effects is considerably smaller when controlling for the richest set of covariates. In the base model, a one standard deviation increase in estimated value-added is associated with a 0.15 standard deviation increase in student test scores distribution and a 9.9 percentage point increase in four-year college enrollment. In models that control for the full set of covariates (including peer, sibling, and student neighborhood controls), these impacts shrink substantially to 0.10 standard deviations and 4.8 percentage points, respectively. We find that test score value-added is higher in schools located in suburbs and towns relative to larger cities, but test score and college value-added are not strongly correlated with average student demographics. We also find that most of the variation in value-added is across schools within school districts rather than between high- and low-value-added districts.

To evaluate the validity of the value-added estimates, we use the tests proposed in Chetty, Friedman and Rockoff (2014*a*). These include a "forecast bias" test that examines the correlation between estimated value-added and covariates excluded from the value-added model. A key advantage of the rich set of controls we use is that, even in our fully-specified model, we can retain "leave-out" variables to evaluate whether estimated value-added is correlated with the omitted variables. For test scores, when we control for peer means of the controls in the base model, there is only minimal correlation between estimated value-added and neighborhood quality. Furthermore, the correlation between the value-added estimates in models with the base and richest set of controls is about 0.9. These findings are consistent with studies using lottery-based variation that show that observational value-added models eliminate much of the selection bias and

[2] Value-added models were developed by researchers attempting to estimate teacher effectiveness, and a large literature examines the validity of value-added models in this context (see for example Rothstein (2017, 2010); Chetty, Friedman and Rockoff (2014*a*); Bacher-Hicks, Kane and Staiger (2014); Bacher-Hicks et al. (2017)).

are useful for evaluating school performance (Angrist et al., 2017; Deming et al., 2014; Angrist et al., 2021).

In contrast, the validity tests suggest a value-added model with standard covariates performs poorly for postsecondary outcomes. Specifically, the value-added estimates from the base model are strongly correlated with a student's own-neighborhood quality, proximity to the nearest public two- and four-year college, and older-sibling college-going. However, when we control for sibling college going, neighborhood quality and school peer averages of the included controls, we find that proximity to college is nearly uncorrelated with four-year college value-added and only weakly correlated with two-year college value-added. We also find that simply adding peer averages to the base specification alone reduces most of the correlation between estimated value-added and neighborhood quality. These results suggest that using a rich set of covariates in value-added models may be sufficient to generate useful estimates of actual school impacts on postsecondary enrollment.

The second part of the paper examines the link between test score value-added and postsecondary enrollment. In our preferred specification using the full set of controls, we find a one-standard deviation increase in a school's math test score value-added is associated with an increase in four-year college enrollment of 1.7 percentage points, while the association with two-year enrollment is close to zero and negative in some specifications. This latter result appears to be driven by heterogeneity in the relationship between test score value-added and college enrollment by student ability. In particular, higher value-added schools increase college enrollment across multiple margins — lower ability students move from no college to two-year enrollment and higher ability students move from two-year to four-year enrollment. The net result is that test score value-added has a small (possibly negative) association with two-year college enrollment.

While test score value-added does predict college enrollment, most of the variation in a school's college enrollment value-added is unrelated to variation in test score value-added. In our preferred specification (i.e., the model with the full set of controls), less than 10 percent of the variance in four-year enrollment value-added is accounted for by test score value-added. This result is consistent with the multidimensionality of school quality (Beuermann et al., 2023; Abdulkadiroğlu et al., 2020). It also suggests that policies that focus solely on test scores miss important components of school effectiveness (Jennings et al., 2015).

Finally, we examine how our value-added estimates relate to measures of school environments based on surveys of parents, students, and teachers. These analyses shed light on potential mechanisms behind differences in school quality. They also provide additional evidence on the validity of the value-added estimates since the perceptions of schools captured by the survey can be viewed as an alternative indicator of school quality. We find that teacher and staff quality, the level of counseling support, and especially school climate are all correlated with test score and four-year college enrollment value-added. In contrast, value-added on two-year college enrollment is negatively correlated with the survey measures of school quality, a pattern consistent with our findings on the relationship between test score value-added and two-year enrollment.

Our study contributes to a growing literature on estimating school quality (see Angrist, Hull and Walters (2022) for a review of this work). One strand uses quasi-random variation generated by features of centralized school assignment systems such as lotteries for oversubscribed schools (Angrist et al., 2017; Beuermann et al., 2023; Abdulkadiroğlu et al., 2020; Angrist et al., 2021). We build on this work by estimating school quality in the more common settings that do not use centralized school assignment. Although our estimates require stronger assumptions than those based on quasi-experimental variation in school assignments, our approach allows us to investigate how school quality varies across a broader set of schools and to analyze how school quality differs between types of schools and localities. A second strand uses observational value-added methods like we do here (Jackson et al., 2020; Hubbard, 2017). We build on this work by adapting tests from the teacher quality literature to evaluate the validity of the estimates and by showing how this validity hinges on the richness of the control variables used, especially for college enrollment. Our results are therefore consistent with Jackson et al. (2020), who show that value-added models that control for neighborhood and peer quality produce causal estimates of school impacts on college enrollment in Chicago.

This paper also builds on studies that examine school impacts on longer-run outcomes. Several studies have analyzed the relationship between impacts on test scores and impacts on longer-run outcomes such as college enrollment and earnings (Abdulkadiroğlu et al., 2020; Dobbie and Fryer, 2020; Mbekeani et al., 2023; Hubbard, 2017).[3] Jackson et al. (2020) further consider short-run impacts on non-cognitive outcomes (in addition to test scores) and their relationship with longer-run outcomes. Deming et al. (2016) study the effect of school accountability policies on both test scores and longer-run outcomes. Our study contributes to this literature by documenting important heterogeneity in the relationship between test score impacts and college enrollment by student ability and how these relationships differ for two- and four-year college enrollment. A second way we contribute to this work is by demonstrating that value-added models for college enrollment may be misleading if they do not adequately account for peer characteristics as well as neighborhood and family environments.

Finally, we contribute to the literature on the factors that drive school quality. Education scholars have long been interested in the practices that characterize effective schools (see Coleman et al. (1966); Phillips (1997); Purkey and Smith (1983); Sammons et al. (1995) for a review of this research). Several recent studies have examined the practices and attributes associated with positive charter school impacts (Angrist, Pathak and Walters, 2013; Hoxby and Murarka, 2009; Dobbie and Fryer Jr, 2013). A number of studies have used

[3]Several papers have studied the link between teacher impacts on test scores and their impacts on non-test score outcomes (Chetty, Friedman and Rockoff, 2014*b*; Jackson, 2018; Backes et al., 2023).

survey data to generate measures of school contexts and related these to student outcomes (Dobbie and Fryer Jr, 2013; Kraft, Marinell and Shen-Wei Yee, 2016; Bloom et al., 2015; Davis and Warner, 2018). Our analyses relating estimated school impacts to survey-based measures of school contexts focus on a much broader set of schools than the charter and urban schools examined in earlier work. In addition, we examine how measures of school contexts correspond to school impacts on postsecondary schooling outcomes.

The remainder of the paper is organized as follows. Section 2 describes the California data we use in the study. Section 3 discusses the empirical methods we use to estimate school quality and how we test the validity of these estimates. Section 4 presents the value-added estimates, including an examination of heterogeneity in school quality. Section 5 investigates the relationship between test score value-added and college enrollment. Section 6 explores mechanisms for differences in school quality using survey-based measures of school environments. Section 7 concludes.

2 Data

Our dataset consists of four 11th-grade cohorts of students from 2014–15 through 2017–18 who attended California public high schools and took the Smarter Balanced Assessment Consortium (SBAC)[4] tests in mathematics and English Language Arts (ELA). In total there are nearly two million students in our data who attended one of California's roughly 1,400 public high schools.

We place a series of restrictions on the analytic sample similar to those found in other school value-added studies. First, we only include schools in the analysis that serve traditional high school grades (since our empirical approach uses middle school test scores as controls and we do not want the lagged test scores to be from the same school as their 11th-grade school), enroll at least 10 students, and are conventional high schools.[5] Next, we only include students with non-missing high school test scores as well as pre-high school test scores and demographic variables used as controls in the value-added models (e.g., race, economic disadvantage and special education status). For students who repeat 11th grade, we use scores from the earliest instance of that grade. For the pre-high school ELA test scores we use scores from 8th grade.[6] For mathematics, we have to use 6th-grade scores because that is the last grade in which all students took the same mathematics exam.[7] After imposing these restrictions, the final sample used to estimate value-added

[4]Scores on the SBAC tests are part of California's accountability system and all public-school students take these tests in grades 3-8 and grade 11.

[5]Examples of non-conventional schools include special education centers and juvenile court schools.

[6]California did not administer tests in the 2013-2014 school year as they were transitioning to the SBAC from the former California Standardized Test (CSTs) used for accountability, so we use 7th grade ELA scores for the 2016-2017 11th-grade cohort.

[7]For the 2014–15 to 2017–2018 cohorts, the pre-high school math scores come from the test used prior to the adoption of the SBAC test, which had 7th and 8th grade math assessments that were tied to the mathematics course a student was taking. All 6th graders took the same test. However, students take standardized mathematics tests under the SBAC that are directly comparable across students, so we use eighth-grade mathematics test scores for the 2018 cohort.

consists of over 1.2 million students. Online Appendix A provides additional information on the restrictions imposed when creating our samples.

Information on postsecondary enrollment comes from linking students in our sample to college enrollment records from the National Student Clearinghouse (NSC). We focus on initial enrollment following high school and define a student to have enrolled in a college if they are observed attending college in the NSC data within two years of 11th grade (i.e. within one year of graduating high school). The analyses below distinguish between enrollment at a four-year and two-year college, with students enrolling in both a two- and four-year college coded as a four-year college enrollee.

We use student-level home addresses and names to augment these data in two important ways.[8] First, we construct sets of siblings by matching students who share a home address and surname in a given year and then linking all students who share a common sibling. Second, we geocode the student addresses in order to construct measures of college proximity and link students to Census tracts. Specifically, we match a student's address in sixth grade to average Census-tract characteristics, such as income, education, and race. For college proximity, we compute the minimum distance to public two- and four-year colleges in California, which has been shown to be strongly predictive of college enrollment and college choice (Card, 1993; Mountjoy, 2022). As described below, we use these additional sibling, neighborhood, and college proximity measures to conduct validity checks and to estimate models while controlling for peer, neighborhood, and family controls.

Table 1 reports summary statistics for the samples used in the study. The first column contains values for the full population of 11th grade test-takers. The second column has data for the "base" value-added sample with the restrictions described above. The third column is for the subsample with data on older siblings and neighborhood quality. Panel A shows summary statistics for our four cohorts of 11th-grade test takers. With the exception of cohort size, the variables in Panel A are the dependent variable (current z-score) or independent variables included in the "base model" version of equation (1).

Mirroring the demographics of the state of California, students in our sample are racially and economically diverse. Comparing the values in the first and second column indicates that students in the base value-added sample have higher prior test scores and are more socioeconomically advantaged in a number of characteristics. For instance, average math scores in column 2 are 0.16 standard deviations higher than in the full sample. These patterns are accentuated when dropping students missing the sibling and neighborhood quality measures. The fact that our value-added sample is positively selected is not surprising and is similar to the large teacher value-added literature that also relies on linking students to prior test scores (Chetty, Friedman and Rockoff, 2014*a*). In section 4 we discuss how this sample selection affects our estimates by comparing estimates from models using the base and restricted samples.

[8]Student addresses were only available in our test score data during the 2002-2003 to 2012-2013 academic years.

Table 1 Panel B shows summary statistics for postsecondary-enrollment for the students in our base sample who we linked to NSC records.[9] College going rates in California are quite high, with 35 percent of all 11th-grade test takers enrolling in a two-year college the year after scheduled high school graduation and 28 percent attending a four-year college. College-going rates among students we use in the value-added estimation are higher; in the most restricted sample (column 3), 36 percent of students attend a two-year college and 37 percent attend a four-year college.

3 Methods

We use a value-added framework to estimate school effectiveness on shorter-run test score outcomes and longer-run college-going outcomes. Value-added on test scores is important since test scores are the primary metrics used in school accountability systems, while value-added on college-going provides an indication of how a school affects students' longer-run well-being.

3.1 Estimating value-added

To measure each high school's value-added on student's outcomes, we estimate models of the form:

$$Y_{ist} = \phi_0 + \phi_1 X_{ist} + \gamma_t + \underbrace{\lambda_{st} + \xi_{st} + \epsilon_{ist}}_{u_{ist}} \tag{1}$$

where Y_{ist} is student i's outcome (e.g., 11th-grade test score or college enrollment) in school s and year t. X_{ist} is a vector of controls (e.g., prior test scores). The γ_t are year fixed effects for each of the four cohorts in our sample. The error term u_{ist} is comprised of three components: school-by-year value-added λ_{st}, a school-by-year common shock ξ_{st}, and a student-specific random term ϵ_{ist}. The key identification assumption is that, conditional on the controls in X_{ist}, student-level heterogeneity (ϵ_{ist}) is uncorrelated with the school a student attends. The following subsection discusses how we test this assumption.

Our goal is to estimate λ_{st}, which captures a school's contribution to student outcomes. A component of this effect is driven by factors controlled (or at least influenced) by school administrators including teacher quality, class size, and the quantity and quality of school support personnel (e.g., counselors). It also likely includes a component that arises from factors that are less malleable to the decisions of school leaders, including peer composition and neighborhood effects (Altonji and Mansfield, 2018).

