Within-School Heterogeneity in Quality: Do Schools Provide Equal Value Added to All Students?

Matthew Naven*

Washington and Lee University

Draft Date: May 22, 2023

Download the most recent version **here**

Abstract

Low-socioeconomic status (SES), minority, and male students perform worse than their high-SES, non-minority, and female peers on standardized tests. This paper investigates how within-school differences in school quality contribute to these educational achievement gaps. Using individual-level data on the universe of public-school students in California, I estimate school quality using a value added methodology that accounts for the fact that students sort to schools on observable characteristics. I allow for within-school heterogeneity by estimating a distinct value added for each school's low-/high-SES, minority/non-minority, and male/female students. Standard value added models suggest that on average schools provide less value added to their low-SES, minority, and male students, particularly on postsecondary enrollment. However, value added models that control for neighborhood, older-sibling, and peer characteristics suggest that schools provide similar value added to low-/high-SES students and minority/non-minority students but more value added to female students. Within-school heterogeneity accounts for 6% of the test-score achievement gap and 22% of the difference in postsecondary enrollment between men and women. [JEL Codes: **I24, H75, I21, I23, J24**]

1 Introduction

Despite efforts to close them in recent decades, achievement gaps in education by socioeconomic status (SES), ethnicity, and sex have persisted. The test score gap between low- and high-income students has increased

*Contact the author at mnaven@wlu.edu or visit www.matthewnaven.com for more information. I am grateful to Scott Carrell, Marianne Bitler, Shu Shen, Michal Kurlaender, Paco Martorell, Sherrie Reed, Katie Shester, and members of the California Education Lab for helpful comments and suggestions. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305E150006 to the Regents of the University of California. The opinions expressed are those of the author and do not necessarily represent views of the Institute, the U.S. Department of Education, or of the agencies providing data. I gratefully acknowledge support from the Russell J. and Dorothy S. Bilinski Educational Foundation Dissertation Writing Fellowship. I would like to thank the participants at the University of California, Davis, the University of Nevada, Reno, the All-California Labor Economics Conference, Sonoma State University, APPAM, Analysis Group, Cornerstone Research, Kennesaw State University, IMPAQ International, the Federal Communications Commission, Washington and Lee University, AEFP, and the APPAM California Regional Student Conference seminars for their insights.

during the past 50 years (Reardon and Robinson, 2008; Reardon, 2011), as has the college-going gap (Bailey and Dynarski, 2011) and the elite college-going gap (Reardon, Baker and Klasik, 2012), even after accounting for academic ability (Belley and Lochner, 2007; Karen, 2002). In addition, the low-/high-income test score gap only decreases slightly as students progress through school (Reardon, 2011).

The white-black and white-Hispanic test score gap is already 0.5 to 0.8 standard deviations by the time students enroll in kindergarten (Reardon and Portilla, 2016), and on average white students score one and a half or more grade levels higher than black and Hispanic students enrolled in socioeconomically-similar school districts (Reardon, 2016). While the black-white gap has shrunk over time, it appears to widen during the school years (Reardon and Robinson, 2008). On the other hand, the Hispanic-white gap varies considerably by origin, but narrows for all subgroups in the earliest grades (Reardon and Galindo, 2009).

Gender differences in education outcomes are less straightforward. Women perform better on average than males on ELA exams (Chatterji, 2006; Cimpian et al., 2016; Fryer Jr and Levitt, 2010; Husain and Millimet, 2009; Lee, Moon and Hegar, 2011; Penner and Paret, 2008; Robinson and Lubienski, 2011; Sohn, 2012), and in the 2018-2019 academic year women accounted for 57% of both postsecondary enrollment (National Center for Education Statistics, 2022*b*) and bachelor's degree receipt (National Center for Education Statistics, 2022*a*). However, men are more likely to study lucrative STEM fields upon enrolling in college (Kahn and Ginther, 2017) and have higher median earnings (American Community Survey, 2019).

While many factors likely play a role in these achievement gaps, one factor that remains unstudied is within-school heterogeneity in the quality that schools provide to their students. Schools are a large component of early human capital formation, and the previous literature on school and teacher quality largely assumes that good schools and teachers are beneficial for *all* students attending those schools. By reporting only a single measure for each school or teacher, these studies calculate the average effect of a school or teacher on their students. Although a measure of overall school quality is undoubtedly important, it is not necessarily informative about whether a school will be effective for a *specific type* of student. Average measures ignore the possibility that schools provide a higher-quality education to certain subgroups of students which may in turn accentuate or attenuate existing achievement gaps.

For example, students have been shown to perform better (Dee, 2004, 2007; Egalite, Kisida and Winters, 2015; Gershenson et al., 2022) and are less likely to be disciplined (Dee, 2005; Lindsay and Hart, 2017; Holt and Gershenson, 2019) when assigned to a same-race or same-sex teacher. Likewise, teachers report higher evaluations of students who share the same race or sex (Dee, 2005). Thus students may perform better when they attend a school with faculty that look like them. Low-income students may benefit from attending a school that is familiar with the domestic issues that this population faces, or they may benefit from being surrounded with high-income peers that are more likely to have institutional knowledge on academic resources. Yet there is a dearth of research on the student-school match[1], despite the fact that there is substantial heterogeneity in teacher characteristics and other inputs at the school level.

This paper explores whether school quality differs by SES, ethnicity, and sex within schools. Using individual-level data on the universe of California public-school students linked to postsecondary records, I estimate school by subgroup quality by applying the value added with drift methodology, as in Chetty, Friedman and Rockoff (2014), to schools. The value added methodology accounts for the fact that students do not randomly sort to schools and that subgroups of students have varying levels of average academic performance. The drift methodology, which allows value added to change from year to year, is particularly suited to the school quality setting, as schools experience faculty and staff turnover that could lead to changes in quality from year to year.

I estimate a school's value added on both standardized test scores and postsecondary enrollment. I perform three separate analyses, one each by SES, ethnicity, and sex, and I allow each school to have two value added estimates: one for each dichotomous group (low-/high-SES, minority/non-minority, and male/female). For example, School A will have one estimate of the value added it provides to low-SES students and one estimate of the value added it provides to high-SES students. Simulations suggest that assuming homogeneity in school value added can incorrectly rank schools with respect to value added in the presence of within-school heterogeneity in value added. Allowing for this within-school heterogeneity when estimating school value added restores the correct ranking of schools.

Results using standard value added models that control for prior test scores and student demographic characteristics suggest that there is indeed within-school heterogeneity in the value added that schools provide to students. These models suggest that within-school heterogeneity in value added on test scores accounts for 1.5% of the ethnicity achievement gap, and 7% of the sex achievement gap. For postsecondary enrollment, within-school heterogeneity in value added explains 14% of the SES gap, 12% of the ethnicity gap, and 32% of the sex gap. However, value added models that include additional controls for the characteristics of the neighborhood in which a student lives, older-sibling characteristics, and peer characteristics suggest that some of the estimated within-school heterogeneity in value added is driven by omitted variable bias. After controlling for these additional variables, within-school heterogeneity in value added on test scores shrinks to 1% of the ethnicity gap and 6% of the sex gap. Within-school heterogeneity in value added on postsecondary enrollment shrinks to 0.5% of the SES gap, 2% of the ethnicity gap, and 22% of the sex gap.