We produce estimates of school value-added using different sets of control variables in X_{ist}. In our

[9] The sample size for the postsecondary outcomes is slightly smaller than the sample used to estimate test score value-added because some records were not able to be linked to data from the National Student Clearinghouse.

most parsimonious "base" model, we control for prior test scores, student age, and indicators for gender, race/ethnicity (Hispanic/Latino, white, Asian, black, "other"), economic disadvantage, limited English proficiency, and disability. For prior test scores, we use cubic polynomials in 8th-grade ELA scores and 6th-grade math scores.[10] Our most saturated specification includes controls for peer, neighborhood, and family controls. To control for neighborhood characteristics, we control for characteristics of the student's Census tract, including measures of the income, race/ethnicity, and educational attainment of Census tract residents.[11] To control for family characteristics, we control for indicator variables measuring whether an individual's older sibling enrolled in either a two-year or four-year college. Finally, to control for peer quality at the school, we include school-by-grade-by-year means of all the individual control variables in the model.

An important contribution of this study is our ability to examine estimates from specifications using different sets of controls. The base specification is similar to many value-added models used in the school quality literature and uses variables widely available in district- or state-level administrative datasets. However, it may not account for all sources of bias arising from sorting of students to schools. By assessing the validity of the estimates across specifications, we ascertain whether unbiased estimates can be obtained using standard value-added controls or if richer controls are necessary (or if bias remains even with our most saturated model). If the estimates from the base and fully-saturated model are highly correlated, this would indicate widely-used school value-added models produce similar results to those from models that use controls that may not always be available to policymakers or researchers.

To estimate equation (1), we employ the value-added with drift procedure developed in Chetty, Friedman and Rockoff (2014*a*) for estimating teacher effectiveness.[12] The first step involves estimating equation (1) and computing the residuals (which we refer to as "student performance residuals"). We then collapse the student performance residuals u_{ist} to the school by year level:

$$\begin{aligned} u_{st} &= \frac{1}{N_{st}} \sum_{i=1}^{N_{st}} [\lambda_{st} + \xi_{st} + \epsilon_{ist}] \\ &= \lambda_{st} + \xi_{st} + \frac{1}{N_{st}} \sum_{i=1}^{N_{st}} \epsilon_{ist} \\ &= \lambda_{st} + \xi_{st} + \bar{\epsilon}_{st} \end{aligned} \tag{2}$$

[10]In results available upon request, our findings are virtually unchanged when we estimate IV models as suggested by Kane, Staiger and Johnson (2024) by instrumenting for 8th-grade ELA scores with 7th-grade ELA scores.

[11]Specifically, we matched student addresses in sixth grade to their specific Census tract, which we then link to the American Community Survey (ACS). The variables we include to control for neighborhood characteristics are percent Asian, percent Hispanic, percent black, proportion high-school dropout, proportion with a bachelor's degree or higher, the percent of families with children under the age of 18 in poverty, and median household income.

[12]In practice, Chetty, Friedman and Rockoff (2014*a*) is a reweighted version of Carrell and West (2010).

where N_{st} is the number of 11th-grade students in school s in year t. Under the assumption that ϵ_{ist} is a mean zero error term and that students do not sort to schools on unobservable characteristics, we have that $\mathbf{E}[\epsilon_{ist}|st] = \mathbf{E}[\epsilon_{ist}] = 0$, thus the average student performance residual at each school in each year $\bar{\epsilon}_{st}$ will converge to zero.

In order to reduce the variation due to common shocks ξ_{st}, our value-added estimates are the predicted value from a regression of u_{st} on $\boldsymbol{u}_{st'}$, where $\boldsymbol{u}_{st'}$ is the vector of $u_{st'}$ for all $t' \neq t$. By construction, the common shocks are uncorrelated with school value-added ($cov(\lambda_{st}, \xi_{st'}) = 0$) and the common shocks are assumed to be uncorrelated across time ($cov(\xi_{st}, \xi_{st'}) = 0$). Therefore, this functions as the first stage of an instrumental variables regression in which we isolate variation in λ_{st} while eliminating variation in ξ_{st}. The variation in λ_{st} under this methodology comes from the assumption that school value-added is correlated from year to year ($cov(\lambda_{st}, \lambda_{st'} \neq 0)$), which is likely given that most schools will not experience complete faculty and staff turnover between years. Chetty, Friedman and Rockoff (2014*a*) and Naven (2022*a*) provide additional methodological details.[13]

3.2 Validity of Value-Added Estimates

The validity of the base value-added measures depends on the strong assumption that, after controlling for pre-high school (i.e., lagged) test scores and student demographics, variation in high school outcomes (e.g., test scores or college going) across schools reflects the influence of schools and not omitted variables. This assumption has been criticized in the teacher effectiveness literature by Rothstein (2009, 2010, 2017), though test score teacher value-added models have been shown to perform well when controlling for lagged achievement (Chetty, Friedman and Rockoff, 2014*a*; Kane and Staiger, 2008). Similarly, while recent studies show that school value-added estimates based on observational methods are useful for policymakers, some studies show that these estimates have some bias, even when controlling for lagged achievement (Angrist et al., 2017).

To assess the validity of our school value-added estimates, we follow Chetty, Friedman and Rockoff (2014*a*) and Rothstein (2017) and perform specification and forecast bias tests on our value-added estimates. Specifically, the specification test examines whether a change in estimated school value-added corresponds to a one-for-one change in student outcomes. If instead there is a larger or smaller than one-for-one change in student outcomes associated with a change in school value-added it would suggest that the value-added

[13]Our value-added estimates differ slightly from Chetty, Friedman and Rockoff (2014*a*) and Naven (2022*a*) because our estimates do not include a school fixed effect when estimating equation (1), which would account for potential correlation between school value-added and the demographic characteristics of students. This is because we want to use across-school comparisons when controlling for test score value-added when estimating school value-added on long-run outcomes that operates independently of test scores in Table 7. Test score value-added estimates that include a school fixed effect in equation (1) have a correlation of 0.99 with the value-added estimates used in this paper.

estimates are biased. Below we present specification tests for both our shorter-run outcomes (i.e. test score value-added) as well as our longer-run outcomes (i.e. two- and four-year college going value-added).

In contrast, the forecast bias test examines whether our value-added estimates are correlated with pre-high school measures (e.g. prior test scores) that are not included in our base value-added model.[14] If our school value-added estimates are uncorrelated with these excluded or "hold-out" variables, it suggests that any bias from omitted variables would have to be from factors that are themselves uncorrelated with the hold-out variables. For the forecast bias tests, as with Chetty, Friedman and Rockoff (2014*a*), we start by using (further) lagged ELA scores as our hold-out variable. For test score value-added, we believe this is a reasonable approach since the hold-out variable is a highly correlated lagged dependent variable. However, for our models estimating college-going value-added, an obvious concern is that using lagged test scores as the hold-out variable is a weaker test, since we are no longer using a lagged dependent variable as the hold out.[15] Since there is no lagged outcome for postsecondary enrollment, we use two factors which are strongly predictive of college enrollment as our hold-out variables for college-going value-added. One is geographic proximity to postsecondary education, specifically the linear distance from a student's high school to the nearest public two- and four-year college (Card, 1993; Mountjoy, 2022). The second is indicator variables for whether the student's *older* sibling attended a two- or four-year college.

4 Value-Added Estimates

4.1 Test Score Value-Added

Panels A and B of Table 2 present results for our estimates of ELA and mathematics test score value-added, respectively. Each column represents a different combination of sample and controls, starting with the least restrictive base model to the most restricted model controlling for peer, neighborhood and family controls. Row 1 shows the standard deviation of the estimated test score value-added across our various specifications, which is a measure of variation in schools' contribution to test scores.[16] The standard deviation of test score value-added ($\sigma_{\hat{\lambda}}$) in the base model is 0.146 for ELA and 0.151 for mathematics. These results indicate

[14]The coefficient from the forecast bias test is an estimate of the value of $\frac{cov(\epsilon_{ist},\hat{\lambda}_{st})}{var(\hat{\lambda}_{st})}$.

[15]Tables 2 and 3 show the F-statistic for the hypothesis that coefficients on the leave-out variables added to equation (1) are jointly zero. In all cases, the F-statistics are large and we strongly reject that the leave-out variables have no predictive power for both test scores and postsecondary outcomes.

[16]We report the standard deviation of the shrunken value-added estimates. This underestimates the standard deviation of the unbiased school impacts because these estimates have been shrunk towards the overall mean of zero. However, a direct estimate of the standard deviation of the true school impacts will include the influence of school-year shocks (ξ_{st}), thus overstating the variation in true school impacts. For the preferred specification with the richest set of controls, the standard deviation of the shrunken estimates of school impacts is about 30-40 percent smaller than the estimated standard deviation of the combined true school impact and school-year shocks for test scores, 33 percent smaller for two-year enrollment, and 25 percent smaller for four-year enrollment.

that a one-standard deviation increase in school effectiveness on test scores is associated with an increase in average student test scores of about 0.15 standard deviations of the student standardized test score distribution. These magnitudes are similar to estimates of school effectiveness from other settings.[17]

Column 2, row 1 shows the standard deviation of value-added when restricting the sample to those students who have information on the full set of covariates (specifically older-sibling college-going, additional lagged test scores, and student neighborhood characteristics).[18] In this sample, the standard deviation for both ELA and math value-added declines slightly to 0.122 and 0.127, respectively. The correlations between the full value-added sample and the restricted sample are high: 0.920 for ELA and 0.900 for math, as shown in Figure 1, panels 1a and 1b. Thus, while the restricted sample is considerably more selected than the base sample, both samples produce similar value-added estimates.

Columns 3-5 of Table 2 show results when sequentially adding peers, neighborhoods, and prior test score leave-out control variables. Specifically, column 3 includes average peer characteristics for all of the covariates in the base specification. Column 4 adds both own and average-peer Census tract characteristics. Column 5 adds both own and average-peer 7th-grade ELA scores and sibling college going. The standard deviation of test score value-added continues to shrink with each successive inclusion of additional controls. For the fully-specified model in column 5, the standard deviation in ELA and math value-added are 0.105 and 0.101, or about 15 to 20 percent smaller than the standard deviation of the value-added in the restricted sample and base controls (column 2). Though the variation shrinks, panels 2a and 2b of Figure 2 show that the correlation between value-added from the base model (column 2) and the model with the richest set of controls (column 5) is high — 0.931 for ELA and 0.876 for math.

Row 2 in Table 2 presents results for the specification test across the various samples and controls. For both ELA and mathematics, all the models perform quite well, indicating a change in estimated school test score value-added corresponds to a one-for-one change in student achievement.

Rows 3-4 present results for the forecast bias test on lagged test scores and sibling college going. Reassuringly, our test score value-added estimates have little correlation with these leave-out variables. Specifically, results in column 4 suggest that only 0.9% and 0.4% of the variation in school value-added for ELA and math, respectively, is due to students sorting to schools on unobservable ability as measured by excluded lagged test scores. Furthermore, students with *lower* excluded test scores sort to schools with *higher* esti-

[17] Hubbard (2017) finds that a standard deviation in school value-added corresponds to 0.23 standard deviations in student test scores, Deming (2014) reports standard deviations ranging from about 0.05 to 0.1, Naven (2022*a*) and Naven (2022*b*) report standard deviations between 0.06 and 0.09, and Angrist et al. (2017) report standard deviations ranging from 0.15 to 0.25. They are also similar to those found in the teacher quality literature which show that a one-standard deviation increase in teacher quality is associated with between a 0.10 and 0.20 increase in student achievement (Kane, Rockoff and Staiger, 2008; Chetty, Friedman and Rockoff, 2014*a*).

[18] For brevity, we only show results from our base model and the model with the richest set of controls. In results not shown, but available upon request, we also estimated models for each combination of samples and leave-out controls.

mated value-added, which would bias our results toward zero. Only 0.1% and 0.2% of the variation in school value-added for ELA and math, respectively, is due to sorting on omitted sibling college going.

Finally, rows 5-6 present results when conducting the forecast bias tests using our Census tract neighborhood characteristics and proximity to the nearest two- and four-year college as the leave-out variables. Results indicate these leave-out variables are correlated with the base value-added estimates, particularly math value-added. For math, the results in column 2 indicate that 8.8% of the variation in math value-added is related to omitted neighborhood characteristics and 1.7% is due to college proximity. However, these correlations shrink substantially when controlling for peer averages of the base controls (2.6% and 1.6%, respectively) but remain statistically significant (column 3). For ELA, very little of the variation in value-added is related to neighborhood characteristics or college proximity once we control for peer averages (column 3).

In column 5 we examine the correlation between college proximity and our fully specified value-added model, which additionally controls for peer, sibling, and neighborhood controls. Results show the correlation with college proximity is negligible and no longer statistically significant, highlighting the potential importance of controlling for peer, family and neighborhood characteristics in value-added models used for school accountability purposes. Column 6 shows that the results from the specification test and the standard deviation of value-added are unchanged when controlling for college proximity.