Thus schools appear to largely provide similar value added to students of all socioeconomic backgrounds

[1] To my knowledge, only two economics papers explore how value added may differ by subgroup. Carrell, Page and West (2010) find that female students assigned to female professors in STEM classes perform better and are more likely to obtain a STEM degree than their fellow female students who are assigned a male professor. Bau (2022) explores how private schools in Pakistan compete in horizontal quality when students respond differentially to match quality by SES.

but slightly more value added to non-minority students and much more value added to female students, particularly with regards to postsecondary enrollment. For male students, the average difference in value added on test scores is equivalent to attending a school 0.09 standard deviations below average and the average difference in value added on postsecondary enrollment is equivalent to attending a school 0.4 standard deviations below average. For minority students, the average differences are equivalent to attending a school 0.08 and 0.1 standard deviations below average, respectively.

I also find that certain school characteristics are correlated with the value added for various subgroups of students in different ways. Schools with more pupil services per student are associated with increases in the likelihood that low-SES and minority students eventually enroll in college. Schools with more English Learner staff per student are associated with relatively higher value added on test scores for low-SES and minority students. Schools with more male teachers are associated with lower value added on test scores for all students but higher value added on postsecondary enrollment for all students except, interestingly, male students. Schools with more minority faculty are associated with lower value added on test scores for high-SES, non-minority, male, and female students, but these negative correlations are completely mitigated for low-SES and minority students.

This paper adds numerous important contributions to the literature on education quality and achievement gaps. First, this paper provides some of the first evidence on whether schools differentially affect test scores and postsecondary enrollment for low-SES, minority, and male students and is the first to extend the value added with drift methodology (Chetty, Friedman and Rockoff, 2014) to schools allowing for within-school heterogeneity. This is particularly relevant given the large postsecondary enrollment gaps for these groups. Second, this paper proves that a ranking system based on a single value added estimate can incorrectly rank the quality of schools if the assumption that schools provide homogenous school value added to their students is violated. This information may be valuable when devising school accountability systems based on value added estimates. Finally, as with Naven (2022) and Carrell et al. (2023), this paper is unique in that it links the universe of public school students in California, which has the largest public school population in the United States, to their postsecondary outcomes. California is a particularly relevant state in which to study this topic because the student population is majority minority and low-SES, and California has a robust postsecondary infrastructure that includes two-year community colleges, teaching universities, and globally-ranked research universities.

2 Simulation

A significant advantage of allowing a school's value added to vary by subgroup is that it corrects for the incorrect ranking of schools if the assumption of a homogeneous school effect is violated. In order to demonstrate this, I simulate a dataset of student observations. Let true value added for school s in year t be μ_{sLt} for low-type students and μ_{sHt} for high-type students. Let student test scores be generated according to the true model in equation (1), with school by subgroup by year value added μ_{sdt}, school by year common shocks θ_{st}, and a noise term ε_{isdt} all contributing.

$$z_{isdt} = \mu_{sdt} + \theta_{st} + \varepsilon_{isdt} \tag{1}$$

For simplicity, let $\mu_{sdt} = \mu_{sd}\ \forall t$. Let there be four types of schools, A, B, C, and D, with equal probability, with true school value added for low- and high-type students distributed as μ_{sLt} and μ_{sHt} in columns 2 and 3 of Table 1. School A and school D are unambiguously the best and worst schools, respectively. School B is particularly effective with low-type students while school C is effective with high-type students. Notice that on average schools provide more value added to high-type students.

Let N_{sL} and N_{sH} be the average number of low- and high-type students in a school, respectively. I assign each school a baseline value of N_{sL} and N_{sH} and allow the yearly population of low- and high-type students to fluctuate within a window of this baseline by multiplying N_{sL} and N_{sH} separately by inflation terms j_{sLt} and j_{sHt}. This allows for yearly variation in the proportion of low-type students within a school. The distributions for θ_{st}, ε_{isdt}, N_{sLt}, N_{sHt}, j_{sLt}, and j_{sHt}, as well as the number of schools and years, are given in Table 2.

Figure 1 plots the value added estimates obtained using the drift methodology in Chetty, Friedman and Rockoff (2014) on the simulated data. Figures 1a, 1b, 1c, and 1d contain the value added estimates under both the assumption of a homogenous effect for all students and a heterogeneous effect for low- and high-type students for schools of type A, B, C, and D, respectively. The horizontal axis plots the proportion of low-type students within the school. The horizontal red lines mark the true value added for each school. For each school type, allowing for a heterogeneous value added effect produces estimates clustered around the true value regardless of what proportion of students are low-type. Imposing the assumption of a homogeneous value added effect, however, causes the value added estimate to fluctuate according the proportion of low-type students. While this issue is less pronounced when schools have similar value added for both low- and high-type students, such as schools A and D, the value added estimate can vary dramatically according to the proportion of low-type students when schools are more effective with a particular type of student, as with schools B and C.

The assumption of homogeneous school value added can also incorrectly rank schools when they are at the extreme end of the distribution of proportion low-type students. Figure 1e plots the homogeneous estimate for schools A and C. Despite the fact that school A is better than school C, since it is equally effective with high-type students but better with low-type students, its estimated effect is worse than school C if A enrolls a large proportion of low-type students and C enrolls a large proportion of high-type students. Allowing the schools to have heterogeneous effects, as in figure 1f, preserves the correct ranking of schools. This ranking error illustrates the importance of studying within-school heterogeneity in value added.

3 Data

My study uses individual-level data on the universe of public school-students in the state of California. Standardized test score information comes from the California Standards Test (CST). Data from the CST spans the 2002-2003 to 2012-2013 school years[2] and tests students in English language arts (ELA) and math during grades 2-11. This paper focuses on the ELA exam because students are tracked to different, non-comparable, math tests starting in the seventh grade. The data also include demographic information on each student, such as sex, race, economic disadvantage status, limited English proficiency status, and whether or not the student has a disability. State student IDs can be used to link students to prior test scores across time. Each cohort consists of about 475,000 students.

I assign students to a binary low- or high-SES status based on whether they are defined as socioeconomically disadvantaged[3] by the California Department of Education (CDE). In order to get an idea of the income level of these students, Figure 2 plots the distribution of total household income in 2017 dollars by socioeconomic disadvantage status from the Survey of Income and Program Participation (SIPP).[4] Economically disadvantaged students live in households with a median income of about $60,000, which is about $36,000 lower that the median income of students in households that are not economically disadvantaged. However, the peak of the income distribution for economically disadvantaged students occurs much lower at about $38,000. I define minority students as Hispanic, black, Native American and two or more races and non-minority students as white and Asian.

Table 3 gives summary statistics for the CST data by subgroup for the base test score value added samples and includes all the dependent and independent variables used in the value added analyses. Appendix Table A.1 shows the limitations that are imposed in order to form the base value added sample, which are similar

[2] Due to the fact that I use test scores from two grades prior as a control variable, I only calculate value added estimates for the years 2004-2005 to 2012-2013.

[3] Defined by the CDE as "a student neither of whose parents have received a high school diploma or a student who is eligible for the free or reduced-price lunch program, also known as the National School Lunch Program (NSLP)."

[4] I exclude observations with a total household income of greater than $250,000 from the figure but not the calculation of the median income, as there is a long low-density right tail in each distribution.

to those made in the teacher value added literature. The vast majority of students in the CST data that cannot be included in the value added estimation are excluded because they lack prior test scores. I exclude grades 2-3 because they lack sufficient prior test scores in order to estimate value added.[5]

Unsurprisingly, SES and minority status are highly correlated. Low-SES students are a little less than 50 percentage points less likely to be white or Asian. Minorities are 50 percentage points more likely to be socioeconomically disadvantaged than their non-minority peers. Low-SES and minority students are also much more likely to be limited English proficient, which is likely due to Spanish-speaking Hispanic students. Low-SES and minority students perform much worse on standardized tests, as their average ELA test scores are about 0.75 standard deviations worse than their peers. Males and females are fairly similar demographically with two exceptions: males are slightly more likely to be diagnosed with a disability than females, and females perform about 0.18 standard deviations better on ELA exams than males.[6] As is the case in other value added studies, the value added sample is positively selected on prior test scores, as students in the value added sample score about 0.09 standard deviations above average on their current test scores.[7] Appendix section A gives more information on the data.

In addition to standard value added models estimated on my base value added samples, I include additional analyses that control for the outcomes of older siblings, American Community Survey (ACS) Census tract characteristics, and peer effects. I match students to Census tracts by geocoding each student's home address in a given year, and I assign students to siblings by linking each student to any other student that shared the same address in the same year and then linking all students that share a mutual sibling. About 13% of students from the test score base sample and 18% of students from the postsecondary enrollment base sample have at least one older sibling and can be matched to the American Community Survey. The ACS Census tract characteristics and older-sibling outcomes for this subsample are presented in appendix Table A.2 and peer summary statistics for this subsample are presented in appendix Table A.3.

Postsecondary data comes from the National Student Clearinghouse (NSC). The NSC data includes enrollment data for the cohorts of students that graduated high school between the spring of 2010 and 2020, inclusive.[8] The NSC provides national enrollment coverage and therefore accounts for both California colleges and universities as well as out-of-state enrollment.

[5] Prior test scores are necessary in order to obtain unbiased estimates when using value added methodologies (Kane and Staiger, 2008; Deming, 2014).

[6] Males perform about 0.02 standard deviations better on math exams in the grades in which students take a common test.

[7] Test scores are standardized to have mean zero and standard deviation one at the grade by year level on the entire population of students taking the CST.

[8] The cohorts matched were actually spring 2009 to spring 2019 11th grade students, because I do not observe high school graduation data nor the students in 12th grade.

4 School Value Added

4.1 Model

In this section I describe a model of student learning in order to better describe which factors contribute to a school's value added measure. The model follows Naven (2022) with the exception that I allow for within-school heterogeneity in value added by subgroup. Suppose that the outcome of a student i of subgroup d in grade g of school s in year t is determined according to equation (2), such that a student's endowment ι_i, contemporaneous learning ℓ_{igdt}, prior learning ℓ_{ikd} depreciated by a factor δ_{kd}, and idiosyncratic school-level shocks θ_{st} all contribute. Assume that students take each grade only once, so that g and t are interchangeable within student (and consequently dropping t subscripts for prior learning).

$$z_{isgdt} = \underbrace{\iota_i}_{\text{Endowment}} + \sum_{k=0}^{g-1} \underbrace{\delta_{kd} \cdot \ell_{ikd}}_{\text{Prior Learning}} + \underbrace{\underbrace{\ell_{igdt}}_{\text{Learning}} + \underbrace{\theta_{st}}_{\text{Shocks}} + \underbrace{\varepsilon_{isgdt}}_{\text{Noise}}}_{r_{isgdt}} \tag{2}$$

Assume that the portion of outcome z_{isgdt} that is due to learning is modeled by equation (3) such that teachers τ_{sgdt}, other school factors ψ_{sdt} (such as principals, counselors, curricula, extracurricular activities, and peers), and non-school factors υ_{igdt} (such as parents, siblings, or neighborhoods) contribute to student learning.

$$\ell_{igdt} = \underbrace{\underbrace{\tau_{sgdt}}_{\text{Teachers}} + \underbrace{\psi_{sdt}}_{\text{School Factors}}}_{\mu_{sgdt}} + \underbrace{\upsilon_{igdt}}_{\text{Outside Factors}} \tag{3}$$

The combined effect of teachers and other school factors on a particular subgroup is a school's value added for that subgroup in a given grade and year, μ_{sgdt}. Aggregating μ_{sgdt} across grades produces a school's total value added for a particular subgroup in a given year, μ_{sdt}, which is the variable of interest in this study. While other studies have investigated the impact of τ_{sgdt} on long-run outcomes, studying school quality allows ψ_{sdt} to also have an impact. This may be particularly important when studying the effects of education on postsecondary enrollment, as high schools are much more likely to have counselors dedicated to the postsecondary decision process and some schools may have better resources on the application process, such as college fairs or mandatory SAT/ACT testing, than others.

Note that by regressing a student's test score in grade g on their test score in grade $g-1$ it is possible to control for ι_i and $\sum_{k=0}^{g-1} \delta_{kd} \cdot \ell_{ikd}$, the performance a student would achieve even in the absence of school input. This leaves us with the residual term r_{isgdt}, which captures the portion of student performance that is not related to a student's prior achievement.

4.2 Methodology

To estimate μ_{sdt}, I extend the value added methodology that allows for drift over time described in Chetty, Friedman and Rockoff (2014) to the school level. The value added methodology accounts for the fact that schools receive students of varying backgrounds.[9] Hence, schools that receive only the lowest performing students should not be penalized for the fact that the students they receive will likely have lower outcomes on average. Instead, they should be evaluated on how much they improve the outcomes of those students, regardless of the students' prior achievement. Thus, a school that improves the test scores of the lowest-performing students would be determined to have a higher value added than a school that made no change to the test scores of the highest performing students, even though the latter school's students may perform better on average.

A school's value added is calculated by first removing the portion of each student's test score that is due to non-school factors. To do so, I regress student test scores z_{isgdt} on cubic polynomials in prior test scores $\boldsymbol{z}_{ig}$ and demographic characteristics $\boldsymbol{X}_{it}$ as in equation (4). The cubic polynomials in prior test scores account for mean reversion and the fact that students with low test scores have more room to improve than students with high test scores. I allow the coefficients on prior test scores $\boldsymbol{\delta}_{gd}$ to differ by grade and subgroup to account for the fact that learning deprecation may differ between grades and subgroups may be on different growth trajectories. I also include cohort fixed effects γ_{gt}. The demographic characteristics $\boldsymbol{X}_{it}$ contain a linear term for age and fixed effects for sex, ethnicity[10], economic disadvantage, limited English proficiency, and disability status.

$$z_{isgdt} = \boldsymbol{z}_{ig}\boldsymbol{\delta}_{gd} + \boldsymbol{X}_{it}\boldsymbol{\beta}_X + \gamma_{gt} + r_{isgdt} \tag{4}$$

As a robustness check, in some analyses I augment equation (4) by including additional ACS, older-sibling, and peer controls, as in equation (5). The ACS controls include the proportion of a student's Census tract that is Hispanic, Asian, black, and some other race, the proportion of a student's Census tract that has below a high school degree, that has a high school degree, that has an associate's degree, and that has a bachelor's degree or higher, the proportion of families in a student's Census tract that fall below the poverty line, and the median household income in a student's Census tract. In the test-score analyses, older-sibling controls include the average test score of a student's older siblings in a given year. In the postsecondary

[9]Value added methodologies were first pioneered in estimating school and hospital quality (Willms and Raudenbush, 1989; McClellan and Staiger, 1999, 2000). Meyer (1997) and Everson (2017) provide some background on the methodology.