4.2 College Going Value-Added

Panels A and B of Table 3 present results for our estimates of two- and four-year college-going value-added, respectively. Similar to the previous section, each column represents a different combination of sample and/or controls. Results in column 1, for the base model, show there is substantial variation in college-going value-added across schools. The standard deviation of the value-added estimates (in percentage points) indicate that a one-standard deviation increase in school effectiveness is associated with an increase in students' enrolling either a two or four-year college by nearly ten percentage points.[19]

Similar to our test score results, as shown in Table 2, when restricting the sample in column 2, the distribution of the base value-added estimates shrinks slightly for both two- and four-year value-added (0.088 and 0.091). Likewise, as shown in Figure 1, panels 1c and 1d, the correlation between estimates when changing the sample is high (0.896 and 0.908), further demonstrating that the restricted and base samples yield similar value-added estimates. Results in columns 3-5 of Table 3 show the standard deviation of college-going value-added shrinks substantially with the inclusion of controls for peers, neighborhoods, and sibling

[19]Note that these results come from separate models, each using a different indicator of postsecondary enrollment as the dependent variable. 2-year and 4-year college enrollment are mutually exclusive such that students who have enrollment records at both types of universities are coded as attending a 4-year university.

college going. For the fully-specified model in column 5, the standard deviation in two- and four-year value-added is nearly cut in half relative to the base model to 0.047 and 0.048, respectively. Figure 2 panels 2c and 2d shows the relationship between value-added obtained using the base controls (i.e., the specification in column 2 of Table 2) and the fully-specified model. The correlation between the two sets of estimates is 0.757 for 2-year enrollment and 0.798 for 4-year enrollment. Though positive, these correlations are weaker than our test score value-added results, further suggesting that controlling for peer and neighborhood controls affects the value-added estimates more for postsecondary enrollment than for test scores.

Row 2 in Table 3 presents results for the specification test across the various samples and controls. For both two- and four-year value-added, all the models perform fairly well, indicating a change in estimated college-going value-added corresponds to a one-for-one change in college enrollment.

Rows 3-4 present results for the forecast bias test on both lagged test scores and sibling college going. Results show both our two- and four-year value-added estimates are virtually uncorrelated with leave-out test scores. We find a statistically significant, but relatively small, correlation between these value-added estimates and our leave out for older-sibling college-going. Specifically, results in column 4 suggest that 2.0% and 2.7% of the variation in two- and four-year value-added is due to students sorting to schools on unobservable ability as measured by sibling college-going.

Rows 5-6 present results when conducting the forecast bias tests using our neighborhood characteristics and proximity to the nearest two- and four-year colleges as the leave-out variables. The magnitudes of the correlations are high for both sets of leave-out variables, even when controlling for peer characteristics in column 3 for our two-year (5.0% and 4.7%, respectively) and four-year (8.2% and 3.6%, respectively) value-added estimates. However, similar to our test score estimates, results in column 5 show the correlations with our college proximity leave-out variables are substantially reduced when controlling for peer, family and neighborhood controls. Though some bias remains for two-year value-added (3.1%), the correlation with four-year value-added is only 1.1% and marginally statistically significant. These results underscore the importance of controlling for peer, family and neighborhood characteristics in value-added models for longer-run outcomes for which lagged outcomes are not available. At the same time, the weak correlation between value-added estimates obtained from the fully-specified model and proximity to college, especially for four-year college value added, suggests that the models with the richest set of controls are useful measures of schools' impacts on postsecondary enrollment.

4.3 Value-Added Heterogeneity

The preceding results indicate that high schools differ considerably in their impacts on test scores and college enrollment. We now examine whether value-added heterogeneity is systematically related to school characteristics such as student demographics, geographic location, and size. Table 4 shows correlations between estimated value-added and school characteristics. For test scores, value-added is higher for schools located in suburbs and towns (relative to larger cities) and also for schools serving fewer students. Schools serving higher rates of minority or low-income students have lower estimated math value-added when using only the base controls (Panel A), but test score value-added is not related to school demographics when generated from models that use the full set of controls (Panel B).

For college enrollment, we find four-year enrollment value-added is higher in smaller schools, but the magnitude is small (a 10 percent increase in enrollment is associated with a reduction in four-year enrollment value-added of less than 0.1 percentage points). There is no evidence that college enrollment value-added is related to student demographics or geography in models that use the full set of controls.

We also examine whether high value-added schools are concentrated in certain school districts. To do so, we examine whether variation in school value-added is primarily an across- or within-district phenomenon. Table 5 shows the R^2 from a regression of estimated value-added on school district fixed-effects. For estimates from models with the full set of controls (Panel B), about 40 percent of the variance in test score and four-year enrollment value-added occurs between districts. In the largest 25 districts in the states, about one-quarter of the variance in value-added is between districts. Overall, these results suggest that school quality varies systematically across districts, but that most of the variation in school quality occurs within school districts.

5 Relationship Between Test Score Value-Added and College Enrollment

The findings in the previous section show that schools have sizable impacts on both test scores and college enrollment. An important question is whether schools that are effective at improving test scores also improve college enrollment. To address this question, we start by examining the raw correlation between college-enrollment value-added and test-score value-added in Figures 3 and 4. As shown in panels 3a, 3b, 4a, and 4b, there is weak negative relationship (correlations range between -0.047 and 0.005) between both a school's ELA and math value-added and their two-year enrollment value-added. However, as shown in panels 3c, 3d, 4c, and 4d, there is a positive relationship (correlations between 0.127 and 0.303) between test-score value-added and four-year enrollment value-added. These results indicate that test score value-added is predictive of college value-added, but also that this relationship differs not just in magnitude but in sign for two- and four-year college enrollment. We now explore these relationships in greater depth.

5.1 Test-score value-added "pass-through"

Next, to examine the degree to which test score value-added is associated with postsecondary enrollment, we employ techniques first developed by Jacob, Lefgren and Sims (2010) and Carrell and West (2010) to estimate the "pass-through" of test score valued-added to college enrollment. Specifically, we consider the following value-added model in equation (3):

$$\begin{aligned} Y_{ist} &= \alpha_0 + \alpha_1 X_{ist} + \gamma_t + \underbrace{\rho\lambda_{st} + \beta_{st} + \theta_{st} + e_{ist}}_{\nu_{ist}} \\ &= \alpha_0 + \alpha_1 X_{ist} + \gamma_t + \underbrace{\pi_{st} + \theta_{st} + e_{ist}}_{\nu_{ist}} \end{aligned} \tag{3}$$

where Y_{ist} is a student i's longer-run outcome who attended high school s in year t. X_{ist} is the same vector of demographic controls as in equation (1) and γ_t are year fixed effects. The error term ν_{ist} is comprised of four components: the pass through of test-score value-added $\rho\lambda_{st}$, the school's contribution to longer-run outcomes that is orthogonal to its contribution to test score gains β_{st}, a school-by-year common shock θ_{st}, and a student-level noise term e_{ist}. The parameter $\pi_{st} \equiv \rho\lambda_{st} + \beta_{st}$ is the contribution of school s in year t to postsecondary schooling.

The primary parameter of interest is ρ, which measures the relationship between a school's contribution to 11th grade test score gains and postsecondary schooling outcomes. In other words, ρ reflects the extent to which test score value-added passes through to long-run outcomes. When ρ is large and positive this indicates schools that generate sizable test score gains also tend to induce significant numbers of students to enroll in college. Knowing the sign and magnitude of this parameter is of interest, for example, for school accountability programs that rely mainly on test performance to evaluate schools.

Estimates of ρ are shown in Table 6.[20] These estimates are scaled so that they reflect the percentage point difference in college enrollment associated with a one-standard deviation change in the shrunken estimates of test score value-added.[21] The first two rows present results for the pass through of ELA value-added, the second two rows present results for the pass through of math value-added, and each column represents a different combination of outcomes, samples and controls. Columns 1-4 show results for two-year enrollment outcomes and columns 5-8 show results for four-year enrollment outcomes.

[20]To estimate ρ we estimate equation (3) by regressing Y_{ist} on X_{ist}, γ_t, and the estimated test score value-added $\hat{\lambda}_{st}$.
[21]This approach is a reweighted equivalent of the methodologies employed by Jacob, Lefgren and Sims (2010) and Carrell and West (2010).

The pattern of results suggest several important findings. First, consistent with results in Figures 5 and 6, there is, if anything, a negative relationship between test score value-added and two-year college enrollment. In fact, results in column 4, indicate that a one standard deviation increase in math test score value-added is associated with a statistically significant 0.6 percentage point decrease in the probability of students enrolling in a two-year college. Second, there is a strong positive relationship between test score value-added and four-year college enrollment. For ELA, the estimated ρ of 0.012 in column 8 indicates that schools one-standard deviation above the mean in ELA test score value-added (i.e., test scores are improved by roughly 0.105 standard deviations) increase four-year college enrollment by roughly 1 percentage point (3.2 percent). For math, the estimates of ρ are somewhat larger (1.7 percentage points, or 4.5 percent, in column 8).[22] Both of these estimates are smaller but roughly comparable in magnitude to that reported by (Jackson et al., 2020).[23] Finally, the estimates of ρ for four-year enrollment become smaller in magnitude when using value-added estimates obtained from models with richer controls This suggests that failure to adequately control for sorting of students to schools might lead to overstated estimates of the pass-through of test score value-added to long-run outcomes.

At first glance, the negative pass through of math test score value-added on two-year college enrollment is somewhat surprising. To further understand the relationship between test score value-added and two-year college enrollment, we conduct two additional analyses. First, we estimate the pass through of test score value-added on *any* college enrollment (i.e., two-year or four-year). The estimates of ρ are 0.010 ($SE = 0.002$) for both math and ELA, indicating overall college enrollment is higher in schools with higher test score value-added.

Second, to get a better understanding of how test score value-added influences the various margins of college enrollment (e.g., no college vs. two-year vs. four-year), we examine heterogeneity in our estimates of ρ across students based on pre-high school student achievement. To do so, we re-estimate equation (3) including interactions between $\hat{\lambda}_{st}$ and decile of a student's pre-high school test score decile. We do these analyses separately for prior math and ELA test scores.[24] Results from this exercise are presented in Figures 5 and 6 for both ELA and math value-added. For two-year enrollment, higher test score value-added schools *increase* two-year enrollment for lower ability students and, for math, *decrease* two-year enrollment for higher ability students. In contrast, higher test score value-added schools increase the probability of four-year college

[22] In results not reported, we show that when we estimate pass through of both math and ELA value-added in the same regression, the pass through of math value-added dominates ELA value-added, as the coefficient on ELA value-added is small and not statistically different than zero. The graphical evidence in Figures 3 and 4 also shows the positive association between test score and college enrollment value-added.

[23] Jackson et al. (2020) find that a one-standard deviation increase in test score value-added is associated with a 2.1 percentage point (6 percent) increase in four-year college enrollment (Table 3).

[24] We use 6th grade scores for math and 8th grade scores for ELA. Deciles are based on the statewide distribution of all test scores in a particular grade-year. In results not reported, we obtain very similar estimates to the "restricted sample, full controls" specification when we also include distance to college as a covariate.

enrollment for students of all abilities, though the effects are largest in the middle and upper end of the ability distribution. Combined, these heterogeneity results suggest that higher test score value-added schools likely improve college enrollment outcomes on both margins. That is, the lowest ability students are moved from no enrollment to two-year college enrollment, while students in the middle and upper end of the ability distribution are moved from two-year to four-year college enrollment.[25]

5.2 Contribution of Test Score Value-Added to College Enrollment Value-Added

The preceding results indicate that schools vary substantially in quality as measured by both standardized test scores and college enrollment, and that schools that generate larger test score gains also tend to generate better longer-run outcomes. An important remaining question is how much of the variation in long-run valued added is explained by the pass through of test score value-added?

We use two approaches to answer this question. First, we compare the variance in college enrollment value-added obtained by estimating equation (1) with the full controls to that obtained by estimating equation (3) with our full set of controls while also including math and ELA test score value-added as additional controls. The ratio of these variances provides a measure of how much of the value-added in college enrollment remains after accounting for the correlation between test score value-added and college enrollment.

Results in panel A of Table 7 for two-year enrollment are consistent with previous findings and indicate that test score value-added pass through accounts for virtually none of the variance in two-year enrollment value-added. Results in panel B for four-year enrollment indicate that between 17 and 27 percent of the variance in four-year college enrollment value-added is explained by the pass through of test score value-added. As such, these findings indicate that, despite test score value-added being correlated with four-year college enrollment, most of the variation in school effects on postsecondary enrollment is due to factors orthogonal to schools' effects on test scores.

As an alternative approach, we also report the unexplained variation $(1-R^2)$ in a school-level regression of college-enrollment value-added on both ELA and math value-added.[26] These results are consistent with our previous findings and again show that test-score value-added is largely uncorrelated with two-year enrollment value-added (panel A) but that between 8 and 16 percent of the of the variation in four-year enrollment value-added is explained by test-score value-added.

[25] In Online Appendix C, we present results examining heterogeneity in test score value-added pass-through across income, race/ethnicity, and gender. Results indicate remarkably consistent pass-through on four-year enrollment across all of these characteristics. For two-year enrollment, results show test score value-added pass-through is negatively correlated with family income.

[26] We weight these regressions by the number of students that contributed to the value-added estimates.

6 Mechanisms: Correlation between Value-Added and Student, Teacher, and Parent Perceptions

Our findings indicate there is substantial variation across schools in both shorter-run test score outcomes as well as longer-run college going outcomes. Additionally, our results show that schools that are better at improving test scores, particularly in mathematics, also improve college enrollment outcomes. Given this, a natural question is "what exactly do higher value-added schools do to improve test scores and/or college outcomes"? To date, relatively little is known regarding the mechanisms that drive school quality. In related work to ours, Naven (2022*a*) shows that school and district-level finance and staffing variables such as school spending or teacher to student ratios are largely uncorrelated with school value-added. However, some of the best evidence to date regarding the effectiveness of various school inputs comes from examining charter schools in Boston and New York. For example, Angrist, Pathak and Walters (2013) find that large test score gains found in Boston urban charter schools are primarily driven by the "No Excuses" model, which emphasizes school discipline and longer school days. Likewise, Hoxby and Murarka (2009) find in New York charters that more effective schools tend to have longer school years, more time devoted to ELA instruction, a "small rewards/small penalties disciplinary policy", and teacher pay which provides performance incentives.