[10]Asian, Hispanic, black, and other; white is omitted.

enrollment analyses, older-sibling controls include the proportion of a student's older siblings that enrolled in a 2-year college and 4-year university. Peer controls are leave-one-out (jackknife) averages of the base controls $\boldsymbol{z}_{ig}$ and $\boldsymbol{X}_{it}$, and, as with a student's own prior test scores, I allow the coefficients on peer prior test scores to differ by grade and subgroup. This allows peers to have differential impacts on different types of students (Carrell, Sacerdote and West, 2013).

$$z_{isgdt} = \boldsymbol{z}_{ig}\boldsymbol{\delta}_{gd} + \boldsymbol{X}_{it}\boldsymbol{\beta}_X + N_{it}\boldsymbol{\beta}_N + F_{it}\boldsymbol{\beta}_F + \bar{\boldsymbol{z}}_{i^-sgt}\boldsymbol{\lambda}_{gd} + \bar{\boldsymbol{X}}_{i^-sgt}\boldsymbol{\beta}_P + \gamma_{gt} + r_{isgdt} \tag{5}$$

For each category of subgroup (SES, ethnicity, and sex) I omit the fixed effects for the corresponding variable when running equation (4), as including these fixed effects would remove potential differences in the average value added that schools provide to each subgroup. For example, when estimating school value added for low- and high-SES students, I omit the economic disadvantage fixed effect, as including it would force the average school value added for low-SES students to be equal to the average school value added for high-SES students.

Because there could be idiosyncratic shocks that are uncorrelated with school quality but influence the performance of all students within a school in each year, such as the proverbial dog barking outside of a school on the day of the test, the residual term r_{isgdt} will contain school by subgroup value added μ_{sdt}, idiosyncratic shocks θ_{st}, and a student-level error term ε_{isgdt} as in equation (6).

$$r_{isgdt} = \mu_{sdt} + \theta_{st} + \varepsilon_{isgdt} \tag{6}$$

Under the assumptions that ε_{isgdt} is a mean zero random error term, students do not sort to schools in each year on unobservable characteristics, and subgroups are not unobservably different from each other, the student-level error terms have expected value zero conditional on school, year, and subgroup, which gives us equation (7).

$$\mathbf{E}[r_{isgdt}|s, d, t] = \mu_{sdt} + \theta_{st} \tag{7}$$

Note that the assumption on the error term here is stronger than in Naven (2022), which only required that students do not sort to schools in each year on unobservable characteristics. When allowing for within-school heterogeneity, it must be the case that on average there are no unobservable differences between students in different subgroups. If this assumption is violated, then there will appear to be differences in the average value added that schools provide to each subgroup when in fact this is just picking up differences in student performance due to unobserved factors that are correlated with subgroup. This is an especially strong assumption due to the fact that I drop the fixed effect that would account for unobservable differences between subgroups, as this fixed effect is collinear with the existence of differences in average school value added by subgroup. Nevertheless, Kane and Staiger (2008) and Deming (2014) provide evidence that demographic characteristics are essentially irrelevant after conditioning on prior test scores when estimating value added on test scores, so the vector of prior test scores will likely be sufficient to satisfy this assumption.

I therefore average the residual r_{isgdt} to the school-by-subgroup-by-year level in order to eliminate the student-level error term. However, because value added by subgroup and idiosyncratic shocks are the same for all students at this level, the average residual will contain both school by subgroup value added and the school-level idiosyncratic shock as in equation (8).

$$\bar{r}_{sdt} = \mu_{sdt} + \theta_{st} \tag{8}$$

In order to reduce the variation from the idiosyncratic shocks while retaining the variation in school value added, I project the average residual in year t onto the average residuals in all other years t' (jackknife projection) as in equation (9).

$$\bar{r}_{sdt} = \bar{\boldsymbol{r}}_{sdt'}\boldsymbol{\beta}_{\bar{r}t'} + \epsilon_{sdt} \tag{9}$$

The value added by subgroup estimates that I use in this paper are the predicted values from equation (9), $\hat{\mu}_{sdt} = \bar{\boldsymbol{r}}_{sdt'}\hat{\boldsymbol{\beta}}_{\bar{r}t'}$. However, I rescale the estimates so that they have mean zero, thus, schools with positive value added are above average and vice versa.[11] The appendix of Naven (2022) outlines additional methodological details under the assumption of homogenous school quality.

This projection strategy has several advantages. Under the assumptions that school by subgroup value added is correlated across years ($\text{cov}(\mu_{sdt}\mu_{sdt'}) \neq 0$), the school-level common shocks are uncorrelated across years ($\text{cov}(\theta_{st}\theta_{st'}) = 0$), and the school-level common shocks are not correlated with school by subgroup value added across years ($\text{cov}(\mu_{sdt}\theta_{st'}) = 0$), the projection will retain variation from school by subgroup value added and remove variation from the common shocks. In practice, the finite sample size in the number of years may lead to violations of the last two assumptions regarding θ_{st}, which is why the projection will reduce the variation from the idiosyncratic shocks instead of completely eliminating it.[12] Appendix section B provides validity tests showing that noise is not substantially driving the distribution of value added

[11] This rescaling has no impact on the results to follow.

[12] If the common shocks are truly idiosyncratic, then the last two assumptions regarding θ_{st} are likely to hold as the number of years goes to infinity. Furthermore, to the extent that good or bad events happen continuously at the same schools, these should be considered part of a school's value added, which further reinforces that the common shocks are idiosyncratic. As for the first assumption, schools will experience some faculty and staff turnover, but school by subgroup value added is likely to be correlated from year to year as the majority of the personnel will remain in the same school from one year to the next.

estimates, that the school by subgroup value added estimates increase student test scores by the correct magnitude, and that forecast bias, which is the proportion of the variance in estimated school by subgroup value added that is due to correlation between unobserved student ability and the school by subgroup value added estimates, is essentially zero.

4.3 Test Score Results

Figures 3, 4, and 5 give summary statistics for the test score value added estimates. Panels (a) and (c) give results when controlling for the base controls described in equation (4). Panels (b) and (d) give results for the subsample of students that have at least one older sibling and can be matched to the ACS and control for ACS Census tract characteristics, average older-sibling test scores, and peer characteristics, as in equation (5).

Figures 3a, 3b, 4a, 4b, 5a, and 5b show the kernel density estimates of school by subgroup value added, $\hat{\mu}_{sdt}$, for each subgroup. The bottom of each graph gives the overall value added mean and standard deviation, the value added mean and standard deviation for each subgroup, the difference in average school value added between subgroups, and a Kolmogorov-Smirnov test for equality of distribution (Kolmogorov, 1933; Smirnov, 1939).

Figures 3c, 3d, 4c, 4d, 5c, and 5d show scatter plots of an individual school's value added for each subgroup of students. Each point represents an observation for a school in a specific year, with the horizontal axis giving the school's value added for high-SES/non-minority/female students and the vertical axis giving the school's value added for low-SES/minority/male students. The red line gives the points at which value added is equal for each subgroup.

4.3.1 Results by Socioeconomic Status

Figure 3a, which controls for standard value added covariates, suggests that on average schools provide similar value added to low-SES students as they do to high-SES students. The average school provides low-SES students 0.003 student-level standard deviations less value added on average. The overall standard deviation of school value added is 0.087, which suggests that low-SES students attend the equivalent of a school that is 0.03 $\left(= \frac{0.003}{0.087}\right)$ standard deviations below average. Thus within-school heterogeneity in value added does not appear to be a large factor in the test score achievement gap between low- and high-SES students.

The results are largely similar after controlling for additional neighborhood, sibling, and peer characteristics, as seen in Figure 3b. Including these additional controls slightly increases the difference in average school value added to 0.007 student-level standard deviations, which is the equivalent of attending a school 0.07 $\left(=\frac{0.007}{0.098}\right)$ standard deviations below average. Within-school heterogeneity in value added thus accounts for 0.8% $\left(=\frac{0.007}{0.690-(-0.185)}\right)$ of the test score achievement gap between low- and high-SES students, which suggests that between-school heterogeneity in value added and factors outside of school are the primary causes of the achievement gap between low- and high-SES students. Figure 3d shows that high-value-added schools tend to provide high value added for all students, as the correlation between value added to low- and high-SES students is 0.826 within schools.

4.3.2 Results by Ethnicity

Figure 4a suggests that schools provide less value added to minority students by 0.011 student-level standard deviations on average. The overall standard deviation of school value added is 0.087, which suggests that minority students attend the equivalent of a school that is 0.13 $\left(=\frac{0.011}{0.087}\right)$ standard deviations below average. However, Figure 4b shows that controlling for additional neighborhood, sibling, and peer characteristics somewhat shrinks the difference in value added that schools provide to minority students relative to non-minority students. Using this expanded set of controls, which are typically not included in standard value added models, decreases the difference in average school value added to 0.008 student-level standard deviations, which is the equivalent of attending a school 0.08 $\left(=\frac{0.008}{0.098}\right)$ standard deviations below average. Within-school heterogeneity in value added thus accounts for 1% $\left(=\frac{0.008}{0.674-(-0.164)}\right)$ of the test score achievement gap between minority and non-minority students. As with the SES achievement gap, it appears that the vast majority of the achievement gap between minority and non-minority students occurs due to differences in non-school factors between minority and non-minority students as well as between-school differences in school value added. Figure 4d shows that high-value-added schools tend to provide high value added for all students, with a correlation of 0.814 between minority and non-minority value added within schools.

4.3.3 Results by Sex

Figure 5a suggests that schools provide less value added to male students by 0.012 student-level standard deviations on average. The overall standard deviation of school value added is 0.085, which suggests that male students attend the equivalent of a school that is 0.14 $\left(=\frac{0.012}{0.085}\right)$ standard deviations below average. Controlling for additional neighborhood, sibling, and peer characteristics, as in Figure 5b, still leaves a meaningful difference in the value added that schools provide to male students relative to female students. While using this expanded set of controls decreases the difference in average school value added to 0.009 student-level standard deviations, which is the equivalent of attending a school 0.09 $\left(=\frac{0.009}{0.097}\right)$ standard deviations below average, the difference in value added that schools provide to male students relative to female students still accounts for 6% $\left(=\frac{0.009}{0.298-0.136}\right)$ of the achievement gap between men and women. Thus it appears that schools operate in a fashion that slightly benefits female students and somewhat contributes to the differences in achievement on ELA exams that we see between men and women. Figure 5d shows a very high correlation of 0.862 between male and female value added within schools, providing further evidence that schools may simply provide education in ways that are more conducive to learning for women.

4.4 Postsecondary Results

Much more important than test scores, however, is the effect of school value added on postsecondary enrollment, because attending college has proven to be a worthwhile investment for both the average and marginal student (Oreopoulos and Petronijevic, 2013). Hoekstra (2009) finds that attending a flagship university increases the earnings of white men by 20%, while Zimmerman (2014) shows that admission to a 4-year university for the marginal student gives a wage premium of 22% and bachelor's degree receipt for the marginal admission increases wages by 90%.

In order to estimate a school's value added on postsecondary enrollment directly, I reestimate equation (4) with an indicator for postsecondary enrollment as the dependent variable instead of a student's test score.[13] I define postsecondary enrollment as enrolling in any institution in the NSC data within one year of high-school graduation.[14] When additionally controlling for ACS, older-sibling, and peer controls, I replace the average test score of a student's older siblings in a given year with the proportion of a student's older siblings that enrolled in a 2-year college and 4-year university.

It should be noted that the assumptions to obtain unbiased estimates of school value added on postsecondary enrollment are stronger than those for school value added on test scores. Value added on test scores relies upon the assumption that prior test scores and demographic characteristics are sufficient to predict how a student would perform on the current year's test, such that any differences in test scores after controlling for these variables are attributable to schools. Prior research shows that this is a valid assumption (Kane and Staiger, 2008; Deming, 2014). Estimating value added on postsecondary enrollment, however, relies upon the assumption that prior test scores and demographic characteristics (plus neighborhood, sibling, and peer characteristics when included) are sufficient to predict the likelihood that a student will attend a postsecondary institution. This assumption may not hold, especially given that Abdulkadiroğlu et al. (2020) finds that the bias of value added estimates on postsecondary enrollment is larger than the bias of value added estimates on test scores at the high school level. Thus the results for school value added on postsecondary

[13]Because each student's enrollment outcome is invariant across grades, I only use observations from 11th grade.
[14]This prevents earlier cohorts of students from having higher potential match rates to the NSC data.

enrollment should be interpreted keeping these caveats in mind.

Table 4 gives summary statistics for the school postsecondary enrollment value added sample. As with test scores and demographic characteristics, there are substantial differences in postsecondary enrollment by SES, ethnicity, and sex. Low-SES students are 18 percentage points less likely to attend a postsecondary institution than high-SES students. The vast majority of this difference comes from differences in four-year university enrollment, as low-SES students are only one percentage point less likely to enroll in a community college than their high-SES peers. Low-SES students are also more likely to attend in-state and public institutions, which is intuitive given that these options tend to be cheaper than out-of-state and private institutions. Similar differences exist between minority and non-minority students. Men are seven percentage points less likely to attend a postsecondary institution and eight percentage points less likely to attend a four-year university than women.

4.4.1 Results by Socioeconomic Status

Figure 6a, which controls for standard value added covariates, suggests that schools provide less value added on postsecondary enrollment to low-SES students by 2.5 percentage points on average. The overall standard deviation of school value added is 0.084, which suggests that low-SES students attend the equivalent of a school that is 0.3 $\left(= \frac{0.025}{0.084}\right)$ standard deviations below average. However, Figure 6b shows that controlling for additional neighborhood, sibling, and peer characteristics essentially eliminates the difference in value added on postsecondary enrollment that schools provide to low-SES students relative to high-SES students. Using this expanded set of controls decreases the difference in average school value added to 0.1 percentage points, which is the equivalent of attending a school 0.03 $\left(= \frac{0.001}{0.040}\right)$ standard deviations below average. Within-school heterogeneity in value added thus accounts for 0.5% $\left(= \frac{0.001}{0.744-0.536)}\right)$ of the postsecondary enrollment gap between low- and high-SES students, which suggests that between-school heterogeneity in value added and factors outside of school are the primary causes of the postsecondary enrollment gap between low- and high-SES students. Figure 6d shows a relatively modest correlation of 0.546 between the value added that schools provide to low- and high-SES students, and there are more schools that provide relatively high value added for only one type of student than with value added on test scores.