To get a better understanding of the potential mechanisms driving our results, we examine the relationship between our value-added estimates and school-level survey responses from student, parent, and teacher surveys from the California School Climate, Health, and Learning Surveys (CalSCHLS).[27] We use these data to generate three school-level average indices measuring (1) school climate, (2) teacher and staff quality, and (3) counseling supports.[28] The "school climate" index measures perceptions about the overall environment of the school — e.g., whether it is welcoming, has a good learning environment, and whether students have a sense of belonging to the school. The teacher- and staff-quality index captures beliefs about whether school staff provide rigorous, high-quality instruction, and whether the staff care about students. The counseling index reflects perceptions of the adequacy of counseling services for students' socioemotional needs and college planning.[29]

[27] Jackson et al. (2020) estimate value-added models using outcomes derived from surveys similar to the CalSCHLS. That approach is not feasible here because the CalSCHLS data we use cannot be linked at the individual level to CDE records. For more information on the surveys see: `https://calschls.org/survey-administration/downloads/#staff_csss`.

[28] Our indices are averages across several questions within each category. The wording of the survey items used to construct these indices is listed in Online Appendix B:. Because the survey is not administered in every year at every school, we take school-level averages across all years of the survey from 2017-19, and standardize these so each index has a standard deviation of one.

[29] The Cronbach's Alpha values are 0.94 and 0.90 for the school climate and the teacher and staff quality indices, respectively, indicating they have high reliability. The counseling support index has a lower alpha (0.66), likely reflecting that it consists of fewer questions (4) than the school climate index (20 questions) and the teacher- and staff-quality index (17 questions).

We relate these indices to value-added by estimating the following equation (4):

$$\hat{\lambda}_s = \phi_0 + \phi_1 X_s + \phi_2 \hat{Index}_s + \epsilon_s \tag{4}$$

Where $\hat{\lambda}_s$ is a value-added estimate (e.g., ELA, math, two-year, or four-year enrollment), X_s is a vector of mean school-level characteristics[30], and $\hat{Index}_s$ is a vector of estimated indices measuring school climate, teacher and staff quality, and counseling support. We estimate models with each index separately as well as a specification that includes all three indices in the same regression.

Results from this exercise are shown in Table 8. Panel A presents results when regressing estimated value-added on each index measure separately, while Panel B presents results when including all three indices in the same regression. For all results we use estimates from our preferred, fully specified, value-added specification that includes sibling, peer, and neighborhood controls.

Starting with Table 8 Panel A, the pattern of results suggest several important findings. First, all three index measures are positively correlated with our test score value-added measures, and school climate and teacher/staff quality are positively correlated with four-year college enrollment value-added. Second, two-year college enrollment value-added only has a weak correlation with our indices. This result provides a validation of our previous findings in which test score value-added had a weak correlation with two-year enrollment value-added.

To further explore the relative importance of each index measure, in Table 8 Panel B we present results when regressing value-added on all three indices in the same regression. Confirming our previous findings and largely consistent with Angrist, Pathak and Walters (2013); Hoxby and Murarka (2009), we find that our index of school climate is the school-level factor most associated with increases in student learning. In the multivariate regression with all three survey indices, we find school climate to be significantly positively correlated with schools' value-added for test score outcomes as well as schools' value-added for four-year college enrollment. In contrast, counseling support and teacher/staff quality are uncorrelated with value-added.

7 Conclusion

This paper examines high school quality in California. We use rich data from California that links K-12 student records to data on college enrollment to examine high schools' contribution to test score performance, postsecondary schooling outcomes, and the relationship between the two. We estimate "value-added" models

[30] We include the means of all baseline demographic characteristics and prior test scores included in our value-added models.

by adapting the procedure Chetty, Friedman and Rockoff (2014*a*) use to estimate teacher effects. We find that value-added models using only standard controls for prior test scores and student demographics are biased, particularly for college enrollment, but that adding rich controls for peer, neighborhood, and family quality eliminates most of this bias.

We additionally examine the link between a school's impact on test scores and its effects on post-secondary enrollment to measure the extent to which test score gains "pass through" to longer-run outcomes. We also estimate value-added models using postsecondary schooling as the outcome to measure a school's effect on college-going. Finally, using results from student and parent surveys, we explore potential mechanisms.

Our results suggest that schools make important contributions to both test scores and college enrollment. A one standard deviation increase in our preferred model of school effectiveness is associated with roughly a 0.10 standard deviation increase in test scores and a 4.8 percentage point increase in the probability of enrolling in a four-year college. We also find that test score impacts are strongly related to improved college going. An one-standard deviation increase in a school's math test score impact is associated with a roughly 1.7 percentage point increase in four-year college attendance.

Importantly, we find that higher test score value-added schools increase college enrollment across multiple margins — lower ability students move from no college to two-year enrollment and higher-ability students move from two-year to four-year enrollment. Using student and parent survey data, we also find that higher value-added schools are associated with better perceptions of school climate, teacher and staff quality, and counseling support.

References

Abdulkadiroğlu, Atila, Joshua D Angrist, Yusuke Narita, and Parag A Pathak. 2017. "Research design meets market design: Using centralized assignment for impact evaluation." *Econometrica*, 85(5): 1373–1432.

Abdulkadiroğlu, Atila, Joshua D Angrist, Yusuke Narita, and Parag A Pathak. 2022. "Breaking ties: Regression discontinuity design meets market design." *Econometrica*, 90(1): 117–151.

Abdulkadiroğlu, Atila, Parag A Pathak, Jonathan Schellenberg, and Christopher R Walters. 2020. "Do parents value school effectiveness?" *American Economic Review*, 110(5): 1502–39.

Altonji, Joseph G, and Richard K Mansfield. 2018. "Estimating group effects using averages of observables to control for sorting on unobservables: School and neighborhood effects." *American Economic Review*, 108(10): 2902–46.

Angrist, Joshua D, Parag A Pathak, and Christopher R Walters. 2013. "Explaining charter school effectiveness." *American Economic Journal: Applied Economics*, 5(4): 1–27.

Angrist, Joshua D, Peter D Hull, Parag A Pathak, and Christopher R Walters. 2017. "Leveraging lotteries for school value-added: Testing and estimation." *The Quarterly Journal of Economics*, 132(2): 871–919.

Angrist, Joshua, Peter Hull, and Christopher R Walters. 2022. "Methods for Measuring School Effectiveness."

Angrist, Joshua, Peter Hull, Parag A Pathak, and Christopher Walters. 2021. "Credible school value-added with undersubscribed school lotteries." *Review of Economics and Statistics*, 1–46.

Bacher-Hicks, Andrew, Mark J Chin, Thomas J Kane, and Douglas O Staiger. 2017. "An evaluation of bias in three measures of teacher quality: Value-added, classroom observations, and student surveys." National Bureau of Economic Research.

Bacher-Hicks, Andrew, Thomas J Kane, and Douglas O Staiger. 2014. "Validating teacher effect estimates using changes in teacher assignments in Los Angeles." National Bureau of Economic Research.

Backes, Ben, James Cowan, Dan Goldhaber, and Roddy Theobald. 2023. "How to Measure a Teacher: The Influence of Test and Nontest Value-Added on Long-Run Student Outcomes." *CALDER Working Paper No. 270-0423-2*.

Beuermann, Diether W, C Kirabo Jackson, Laia Navarro-Sola, and Francisco Pardo. 2023. "What is a good school, and can parents tell? Evidence on the multidimensionality of school output." *The Review of Economic Studies*, 90(1): 65–101.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen. 2015. "Does management matter in schools?" *The Economic Journal*, 125(584): 647–674.

Card, David. 1993. "Using geographic variation in college proximity to estimate the return to schooling."

Carrell, Scott E, and James E West. 2010. "Does professor quality matter? Evidence from random assignment of students to professors." *Journal of Political Economy*, 118(3): 409–432.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014*a*. "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates." *The American Economic Review*, 104(9): 2593–2632.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014*b*. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *The American Economic Review*, 104(9): 2633–2679.

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York. 1966. *Equality of Educational Opportunity.* Washington, D.C.:U.S. Department of Health, Education, and Welfare, Office of Education.

Davis, Jonathan Ryan, and Nathan Warner. 2018. "Schools matter: The positive relationship between New York City high schools' student academic progress and school climate." *Urban Education*, 53(8): 959–980.

Deming, David J. 2014. "Using school choice lotteries to test measures of school effectiveness." *The American Economic Review*, 104(5): 406–411.

Deming, David J, Justine S Hastings, Thomas J Kane, and Douglas O Staiger. 2014. "School choice, school quality, and postsecondary attainment." *American Economic Review*, 104(3): 991–1013.

Deming, David J, Sarah Cohodes, Jennifer Jennings, and Christopher Jencks. 2016. "School accountability, postsecondary attainment, and earnings." *Review of Economics and Statistics*, 98(5): 848–862.

Dobbie, Will, and Roland G Fryer. 2020. "Charter schools and labor market outcomes." *Journal of Labor Economics*, 38(4): 915–957.

Dobbie, Will, and Roland G Fryer Jr. 2013. "Getting beneath the veil of effective schools: Evidence from New York City." *American Economic Journal: Applied Economics*, 5(4): 28–60.

Everson, Kimberlee C. 2017. "Value-added modeling and educational accountability: are we answering the real questions?" *Review of Educational Research*, 87(1): 35–70.

Goldstein, Dana. 2018. "New York's Schools Chancellor Is Talking About Integration. Can He Make It Happen?" *The New York Times*.

Hoxby, Caroline M, and Sonali Murarka. 2009. "Charter schools in New York City: Who enrolls and how they affect their students' achievement." National Bureau of Economic Research.

Hubbard, Daniel. 2017. "More Gains Than Score Gains? High School Quality and College Success."

Jackson, C. Kirabo. 2018. "What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes." *Journal of Political Economy*, 126(5): 2072–2107.

Jackson, C. Kirabo, Shanette C. Porter, John Q. Easton, Alyssa Blanchard, and Sebastián Kiguel. 2020. "School Effects on Socioemotional Development, School-Based Arrests, and Educational Attainment." *American Economic Review: Insights*, 2(4): 491–508.

Jacob, Brian A, Lars Lefgren, and David P Sims. 2010. "The persistence of teacher-induced learning." *Journal of Human resources*, 45(4): 915–943.

Jennings, Jennifer L, David Deming, Christopher Jencks, Maya Lopuch, and Beth E Schueler. 2015. "Do differences in school quality matter more than we thought? New evidence on educational opportunity in the twenty-first century." *Sociology of Education*, 88(1): 56–82.

Kane, Thomas J, and Douglas O Staiger. 2008. "Estimating teacher impacts on student achievement: An experimental evaluation." National Bureau of Economic Research.

Kane, Thomas J, Douglas O Staiger, and Brian D Johnson. 2024. "Why Does Value-Added Work? Implications of a Dynamic Model of Student Achievement." *NBER Conference Paper*.

Kane, Thomas J, Jonah E Rockoff, and Douglas O Staiger. 2008. "What does certification tell us about teacher effectiveness? Evidence from New York City." *Economics of Education review*, 27(6): 615–631.

Kraft, Matthew A, William H Marinell, and Darrick Shen-Wei Yee. 2016. "School organizational contexts, teacher turnover, and student achievement: Evidence from panel data." *American Educational Research Journal*, 53(5): 1411–1449.

Mbekeani, Preeya P, John Papay, Ann Mantil, and Richard J Murnane. 2023. "Understanding High Schools' Effects on Longer-Term Outcomes."

Meyer, Robert H. 1997. "Value-added indicators of school performance: A primer." *Economics of education Review*, 16(3): 283–301.

Mountjoy, Jack. 2022. "Community colleges and upward mobility." *American Economic Review*, 112(8): 2580–2630.

Naven, Matthew. 2022*a*. "Human-Capital Formation During Childhood and Adolescence: Evidence from School Quality and Postsecondary Success in California."

Naven, Matthew. 2022*b*. "Within-School Heterogeneity in Quality: Do Schools Provide Equal Value Added to All Students?"

Phillips, Meredith. 1997. "What Makes Schools Effective? A Comparison of the Relationships of Communitarian Climate and Academic Climate to Mathematics Achievement and Attendance During Middle School." *American Educational Research Journal*, 34: 633–662.

Purkey, Stewart C, and Marshall S Smith. 1983. "Effective schools: A review." *The elementary school journal*, 83(4): 427–452.

Rothstein, Jesse. 2009. "Student sorting and bias in value-added estimation: Selection on observables and unobservables." *Education finance and policy*, 4(4): 537–571.

Rothstein, Jesse. 2010. "Teacher quality in educational production: Tracking, decay, and student achievement." *The Quarterly Journal of Economics*, 125(1): 175–214.

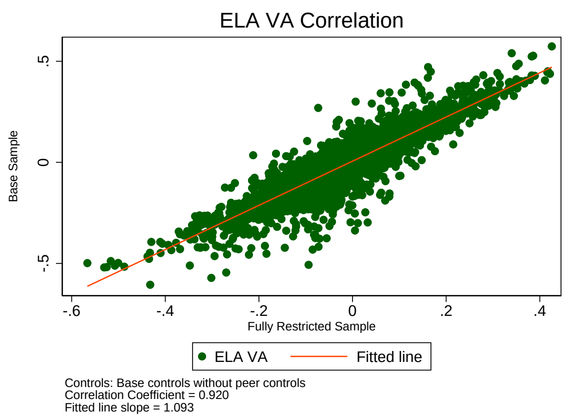
Rothstein, Jesse. 2017. "Measuring the impacts of teachers: Comment." *American Economic Review*, 107(6): 1656–1684.