4.4.2 Results by Ethnicity

Figure 7a suggests that schools provide less value added on postsecondary enrollment to minority students by 2 percentage points on average. The overall standard deviation of school value added is 0.088, which suggests that minority students attend the equivalent of a school that is 0.2 $\left(= \frac{0.020}{0.088}\right)$ standard deviations below average.

Controlling for additional neighborhood, sibling, and peer characteristics considerably shrinks the difference in value added on postsecondary enrollment that schools provide to minority students relative to non-minority students. Using this expanded set of controls decreases the difference in average school value added to 0.5 percentage points, which is the equivalent of attending a school 0.1 $\left(=\frac{0.005}{0.043}\right)$ standard deviations below average. Within-school heterogeneity in value added thus accounts for 2% $\left(=\frac{0.005}{0.749-0.537)}\right)$ of the postsecondary enrollment gap between minority and non-minority students, which suggests that within-school differences play a very small role in the postsecondary enrollment gap between minority and non-minority students. Figure 7d shows a large degree of dispersion in value added within schools after including these additional controls. While school value added for minority and non-minority students has a positive correlation of 0.448, there are many schools that provide above average value added to minority students and below average value added to non-minority students, and vice versa. Of all three analyses, the correlation in value added on postsecondary enrollment between minority and non-minority students is the lowest, suggesting that schools may have comparative advantages with particular ethnicities more-so than they do for particular SESs or sexes.

4.4.3 Results by Sex

Figure 8a suggests that schools provide less value added on postsecondary enrollment to male students by 2.3 percentage points on average. The overall standard deviation of school value added is 0.082, which suggests that male students attend the equivalent of a school that is 0.3 $\left(=\frac{0.023}{0.082}\right)$ standard deviations below average. The difference in value added on postsecondary enrollment that schools provide to male students relative to female students remains large even after controlling for additional neighborhood, sibling, and peer characteristics, as seen in Figure 8b. Using this expanded set of controls decreases the difference in average school value added to 1.5 percentage points, which is the equivalent of attending a school 0.4 $\left(=\frac{0.015}{0.042}\right)$ standard deviations below average. Within-school heterogeneity in value added thus accounts for 22% $\left(=\frac{0.015}{0.668-0.601)}\right)$ of the postsecondary enrollment gap between low- and high-SES students, which suggests that schools play a significant role in maintaining the disparities in postsecondary enrollment between men and women. Figure 8d shows that value added on postsecondary enrollment for male and female students is more correlated than for other subgroups, which, combined with the gap in average value added, suggests that schools are simply better at increasing the likelihood that women attend college than they are for men.

5 School Characteristics

Finally, I explore which school characteristics are correlated with school value added. While these regressions are not causal, they provide a description of what high-value-added schools have in common and whether there are differences in the characteristics of high-value-added schools by subgroup. This analysis may therefore provide clues of some effective characteristics that could be explored in a causal framework in future studies.

I run trivariate regressions of school by subgroup value added on school-level inputs as in equation (10). The independent variables in each regression are a school characteristic X_{st}, a subgroup fixed effect ϕ_d, and the school characteristic interacted with the subgroup fixed effect $\phi_d \times X_{st}$. The school characteristics used for X_{st} are the number of full-time equivalent (FTE) teachers per student, FTE pupil services staff[15] per student, English-learner staff per student, proportion teachers with three years or less experience, proportion teachers with full credentials, proportion male teachers, proportion minority teachers, instruction expenditures[16] per student, pupil services expenditures[17] per student, ancillary services expenditures[18] per student, other expenditures[19] per student, and general administration expenditures[20] per student. I drop the top and bottom 2.5% of each school characteristic in order to account for outliers and potential errors in the data that schools report.

$$\hat{\mu}_{sdt} = X_{st}\beta_1 + \phi_d \times X_{st}\beta_2 + \phi_d + \varepsilon_{st} \tag{10}$$

Table 5 shows the correlations between school by subgroup value added and school characteristics. The value added estimates used in this table control for the ACS, older-sibling, and peer controls in equation (5). Appendix Table C.1 shows the results when using the value added estimates that control for the base controls in equation (4). Columns one, three, and five use test score value added by subgroup as the dependent variable, while columns two, four, and six use postsecondary enrollment value added by subgroup as the dependent variable. Columns one and two give value added by SES, columns three and four give value added by ethnicity, and columns five and six give value added by sex. Each column pertains to a distinct set of value added estimates. Every two rows within each column, which are separated by solid black horizontal

[15] This includes counselors, psychologists, librarian/library/media teachers, social workers, nurses, and speech/language/hearing specialists.

[16] This combines the "instruction" and "instruction-related services" function classifications.

[17] This includes guidance and counseling services, psychological services, health services, speech pathology and audiology services, pupil testing services, pupil transportation, food services, and other pupil services.

[18] School-sponsored activities during or after the school day that are generally designed to provide students with experiences such as motivation, enjoyment, and improvement of skills in either a competitive or noncompetitive setting. This includes school-sponsored co-curricular activities, school-sponsored athletics, and other ancillary services.

[19] This combines the "community services", "enterprise", "plant services", and "other outgo" function classifications.

[20] Agency-wide administrative activities that are accounted for in the general fund.

lines, are a separate regression.

While emphasizing that these regressions are not causal, there are many interesting differential correlations by subgroup. Schools with more teachers per student (i.e. smaller class sizes) have higher value added on test scores for all students, although the effect is more positive for high-SES and non-minority students. However, more teachers per student is negatively correlated with value added on postsecondary enrollment for high-SES students and positively correlated for low-SES students.

Schools with more pupil services, which includes counselors, per student are associated with lower value added on test scores for all students but significantly increase the likelihood that low-SES and minority students eventually enroll in college. Unsurprisingly, given the fact that most of California's low-SES and minority students are Hispanic, schools with more English Learner staff per student are associated with relatively higher value added on test scores for low-SES and minority students, although this does not translate to higher value added on postsecondary enrollment.

I also find that while having more teachers with greater than three years of experience and more teachers with full credentials is associated with higher value added on test scores for all types of students, low-SES and minority students benefit less than their high-SES and non-minority peers in schools with high proportions of teachers with full credentials. Somewhat surprisingly, having more teachers with three years or less of experience is correlated with higher value added on postsecondary enrollment for all types of students.

With regards to the identity of the school's faculty, schools with more male teachers provide lower value added on test scores to all students, with relatively smaller effects for low-SES and minority students, but higher value added on postsecondary enrollment for all students except, interestingly, male students. Schools with more minority faculty are associated with lower value added on test scores for high-SES, non-minority, male, and female students, but these negative associations are completely mitigated for low-SES and minority students. Schools with more minority faculty are associated with higher value added on postsecondary enrollment for all students except male students.