Sammons, Pam, et al. 1995. *Key characteristics of effective schools: A review of school effectiveness research.* ERIC.

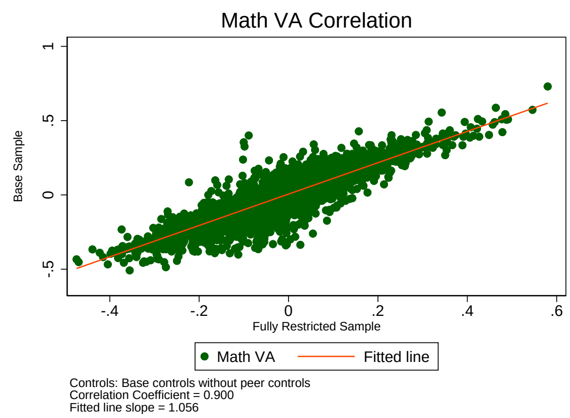
Veiga, Christina. 2021. "NYC reconsiders elimination of high school attendance zones, borough admissions priorities." *Chalkbeat New York*.

Willms, J Douglas, and Stephen W Raudenbush. 1989. "A longitudinal hierarchical linear model for estimating school effects and their stability." *Journal of educational measurement*, 26(3): 209–232.

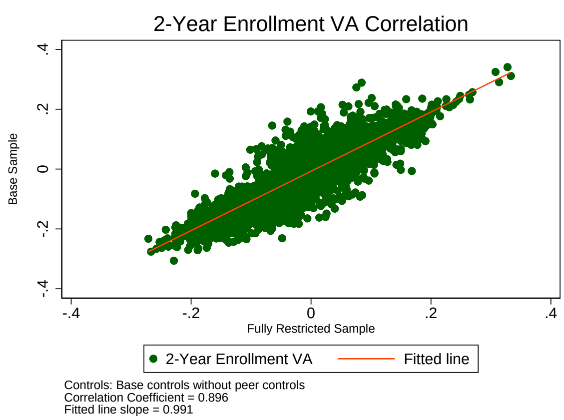
Figure 1: Value-Addded Correlations: Base Controls, Base vs. Restricted Sample



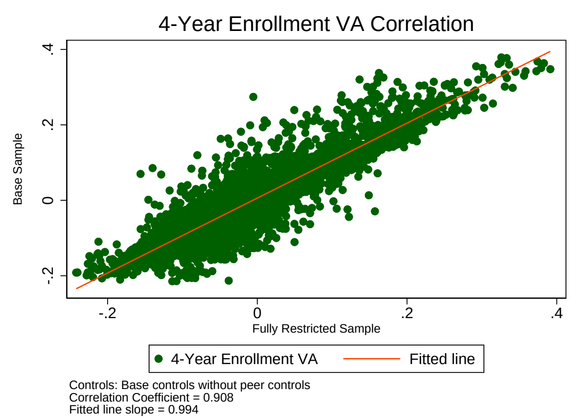
(a) ELA



(b) Math



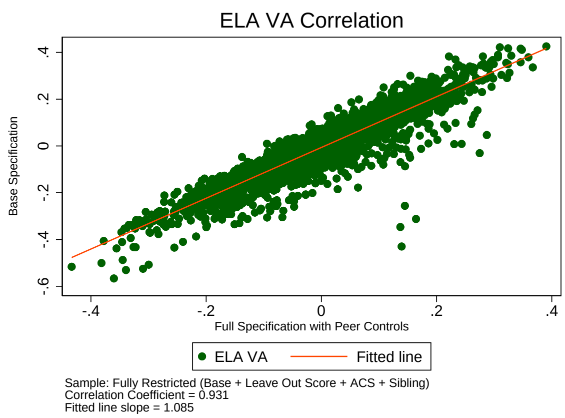
(c) 2-Year Enrollment



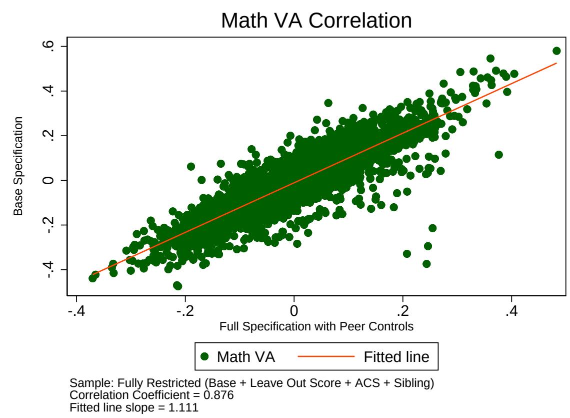
(d) 4-Year Enrollment

Each data point represents a high school's value-added in a given year for two different samples. Value-added is estimated using the base set of controls described in equation (1) for both samples. The vertical axis gives value-added estimates for the base value-added sample. The horizontal axis gives value-added estimates for the restricted value-added sample that can be matched to peer, neighborhood, and sibling characteristics. Data comes from public schools in the state of California between the 2014-2015 and 2016-2018 school years.

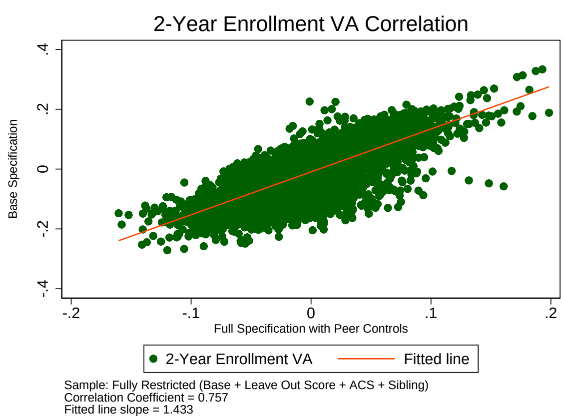
Figure 2: Value-Added Correlations: Restricted Sample, Base vs. Full Controls



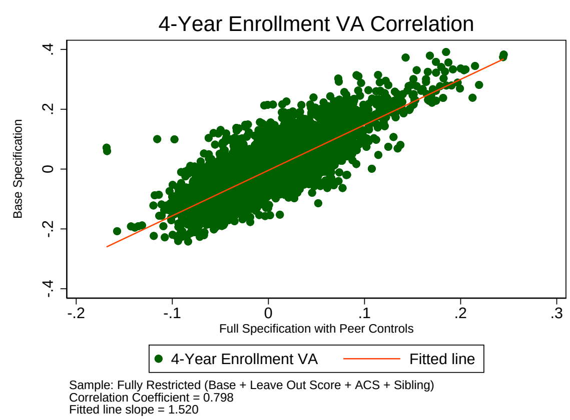
(a) ELA



(b) Math



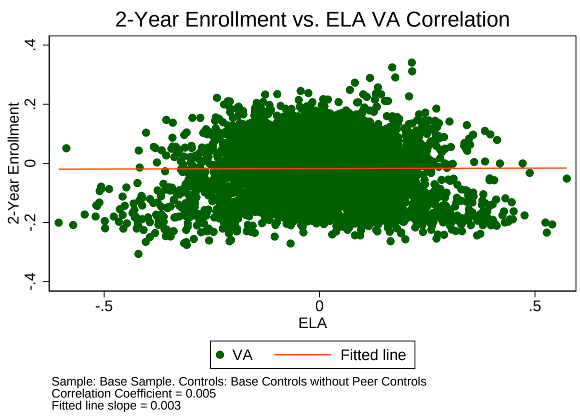
(c) 2-Year Enrollment



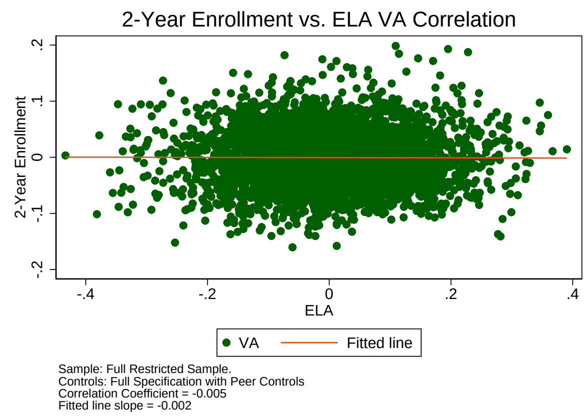
(d) 4-Year Enrollment

Each data point represents a high school's value-added in a given year for two different specifications. Value-added is estimated using the restricted value-added sample that can be matched to peer, neighborhood, and sibling characteristics for both specifications. The vertical axis gives value-added estimates controlling for the base value-added controls described in equation (1). The horizontal axis gives value-added estimates that additionally control for peer, neighborhood, and sibling characteristics. Data comes from public schools in the state of California between the 2014-2015 and 2016-2018 school years.

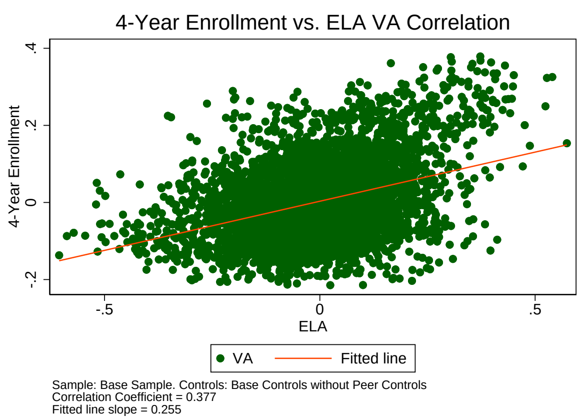
Figure 3: ELA Test Score Value-Added vs. Postsecondary Enrollment Value-Added



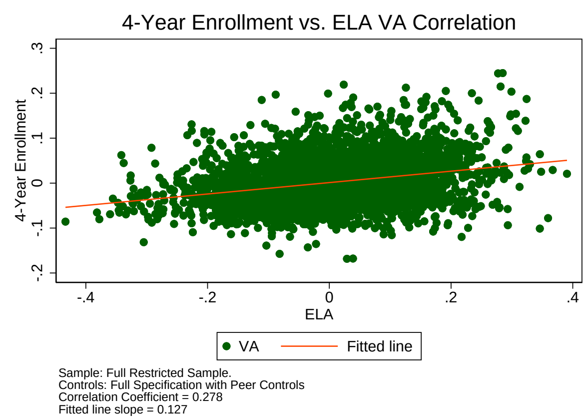
(a) Base Sample, Base Controls



(b) Restricted Sample, Full Controls



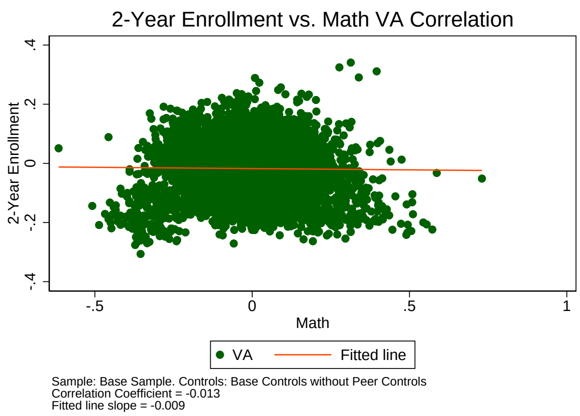
(c) Base Sample, Base Controls



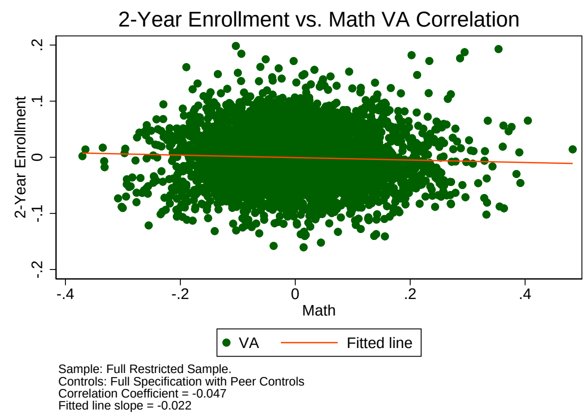
(d) Restricted Sample, Full Controls

Each data point represents a high school's value-added in a given year for two different outcomes. In Figures 3a and 3b, the horizontal axis gives ELA value-added and the vertical axis gives 2-year enrollment value-added. In Figures 3c and 3d, the horizontal axis gives ELA value-added and the vertical axis gives 4-year enrollment value-added. In Figures 3a and 3c, value-added is estimated on the base value-added sample using the base set of controls described in equation (1). In Figures 3b and 3d, value-added is estimated on the restricted value-added sample and additionally controls for peer, neighborhood, and sibling characteristics. Data comes from public schools in the state of California between the 2014-2015 and 2016-2018 school years.