With respect to how schools spend their finances, there are fewer clear patterns. The two most consistent patterns are that increases in pupil services expenditures are associated with lower value added on both test scores for all students, but that the correlation is much less negative for low-SES and minority students. I also find that schools with more ancillary expenditures per student are associated with lower test score value added for all students, with a relatively less negative correlation for male students, and an increase in postsecondary enrollment value added for low-SES, minority, and male students. This could potentially be due to increases in the likelihood of receiving an athletic scholarship to a university.

6 Conclusion

Achievement gaps in education by SES, ethnicity, and sex continue to persist despite efforts to close them. This paper explores to what extent within-school heterogeneity in school quality by SES, ethnicity, and sex contribute to the existence of these achievement gaps. I investigate this issue by performing three separate analyses in which I allow each school to have distinct measures of school quality for low-/high-SES, minority/non-minority, and male/female students. I therefore do not impose the assumption that schools have an equal impact on all students enrolled in the school but instead allow for the possibility that schools provide more quality to certain subgroups of students.

Using data on the universe of public school students in California, I estimate school by subgroup quality on both standardized test scores and postsecondary enrollment by applying the value added with drift methodology, as in Chetty, Friedman and Rockoff (2014), to schools. The value added methodology accounts for the fact that students do not randomly sort to schools and subgroups of students have varying levels of average academic performance. Allowing for within-school heterogeneity by subgroup relaxes the assumption that schools must have the same impact on all types of students enrolled in the school. Simulations suggest that assuming homogeneity in school value added incorrectly ranks schools in the presence of within-school heterogeneity in value added. Allowing for this within-school heterogeneity when estimating school value added restores the correct ranking of schools. To my knowledge this paper is the first to allow for within-school heterogeneity in how much schools increase the postsecondary enrollment of their students.

While standard value added models suggest some within-school heterogeneity in value added on test scores may exist, models with additional neighborhood, sibling, and peer controls suggest that within-school heterogeneity plays essentially no role in the SES and ethnicity achievement gaps and a modest role in the sex achievement gap. The difference in the average value added on test scores provided to male and female students is equivalent to attending a school that is 0.09 standard deviations below average, and this difference accounts for 6% of the test score achievement gap.

Within-school heterogeneity in value added on postsecondary enrollment plays a somewhat larger role in enrollment gaps than does within-school heterogeneity in value added on test scores in achievement gaps. The difference in the average value added on postsecondary enrollment provided to male and female students is equivalent to attending a school that is 0.4 standard deviations below average, and this difference accounts for 22% of the enrollment gap. The difference for minority and non-minority students is equivalent to attending a school that is 0.1 standard deviations below average and accounts for 2% of the enrollment gap, and the difference for low- and high-SES students is equivalent to attending a school that is 0.03 standard deviations below average and accounts for 0.5% of the enrollment gap.

When correlating school value added measures with school characteristics, I find that certain school characteristics are correlated with the value added for various subgroups of students in different ways. Low student-teacher ratios are associated with relatively higher value added on test scores for high-SES and non-minority students but relatively higher value added on postsecondary enrollment for low-SES students. English learner staff per student is associated with relatively higher value added on test scores for low-SES and minority students. A higher proportion of male teachers is associated with lower value added on test scores and higher value added on postsecondary enrollment, with the exception that having more male teachers is uncorrelated with value added on postsecondary enrollment for male students. Low-SES and minority students are unaffected by a higher proportion of minority teachers with regards to test scores, while value added for other groups is negatively correlated with a higher proportion of minority teachers.

While within-school heterogeneity in quality explains little of the test score or postsecondary enrollment gaps by SES or ethnicity, within-school differences in quality explain a modest portion of the test score achievement gap between men and women and a large portion of the enrollment gap between men and women. Particularly given that men are less likely to enroll in college and that college graduates earn higher wages in the labor market, schools should examine whether they are tailoring their instruction and other services to students who are already succeeding to assure that they don't leave other types of students behind.

References

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

American Community Survey. 2019. "Full-Time, Year-Round Workers & Median Earnings by Sex & Occupation."

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.

Bailey, Martha J, and Susan M Dynarski. 2011. "Gains and gaps: Changing inequality in US college entry and completion." National Bureau of Economic Research.

Bau, Natalie. 2022. "Estimating an equilibrium model of horizontal competition in education." *Journal of Political Economy*, 130(7): 1717–1764.

Belley, Philippe, and Lance Lochner. 2007. "The changing role of family income and ability in determining educational achievement." *Journal of Human capital*, 1(1): 37–89.

Bitler, Marianne, Sean P Corcoran, Thurston Domina, and Emily K Penner. 2021. "Teacher effects on student achievement and height: A cautionary tale." *Journal of Research on Educational Effectiveness*, 14(4): 900–924.

Carrell, Scott E, Bruce I Sacerdote, and James E West. 2013. "From natural variation to optimal policy? The importance of endogenous peer group formation." *Econometrica*, 81(3): 855–882.

Carrell, Scott E, Marianne E Page, and James E West. 2010. "Sex and science: How professor gender perpetuates the gender gap." *The Quarterly Journal of Economics*, 125(3): 1101–1144.

Carrell, Scott, Michal Kurlaender, Paco Martorell, Matthew Naven, and Christina Sun. 2023. "Do Schools Matter? Measuring the Impact of California High Schools on Test Scores and Postsecondary Enrollment."

Chatterji, Madhabi. 2006. "Reading achievement gaps, correlates, and moderators of early reading achievement: Evidence from the Early Childhood Longitudinal Study (ECLS) kindergarten to first grade sample." *Journal of educational psychology*, 98(3): 489.

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

Cimpian, Joseph R, Sarah T Lubienski, Jennifer D Timmer, Martha B Makowski, and Emily K Miller. 2016. "Have gender gaps in math closed? Achievement, teacher perceptions, and learning behaviors across two ECLS-K cohorts." *AERA Open*, 2(4): 2332858416673617.

Dee, Thomas S. 2004. "Teachers, race, and student achievement in a randomized experiment." *Review of Economics and Statistics*, 86(1): 195–210.

Dee, Thomas S. 2005. "A teacher like me: Does race, ethnicity, or gender matter?" *American Economic Review*, 95(2): 158–165.

Dee, Thomas S. 2007. "Teachers and the gender gaps in student achievement." *Journal of Human Resources*, 42(3): 528–554.

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

Egalite, Anna J, Brian Kisida, and Marcus A Winters. 2015. "Representation in the classroom: The effect of own-race teachers on student achievement." *Economics of Education Review*, 45: 44–52.

Everson, Kimberlee C. 2017. "Value-Added Modeling and Educational Accountability: Are We Answering the Real Questions?" *Review of Educational Research*, 87(1): 35–70.

Fryer Jr, Roland G, and Steven D Levitt. 2010. "An empirical analysis of the gender gap in mathematics." *American Economic Journal: Applied Economics*, 2(2): 210–40.

Gershenson, Seth, Cassandra MD Hart, Joshua Hyman, Constance A Lindsay, and Nicholas W Papageorge. 2022. "The long-run impacts of same-race teachers." *American Economic Journal: Economic Policy*, 14(4): 300–342.