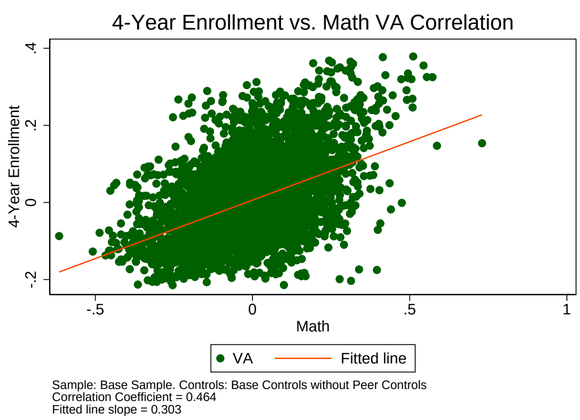
Figure 4: Math Test Score Value-Added vs. Postsecondary Enrollment Value-Added



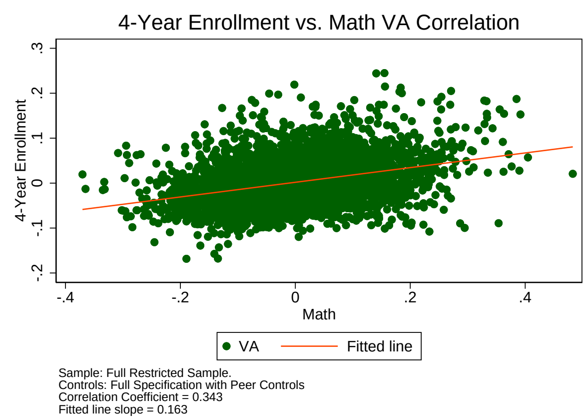
(a) Base Sample, Base Controls



(b) Restricted Sample, Full Controls



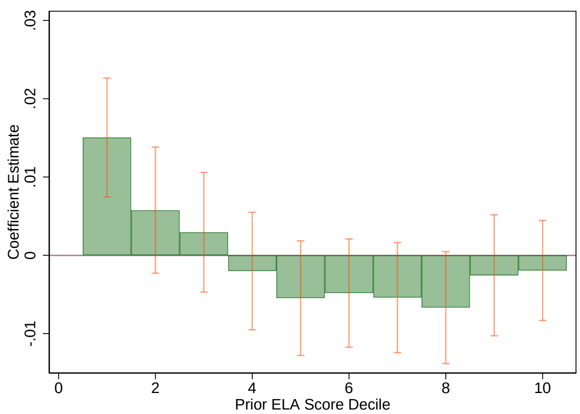
(c) Base Sample, Base Controls



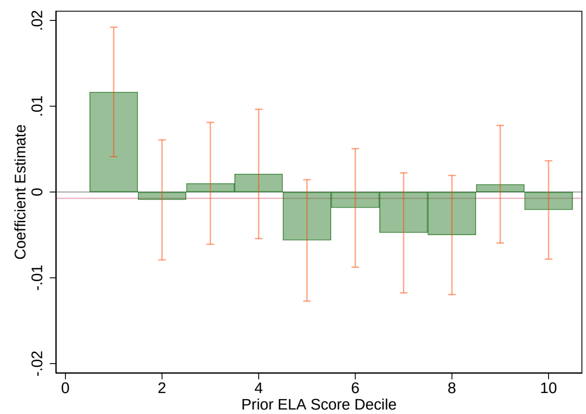
(d) Restricted Sample, Full Controls

Each data point represents a high school's value-added in a given year for two different outcomes. In Figures 4a and 4b, the horizontal axis gives math value-added and the vertical axis gives 2-year enrollment value-added. In Figures 4c and 4d, the horizontal axis gives math value-added and the vertical axis gives 4-year enrollment value-added. In Figures 4a and 4c, value-added is estimated on the base value-added sample using the base set of controls described in equation (1). In Figures 4b and 4d, value-added is estimated on the restricted value-added sample and additionally controls for peer, neighborhood, and sibling characteristics. Data comes from public schools in the state of California between the 2014-2015 and 2016-2018 school years.

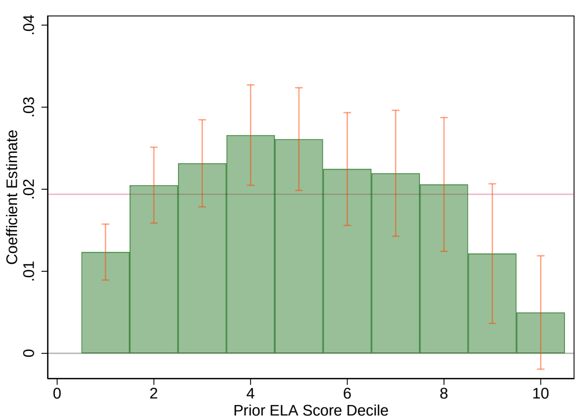
Figure 5: Heterogeneity of ELA Value-Added on Postsecondary Enrollment



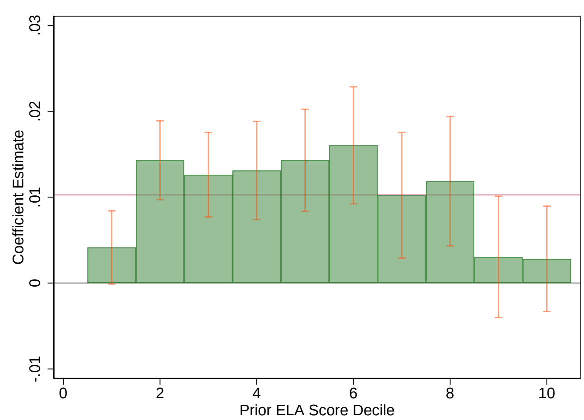
(a) 2-Year Enrollment: Base Sample, Base Controls



(b) 2-Year Enrollment: Restricted Sample, Full Controls



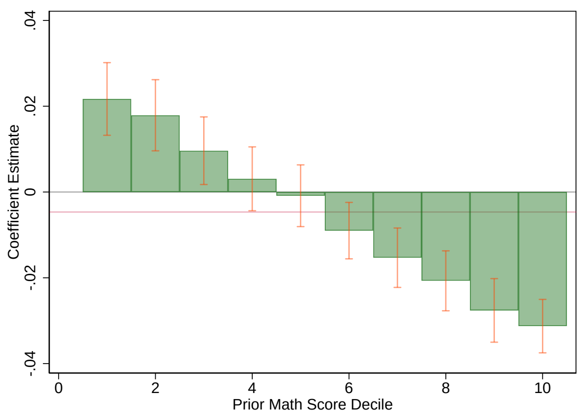
(c) 4-Year Enrollment: Base Sample, Base Controls



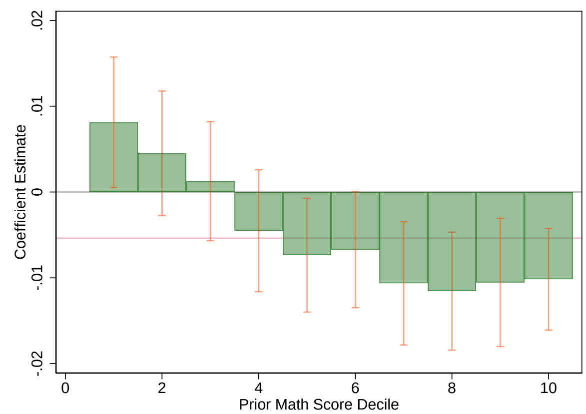
(d) 4-Year Enrollment: Restricted Sample, Full Controls

Each figure represents one regression. Each bar represents the coefficient estimate from a regression of college enrollment on ELA value-added interacted with a student's prior ELA score decile. Each regression also includes the controls included in the estimation of school value-added. The plunger associated with each bar gives the 95% confidence interval for the coefficient estimate. In Figures 5a and 5b, the dependent variable is 2-year college enrollment. In Figures 5c and 5d, the dependent variable is 4-year university enrollment. In Figures 5a and 5c, value-added is estimated on the base value-added sample using the base set of controls described in equation (1). In Figures 5b and 5d, value-added is estimated on the restricted value-added sample and additionally controls for peer, neighborhood, and sibling characteristics. Data comes from public schools in the state of California between the 2014-2015 and 2016-2018 school years.

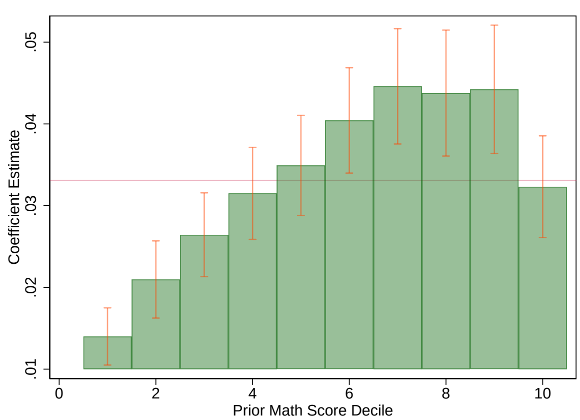
Figure 6: Heterogeneity of Math Value-Added on Postsecondary Enrollment



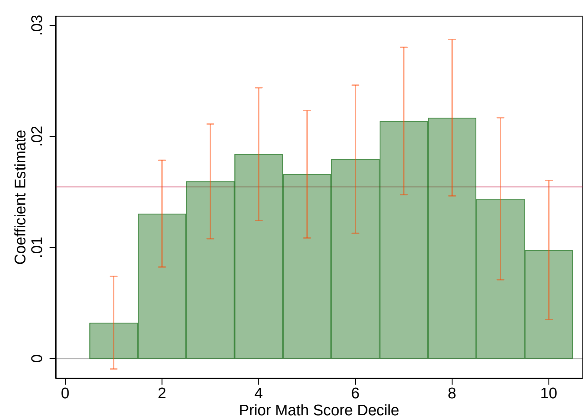
(a) 2-Year Enrollment: Base Sample, Base Controls



(b) 2-Year Enrollment: Restricted Sample, Full Controls



(c) 4-Year Enrollment: Base Sample, Base Controls



(d) 4-Year Enrollment: Restricted Sample, Full Controls

Each figure represents one regression. Each bar represents the coefficient estimate from a regression of college enrollment on math value-added interacted with a student's prior math score decile. Each regression also includes the controls included in the estimation of school value-added. The plunger associated with each bar gives the 95% confidence interval for the coefficient estimate. In Figures 6a and 6b, the dependent variable is 2-year college enrollment. In Figures 6c and 6d, the dependent variable is 4-year university enrollment. In Figures 6a and 6c, value-added is estimated on the base value-added sample using the base set of controls described in equation (1). In Figures 6b and 6d, value-added is estimated on the restricted value-added sample and additionally controls for peer, neighborhood, and sibling characteristics. Data comes from public schools in the state of California between the 2014-2015 and 2016-2018 school years.

Table 1: Summary Statistics

Panel A: 11th Grade Characteristics			
	Full Sample	Base Sample	Restricted Sample
11th Graders per School	440	475	498
Age in Years	16.8	16.7	16.7
Male	0.512	0.493	0.490
Hispanic or Latino	0.525	0.527	0.499
White	0.253	0.251	0.265
Asian	0.127	0.140	0.169
Black or African American	0.062	0.051	0.037
Other Race	0.042	0.031	0.030
Economic Disadvantage	0.561	0.545	0.507
Limited English Proficiency Status	0.101	0.053	0.040
Disabled	0.101	0.048	0.029
ELA Z-Score	0.000	0.175	0.291
Math Z-Score	-0.000	0.161	0.305
Prior ELA Z-Score	0.021	0.122	0.238
Prior Math Z-Score	0.022	0.127	0.259
Observations	1,913,221	1,232,262	410,947
Panel B: Postsecondary Outcomes			
Enrolled at a 2-Year College	0.347	0.378	0.363
Enrolled at a 4-Year University	0.278	0.332	0.374
Observations	1,822,742	1,208,514	409,286

Values are the mean of the variable listed on the left. Data comes from public schools in the state of California between the 2014-2015 and 2016-2018 school years. "Full Sample" refers to all 11th-grade students, "Base VA Sample" refers to the subset of students that meet the restrictions for our base value-added sample, and "Restricted VA Sample" refers to the subset of the base value-added sample that can be matched to peer, neighborhood, and sibling characteristics. Panel A contains the sample used to estimate ELA test score value-added (summary statistics for the sample used to estimate math value-added are nearly identical; the variable "math z-score" is the dependent variable in the math value-added models). With the exception of 11th graders per school, the variables in Panel A are the base controls used in the estimation of school value-added. Panel B contains the subset of panel A students who could be linked to the NSC data. The base ELA test score value-added sample contains 1,232,262 observations and 1,208,514 of these were linked to the NSC data. The restricted ELA test score value-added sample contains 410,947 observations and 409,286 of these were linked to the NSC data.

Table 2: ELA and Math Value-Added Distributions and Validity Tests

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: ELA						
Sample	Base	Restricted	Restricted	Restricted	Restricted	Restricted
Controls	Base	Base	Base	Base + ACS	Base + LO score + sib + ACS	Base + LO score + sib + ACS + Distance
Peer Controls	No	No	Yes	Yes	Yes	Yes
SD of VA	0.146	0.122	0.109	0.105	0.105	0.105
Spec Test	0.974	1.003	1.009	1.008	1.007	1.007
	(0.018)	(0.026)	(0.027)	(0.028)	(0.028)	(0.028)
FB: LO Score		–0.007*	–0.010**	–0.009*		
		(0.004)	(0.003)	(0.003)		
		{15000.0}	{15000.0}	{15000.0}		
FB: Sibling		0.001***	0.001***	0.001***		
		(0.000)	(0.000)	(0.000)		
		{65.7}	{67.5}	{61.4}		
FB: ACS		0.023***	0.005**			
		(0.004)	(0.002)			
		{87.3}	{37.2}			
FB: Distance	0.002	0.005	0.007	0.005	0.002	
	(0.001)	(0.003)	(0.004)	(0.003)	(0.003)	
	{59.6}	{53.9}	{75.2}	{48.2}	{28.8}	
Panel B: Math						
Sample	Base	Restricted	Restricted	Restricted	Restricted	Restricted
Controls	Base	Base	Base	Base + ACS	Base + LO score + sib + ACS	Base + LO score + sib + ACS + Distance
Peer Controls	No	No	Yes	Yes	Yes	Yes
SD of VA	0.151	0.127	0.109	0.101	0.101	0.101
Spec Test	0.946***	0.993	1.024	1.021	1.022	1.021
	(0.013)	(0.018)	(0.019)	(0.021)	(0.020)	(0.020)
FB: LO Score		0.001	–0.004**	–0.004**		
		(0.001)	(0.001)	(0.001)		
		{1888.2}	{1585.9}	{1512.3}		
FB: Sibling		0.002***	0.003***	0.002***		
		(0.000)	(0.000)	(0.000)		
		{205.2}	{216.6}	{194.8}		
FB: ACS		0.088***	0.026***			
		(0.012)	(0.005)			
		{521.6}	{151.5}			
FB: Distance	0.013***	0.017***	0.016***	0.006*	0.004	
	(0.003)	(0.004)	(0.004)	(0.003)	(0.002)	
	{581.1}	{238.3}	{175.3}	{69.0}	{46.9}	

Each column represents a separate set of value-added estimates. Panel A reports results for ELA value-added, and Panel B reports results for math value-added. The first row of each panel denotes the sample used to estimate value-added. "Base" refers to the base value-added sample. "Restricted" refers to the subset of students that can additionally be matched to peer, neighborhood, and sibling characteristics. The second row of each panel denotes the controls used in the estimation of value-added. "Base" includes the base set of controls described in equation (1). "ACS" refers to American Community Survey Census tract characteristics. "LO score" refers to one additional leave-out prior score. "Sib" refers to indicator variables for whether an older sibling attended a 2-year college or 4-year university. "Distance" refers to the distance from a student's high school to the nearest 2-year college and 4-year university. The third row of each panel indicates whether peer (jackknife) averages of the controls are included as additional controls. The fourth row of each panel contains the standard deviation of the school-year value-added estimates. The fifth row of each panel contains the coefficient for a bivariate regression of test score residuals u_{ist} on school value-added $\hat{\lambda}_{st}$. Statistical inference is conducted under the null hypothesis that the coefficient equals 1. The sixth-ninth rows of each panel contain the coefficient for a bivariate regression of test scores, as predicted by residualized excluded observables $\hat{u}_{ist}$, on school value-added $\hat{\lambda}_{st}$. Statistical inference is conducted under the null hypothesis that the coefficient equals 0. Standard errors clustered at the school level are presented in parentheses. The F-statistic for the excluded variables from the regression of test scores on excluded variables is presented in brackets. Data comes from public schools in the state of California between the 2014-2015 and 2016-2018 school years. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

Table 3: 2-Year and 4-Year Enrollment Value-Added Distributions and Validity Tests

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: 2 Year Enrollment						
Sample	Base	Restricted	Restricted	Restricted	Restricted	Restricted
Controls	Base	Base	Base	Base + ACS	Base + LO score + sib + ACS	Base + LO score + sib + ACS + Distance
Peer Controls	No	No	Yes	Yes	Yes	Yes
SD of VA	0.098	0.088	0.066	0.061	0.047	0.046
Spec Test	0.985**	0.992	0.996	1.005	1.027	1.029
	(0.006)	(0.008)	(0.012)	(0.013)	(0.019)	(0.019)
FB: LO Score		0.000	0.001	0.001		
		(0.000)	(0.000)	(0.000)		
		{118.2}	{153.4}	{133.5}		
FB: Sibling		0.020***	0.021***	0.020***		
		(0.001)	(0.001)	(0.001)		
		{994.6}	{922.3}	{874.0}		
FB: ACS		0.147***	0.050***			
		(0.009)	(0.007)			
		{484.0}	{172.1}			
FB: Distance	0.046***	0.031***	0.047***	0.045***	0.031***	
	(0.008)	(0.005)	(0.008)	(0.009)	(0.008)	
	{1023.1}	{249.6}	{285.0}	{232.5}	{134.6}	
Panel B: 4 Year Enrollment						
Sample	Base	Restricted	Restricted	Restricted	Restricted	Restricted
Controls	Base	Base	Base	Base + ACS	Base + LO score + sib + ACS	Base + LO score + sib + ACS + Distance
Peer Controls	No	No	Yes	Yes	Yes	Yes
SD of VA	0.099	0.091	0.075	0.063	0.048	0.048
Spec Test	1.004	1.025***	1.028**	1.025*	1.041*	1.042*
	(0.005)	(0.006)	(0.010)	(0.011)	(0.018)	(0.018)
FB: LO Score		0.004**	0.001	0.000		
		(0.001)	(0.001)	(0.001)		
		{1048.9}	{990.3}	{880.2}		
FB: Sibling		0.026***	0.026***	0.027***		
		(0.001)	(0.001)	(0.001)		
		{1835.2}	{1772.9}	{1670.3}		
FB: ACS		0.174***	0.082***			
		(0.014)	(0.010)			
		{686.0}	{429.8}			
FB: Distance	0.067***	0.043***	0.036***	0.027***	0.011*	
	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)	
	{1794.3}	{417.6}	{267.0}	{148.0}	{58.9}	

Each column represents a separate set of value-added estimates. Panel A reports results for 2-year enrollment value-added, and Panel B reports results for 4-year enrollment value-added. The first row of each panel denotes the sample used to estimate value-added. "Base" refers to the base value-added sample. "Restricted" refers to the subset of students that can additionally be matched to peer, neighborhood, and sibling characteristics. The second row of each panel denotes the controls used in the estimation of value-added. "Base" includes the base set of controls described in equation (1). "ACS" refers to American Community Survey Census tract characteristics. "LO score" refers to one additional leave-out prior score. "Sib" refers to indicator variables for whether an older sibling attended a 2-year college or 4-year university. "Distance" refers to the distance from a student's high school to the nearest 2-year college and 4-year university. The third row of each panel indicates whether peer (jackknife) averages of the controls are included as additional controls. The fourth row of each panel contains the standard deviation of the school-year value-added estimates. The fifth row of each panel contains the coefficient for a bivariate regression of test score residuals u_{ist} on school value-added $\hat{\lambda}_{st}$. Statistical inference is conducted under the null hypothesis that the coefficient equals 1. The sixth-ninth rows of each panel contain the coefficient for a bivariate regression of test scores, as predicted by residualized excluded observables $\hat{u}_{ist}$, on school value-added $\hat{\lambda}_{st}$. Statistical inference is conducted under the null hypothesis that the coefficient equals 0. Standard errors clustered at the school level are presented in parentheses. The F-statistic for the excluded variables from the regression of test scores on excluded variables is presented in brackets. Data comes from public schools in the state of California between the 2014-2015 and 2016-2018 school years. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

Table 4: Correlation between VA and school characteristics

	(1) ELA	(2) Math	(3) 2 Year Enr	(4) 4 Year Enr
Panel A: Restricted sample with base controls				
Percent Black or Hispanic	–0.000	–0.075**	–0.064**	0.020
	(0.012)	(0.013)	(0.011)	(0.010)
Percent Economically Disadvantaged	–0.020	–0.094**	–0.060**	–0.006
	(0.013)	(0.013)	(0.011)	(0.011)
Suburb	0.033**	0.021**	0.006	–0.002
	(0.007)	(0.008)	(0.006)	(0.005)
Town	0.033**	0.013	0.022*	–0.041**
	(0.011)	(0.012)	(0.010)	(0.007)
Rural	0.002	–0.016	0.002	–0.037**
	(0.014)	(0.015)	(0.010)	(0.009)
Log Enrollment	–0.014**	–0.005	0.024**	–0.014**
	(0.005)	(0.006)	(0.004)	(0.004)
Panel B: Restricted sample with full controls				
Percent Black or Hispanic	0.002	0.005	0.013*	–0.004
	(0.011)	(0.010)	(0.006)	(0.005)
Percent Economically Disadvantaged	–0.005	0.002	0.007	–0.005
	(0.011)	(0.010)	(0.006)	(0.005)
Suburb	0.024**	0.013*	–0.004	0.004
	(0.007)	(0.006)	(0.003)	(0.003)
Town	0.026*	0.021*	0.004	–0.004
	(0.010)	(0.010)	(0.006)	(0.004)
Rural	–0.002	–0.006	–0.010	–0.003
	(0.013)	(0.013)	(0.006)	(0.005)
Log Enrollment	–0.020**	–0.013**	0.003	–0.009**
	(0.004)	(0.004)	(0.002)	(0.002)

Notes: Table entries are coefficients from regressions of estimated value-added on the indicated school characteristic, weighted by school enrollment. Standard errors in parentheses. Each row represents results from a separate bivariate regression except for geographic location, which comes from a regression of value-added on indicators for the school's geographic location (urban is the reference group). Panel A uses value-added from models with the restricted sample and covariates used in column 2 in Tables 2 and 3. Panel B uses value-added from models with the restricted sample and controls used in the specification for column 5 in Tables 2 and 3. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

Table 5: Fraction of VA Variance Across District

	(1) ELA	(2) Math	(3) 2 Year Enr	(4) 4 Year Enr
Panel A: Restricted sample with base controls				
At least 2 schools in district	0.420	0.509	0.700	0.582
25 largest districts (7+ schools)	0.258	0.342	0.580	0.389
Panel B: Restricted sample with full controls				
At least 2 schools in district	0.404	0.401	0.602	0.410
25 largest districts (7+ schools)	0.248	0.264	0.362	0.230

Notes: Table entries are the R^2 from a regression of value-added on school district fixed-effects, weighted by school enrollment. Panel A uses value-added from models with the restricted sample and covariates used in column 2 in Tables 2 and 3. Panel B uses value-added from models with the restricted sample and controls used in the specification for column 5 in Tables 2 and 3. In each panel, the sample used in the first row includes districts with at least two high schools that have estimated value-added. The sample used in the second row consists of schools in the 25 largest school districts in California (which all have at least 7 schools with value-added estimates).

Table 6: Pass through of Test Score Value-Added to Postsecondary Enrollment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2-Year Enrollment				4-Year Enrollment			
ELA VA	0.001	–0.001	–0.008**	–0.001	0.022***	0.021***	0.019***	0.012***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)
N	1,273,999	427,979	427,979	427,793	1,273,999	427,979	427,979	427,793
Math VA	–0.003	–0.012**	– 0.017***	–0.006*	0.035***	0.038***	0.029***	0.017***
	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
N	1,273,929	427,963	427,963	427,781	1,273,929	427,963	427,963	427,781
Peer	N	N	Y	Y	N	N	Y	Y
Regression Controls	Match VA	Match VA	Match VA	Match VA	Match VA	Match VA	Match VA	Match VA
VA Sample	Base	Restricted	Restricted	Restricted	Base	Restricted	Restricted	Restricted
VA Controls	Base	Base	Base	Leave Out Score + ACS + Sibling	Base	Base	Base	Leave Out Score + ACS + Sibling

Each column represents a separate regression of college enrollment on test score value-added. ELA and math are estimated separately. Each estimate gives the coefficient estimate from a regression of college enrollment on test score value-added. Standard errors clustered at the school level are presented in parentheses. Each regression also includes all of the controls used in the estimation of value-added. The first row at the bottom indicates whether peer (jackknife) averages of the controls are included as additional controls. The third row at the bottom gives the value-added sample. "Base" refers to the base value-added sample. "Restricted" refers to the subset of students that can additionally be matched to peer, neighborhood, and sibling characteristics. The fourth row at the bottom gives the controls used in the estimation of value-added. "Base" includes the base set of controls described in equation (1). "ACS" refers to American Community Survey Census tract characteristics. "Leave Out Score" refers to one additional leave-out prior score. "Sibling" refers to indicator variables for whether an older sibling attended a 2-year college or 4-year university. "Distance" refers to the distance from a student's high school to the nearest 2-year college and 4-year university. Data comes from public schools in the state of California between the 2014-2015 and 2016-2018 school years. * $p < 0.05$ $**p < 0.01$ *** $p < 0.001$.

Table 7: Variance in Postsecondary Enrollment Value-Added Accounted for by Test Score Value-Added

	(1)	(2)	(3)	(4)	(5)
Panel A: 2 Year Enrollment					
Total Var	0.0095	0.0078	0.0044	0.0037	0.0022
Var Net of Test Score VA	0.0094	0.0074	0.0041	0.0036	0.0022
Net Var/Total Var	0.9915	0.9489	0.9236	0.9642	0.9813
1 - R^2	0.9916	0.9709	0.9449	0.9722	0.9872
Panel B: 4 Year Enrollment					
Total Var	0.0097	0.0082	0.0056	0.0039	0.0023
Var Net of Test Score VA	0.0072	0.0060	0.0042	0.0032	0.0019
Net Var/Total Var	0.7384	0.7253	0.7464	0.8150	0.8259
1 - R^2	0.8446	0.8373	0.8549	0.9037	0.9225
Sample	Base	Restricted	Restricted	Restricted	Restricted
Controls	Base	Base	Base	Base + ACS	Base + LO score + sib + ACS
Peer Controls	No	No	Yes	Yes	Yes

Each column represents a separate set of value-added estimates. Panel A reports results for 2-year enrollment value-added, and Panel B reports results for 4-year enrollment value-added. The first row of each panel gives the variance of school value-added on college enrollment. The second row of each panel gives the variance of school value-added on college enrollment obtained from a model that additionally controls for ELA and math test score value-added. The third row of each panel gives the ratio of the second row to the first row. The fourth row of each panel gives the value of $1 - R^2$ from a regression of enrollment value-added on ELA and math test score value-added, weighted by the number of students that contributed to the value-added estimates. The first row at the bottom denotes the sample used to estimate value-added. "Base" refers to the base value-added sample. "Restricted" refers to the subset of students that can additionally be matched to peer, neighborhood, and sibling characteristics. The second row at the bottom denotes the controls used in the estimation of value-added. "Base" includes the base set of controls described in equation (1). "ACS" refers to American Community Survey Census tract characteristics. "LO score" refers to one additional leave-out prior score. "Sib" refers to indicator variables for whether an older sibling attended a 2-year college or 4-year university. The third row at the bottom indicates whether peer (jackknife) averages of the controls are included as additional controls. Data comes from public schools in the state of California between the 2014-2015 and 2016-2018 school years.

Table 8: Relationship Between Value-Added and School Climate Survey Indices

	(1)	(2)	(3)	(4)
Panel A: Separate Regressions for Each Index				
School Climate	0.20***	0.19***	–0.06	0.18***
	(0.04)	(0.04)	(0.04)	(0.04)
Teacher and Staff Quality	0.15***	0.17***	–0.05	0.13***
	(0.04)	(0.04)	(0.04)	(0.04)
Counseling Support	0.18***	0.21***	0.02	0.07
	(0.05)	(0.05)	(0.05)	(0.05)
Panel B: Regressions Including All Indices				
School Climate	0.21***	0.14**	–0.07	0.22***
	(0.07)	(0.07)	(0.07)	(0.07)
Teacher and Staff Quality	–0.06	0.01	–0.03	–0.03
	(0.07)	(0.07)	(0.07)	(0.07)
Counseling Support	0.09	0.12**	0.08	–0.04
	(0.06)	(0.06)	(0.06)	(0.06)
VA	ELA	Math	2 Year Enrollment	4 Year Enrollment
Controls	Base + LOS + Sibling + ACS	Base + LOS + Sibling + ACS	Base + LOS + Sibling + ACS	Base + LOS + Sibling + ACS
Peer	Yes	Yes	Yes	Yes

In Panel A, each cell represents a separate regression of school value-added on a single school survey index. In Panel B, each column represents a separate regression of school value-added on all three school survey indices. School value-added estimates are averaged across the years 2015-2018 and school survey indices are averaged across the years 2017-2019. Each regression contains school-level averages of the base value-added controls. The first row at the bottom gives the type of school value-added. The second row at the bottom gives the controls used in the estimation of value-added. “Base” includes the base set of controls described in equation (1). “ACS” refers to American Community Survey Census tract characteristics. “LOS” refers to one additional leave-out prior score. “Sibling” refers to indicator variables for whether an older sibling attended a 2-year college or 4-year university. The third row at the bottom indicates whether peer (jackknife) averages of the controls are included as additional controls. Data comes from public schools in the state of California between the 2014-2015 and 2016-2018 school years. * $p < 0.05$) ** $p < 0.01$ *** $p < 0.001$.

Online Appendix A Sample Creation

Columns one and three of Table A.1 provide the number of observations in the data conditional on a set of restrictions implemented in order to form the value added sample. Column one provides results for the English language arts (ELA) sample and column three provides results for the math sample. The rows are additive, such that the first row contains all observations, the second row imposes one restriction, the third row imposes two restrictions, and so forth. The first row denotes the total number of 11th-grade students in the California Assessment of Student Performance and Progress (CAASPP) dataset. This constitutes the "full" sample in Table 1.

The second row keeps students who attend high schools that serve grades 9–12. The third row keeps only the first time that a student attempted a grade, and thus drops observations in which a student is repeating a grade. The fourth row keeps only students at "conventional" schools. This includes schools in the following categories defined by the California Department of Education (CDE): Preschool, Elementary School (Public), Elementary School in 1 School District (Public), Intermediate/Middle Schools (Public), Junior High Schools (Public), K–12 Schools (Public), High Schools (Public), and High Schools in 1 School District (Public).[31] The fifth row drops any schools that enroll 10 11th-grade students or fewer in a given year. The sixth row drops students who are missing a test score in the specific subject for which value added is calculated. The seventh row drops students who are missing any of the demographic controls. The eighth row drops students who are missing the prior test scores used as control variables. The ninth row drops students if fewer than seven observations can be used to estimate value added for their school by year cell, which insures that all value added estimates are based on at least seven observations. The sample in row nine constitutes our "base" sample in Table 1.

The tenth row drops students who are missing 7th grade test scores, which are used as a leave-out variable to estimate forecast bias. The 11th row drops students who could not be matched to older siblings. The 12th row drops students who could not be matched to Census-block data from the American Community Survey (ACS) based on their home address. The final row constitutes our "restricted" sample in Table 1.

Columns two and four of Table A.1 provide the average ELA and math test score, respectively, of the sample. Test scores are standardized to have mean zero and standard deviation one on the population of 11th grade students, so the average test score in the first row is effectively zero. Our base and restricted samples are positively selected, with an average test score about 0.17 (base) and 0.3 (restricted) standard deviations above average. The restrictions that most impact the composition of our sample are dropping

[31] This drops students in the following categories: Special Education Schools (Public), County Community, Youth Authority Facilities (CEA), Opportunity Schools, Juvenile Court Schools, Other County or District Programs, State Special Schools, Alternative Schools of Choice, Continuation High Schools, District Community Day Schools, Adult Education Centers, and Regional Occupational Center/Program (ROC/P).

students who attend "non-conventional" schools, dropping students who lack prior test scores, dropping students who could not be matched to an older sibling, and dropping students who could not be matched to ACS data.

Table A.1: Sample Counts

	ELA		Math	
	# of Students	Z-Score Mean	# of Students	Z-Score Mean
All Students	1,913,221	6.70e-07	1,913,221	-1.05e-07
+ 9-12 School	1,790,429	.00305	1,790,429	.00863
+ First Test Score for Grade	1,752,542	.00678	1,752,542	.0122
+ Conventional School	1,586,766	.0812	1,586,766	.086
+ 11th Graders per School > 10	1,585,957	.0813	1,585,957	.0862
+ Nonmissing Subject Test Score	1,494,208	.0813	1,490,647	.0862
+ Nonmissing Demographic Controls	1,486,580	.0817	1,482,922	.0866
+ Nonmissing Prior Test Scores	1,232,514	.174	1,227,174	.161
+ School VA Sample Size ≥ 7	1,232,262	.175	1,226,915	.161
+ Leave Out Scores	1,190,916	.201	1,185,842	.185
+ Leave Out Scores and Sibling	662,094	.259	659,398	.266
+ Leave Out Scores, Sibling, and ACS	410,947	.291	409,332	.305

Values are counts of the number of observations in each sample along with the average test score for the sample. Each row is additive, so the restrictions from all prior rows are also present in the current row. Data comes from public schools in the state of California between the 2014-2015 and 2016-2018 school years.

Online Appendix B: Survey Questions

The survey questions for our school climate, teacher and staff quality, and counseling support indices in equation (4) are the following:

B.1 School Climate Index

Questions for Parents

Please indicate how much you agree or disagree with the following statements about this school. (Strongly agree, agree, disagree, strongly disagree, don't know/NA)

1. This school promotes academic success for **all** students
2. This school is a supportive and inviting place for students to learn
3. This school allows input and welcomes parents' contributions
4. This school encourages me to be an active partner with the school in educating my child

Questions for Students

How strongly do you agree or disagree with the following statements? (strongly disagree, disagree, neither disagree nor agree, agree, strongly agree)

1. I feel close to people in this school
2. I am happy to be at this school
3. I feel like I am part of this school
4. The teachers at this school treat students fairly
5. I feel safe in my school
6. My school is usually clean and tidy
7. Teachers at this school communicate with parents about what students are expected to learn in class
8. Parents feel welcome to participate at this school
9. School staff take parent concerns seriously

Questions for Staff

Please indicate how much you agree or disagree with the following statements about your school. (Strongly agree, agree, disagree, strongly disagree)

1. This school encourages students to enroll in rigorous courses (such as honors and AP), regardless of their race, ethnicity, or nationality
2. This school has high expectations for all students regardless of their race, ethnicity, or nationality
3. In this school, adults feel a responsibility to improve this school
4. This school motivates students to learn

Please indicate how much you agree or disagree with the following statements about your school. (Strongly agree, agree, disagree, strongly disagree)

1. Students are motivated to learn
2. Teachers at this school communicate with parents about what their children are expected to learn in class

How much of a problem AT THIS SCHOOL is... (insignificant problem, mild problem, moderate problem, severe problem)

1. Cutting classes or being truant?

B.2 Teacher and Staff Quality Index

Questions for Parents

Please indicate how much you agree or disagree with the following statements about this school. (Strongly agree, agree, disagree, strongly disagree, don't know/NA)

1. This school provides high quality instruction to my child
2. This school motivates students to learn
3. This school has teachers who go out of their way to help students
4. This school has adults who really care about students
5. This school has high expectations for all students

Questions for Students

At my school, there is a teacher or some other adult... (Not at all true, a little true, pretty much true, very much true)

1. Who really cares about me
2. Who tells me when I do a good job
3. Who notices when I am not there
4. Who always wants me to do my best
5. Who listens to me when I have something to say
6. Who believes that I will be a success

Questions for Staff

Do you feel you need more professional development, training, mentorship, or other support to do your job in any of the following areas?

1. Meeting academic standards
2. Evidence-based methods of instruction
3. Positive behavioral support and classroom management
4. Closing the achievement gap
5. Meeting social, emotional, and developmental needs of youth (e.g. resilience promotion)
6. Creating a positive school climate

B.3 Counseling Support Index

Questions for Parents

Please indicate how much you agree or disagree with the following statements about this school. (Strongly agree, agree, disagree, strongly disagree, don't know/NA)

1. This school provides quality counseling or other ways to help students with social or emotional needs

How well has this child's school been doing the following things during the school year? (very well, just okay, not very well, does not do it at all, don't know/NA)

1. Providing information on how to help your child plan for college or vocational school

Questions for Staff

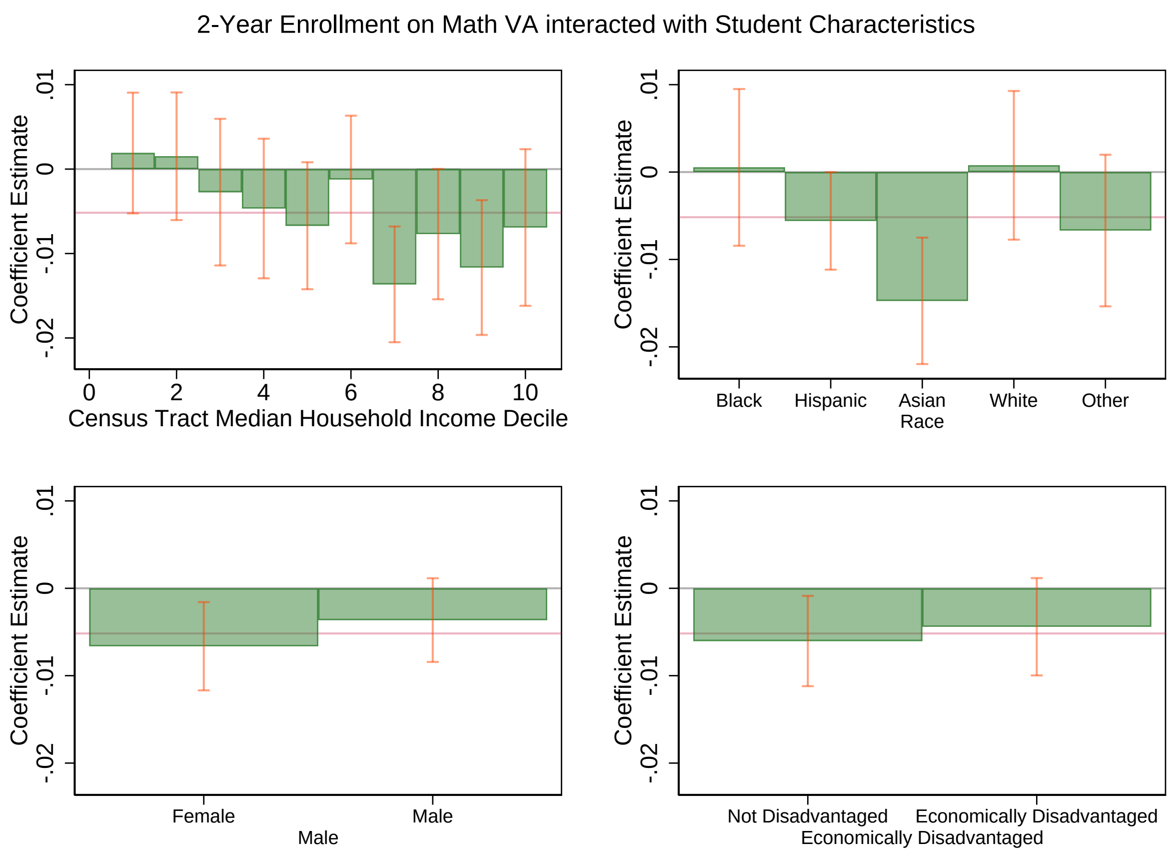
Please indicate how much you agree or disagree with the following statements about your school. (Strongly agree, agree, disagree, strongly disagree)

1. This school provides adequate counseling and support services for students

Please indicate how much you agree or disagree with the following statements about your school. (Strongly agree, agree, disagree, strongly disagree)

1. This school provides counseling or other ways to help students with their social-emotional needs

Online Appendix C: Pass Through Heterogeneity by Student Characteristics



4-Year Enrollment on Math VA interacted with Student Characteristics