Hoekstra, Mark. 2009. "The effect of attending the flagship state university on earnings: A discontinuity-based approach." *The Review of Economics and Statistics*, 91(4): 717–724.

Holt, Stephen B, and Seth Gershenson. 2019. "The impact of demographic representation on absences and suspensions." *Policy Studies Journal*, 47(4): 1069–1099.

Husain, Muna, and Daniel L Millimet. 2009. "The mythical 'boy crisis'?" *Economics of Education Review*, 28(1): 38–48.

Kahn, Shulamit, and Donna Ginther. 2017. "Women and STEM." National Bureau of Economic Research.

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

Karen, David. 2002. "Changes in access to higher education in the United States: 1980-1992." *Sociology of Education*, 191–210.

Kolmogorov, Andrey. 1933. "Sulla determinazione empirica di una lgge di distribuzione." *Inst. Ital. Attuari, Giorn.*, 4: 83–91.

Lee, Joohi, Sungseek Moon, and Rebecca L Hegar. 2011. "Mathematics skills in early childhood: Exploring gender and ethnic patterns." *Child Indicators Research*, 4(3): 353–368.

Lindsay, Constance A, and Cassandra MD Hart. 2017. "Exposure to same-race teachers and student disciplinary outcomes for Black students in North Carolina." *Educational Evaluation and Policy Analysis*, 39(3): 485–510.

McClellan, Mark, and Douglas Staiger. 1999. "The quality of health care providers." National bureau of economic research.

McClellan, Mark, and Douglas Staiger. 2000. "Comparing the quality of health care providers." In *Frontiers in Health Policy Research, Volume 3.* 113–136. MIT Press.

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

National Center for Education Statistics. 2022*a*. "Bachelor's degrees conferred by postsecondary institutions, by race/ethnicity and sex of student: Selected academic years, 1976-77 through 2020-21."

National Center for Education Statistics. 2022*b*. "Total fall enrollment in degree-granting postsecondary institutions, by attendance status, sex of student, and control of institution: Selected years, 1947 through 2031."

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

Oreopoulos, Philip, and Uros Petronijevic. 2013. "Making college worth it: A review of research on the returns to higher education." National Bureau of Economic Research.

Penner, Andrew M, and Marcel Paret. 2008. "Gender differences in mathematics achievement: Exploring the early grades and the extremes." *Social Science Research*, 37(1): 239–253.

Reardon, Sean F. 2011. "The widening academic achievement gap between the rich and the poor: New evidence and possible explanations." *Whither opportunity*, 91–116.

Reardon, Sean F. 2016. "School district socioeconomic status, race, and academic achievement." *Stanford Center for Educational Policy Analysis.*

Reardon, Sean F, and Claudia Galindo. 2009. "The Hispanic-White achievement gap in math and reading in the elementary grades." *American Educational Research Journal*, 46(3): 853–891.

Reardon, Sean F, and Joseph P Robinson. 2008. "Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps." *Handbook of research in education finance and policy*, 497–516.

Reardon, Sean F, and Ximena A Portilla. 2016. "Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry." *AERA Open*, 2(3): 2332858416657343.

Reardon, Sean F, Rachel B Baker, and Daniel Klasik. 2012. "Race, income, and enrollment patterns in highly selective colleges, 1982-2004."

Robinson, Joseph Paul, and Sarah Theule Lubienski. 2011. "The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings." *American Educational Research Journal*, 48(2): 268–302.

Rothstein, Jesse. 2017. "Measuring the Impacts of Teachers: Comment." *American Economic Review*, 107(6): 1656–84.

Smirnov, Nikolai V. 1939. "Estimate of deviation between empirical distribution functions in two independent samples." *Bulletin Moscow University*, 2(2): 3–16.

Sohn, Kitae. 2012. "A new insight into the gender gap in math." *Bulletin of Economic Research*, 64(1): 135–155.

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.

Zimmerman, Seth D. 2014. "The returns to college admission for academically marginal students." *Journal of Labor Economics*, 32(4): 711–754.

A Data

Table A.1 gives the number of observations in the CST data conditional on a set of restrictions implemented in order to form the value added 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, etc. The first row denotes the total number of observations in the CST dataset. The second row keeps students who have information on test scores, as opposed to just demographic characteristics. 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. I impose this restriction because students repeating a grade are tested on material for which they have already been tested at least once. The fourth row keeps only students at "conventional" schools. This includes schools in the following categories defined by the 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).[21] The fifth row drops any

[21] 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,

schools that enroll 10 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 and ninth rows drop students who are missing test scores from one grade and two grades prior, respectively. This restriction is the cause of the vast majority of observations which are excluded from the value added estimates. The tenth row drops observations for which peer averages of the control variables could not be calculated. The eleventh row drops students if fewer than seven observations can be used to estimate value added for their school by subgroup by year cell, which insures that all value added estimates are based on at least seven observations.

Table A.2 gives summary statistics by subgroup for the subsample of the test score value added sample that has at least one older sibling and can be matched to the American Community Survey. The table includes the dependent variable and the ACS and sibling controls used in the value added analyses. Low-SES and minority students live in neighborhoods with more Hispanic, black, and other-race residents, fewer white and Asian residents, more residents who dropped out of high school or whose highest degree is a high school diploma, fewer residents with an associates, bachelors, or higher degree, more families below the poverty line, and lower median household incomes than the neighborhoods that high-SES and non-minority students live in. Unsurprisingly, there are few differences in neighborhood characteristics between male and female students, as students do not sort to neighborhoods based on sex in the same way that students sort to neighborhoods based on SES or race.

Table A.3 gives summary statistics for school by grade level peer (jackknife) averages for the subsample of the test score value added sample that has at least one older sibling and can be matched to the American Community Survey. This table provides evidence of the sorting of students that occurs at the school level. The average low-SES student attends a school where 71% of their peers are also economically disadvantaged and 73% of their peers are minorities. In contrast, the average high-SES student attends a school where 30% of their peers are economically disadvantaged and 38% are minorities. These stark differences are similar between minority and non-minority students. The average minority student attends a school with 68% low-SES students and 73% minority students, while the average non-minority student attends a school with 34% low-SES students and 37% minority students. Males and females attend schools with similar peers, as one would expect given that school sorting occurs at the residential level on socioeconomic and ethnic factors but not according to student gender.

Table A.4 gives summary statistics for the subsample of the school postsecondary enrollment value added sample that has at least one older sibling and can be matched to the American Community Survey. The

Alternative Schools of Choice, Continuation High Schools, District Community Day Schools, Adult Education Centers, and Regional Occupational Center/Program (ROC/P).

older siblings of low-SES and minority students are less likely to enroll in 4-year universities than the older siblings of high-SES and non-minority students, although the older siblings of all students are similarly likely to enroll in 2-year colleges.

B Validity Tests

There are three potential concerns regarding the validity of the value added estimates. The first is that the estimates may be picking up noise due to sampling error and small sample variability. This would be the case if test scores are sufficiently noisy that student-level residual test scores, ε_{isgdt}, do not average out to zero in each school by subgroup cell, even when schools have no effect on student performance (Bitler et al., 2021). If this were the case, we would attribute value added to schools when we were in fact just observing sampling error.

In order to measure how much of the estimated variation in school by subgroup value added is due to noise, I calculate school by subgroup value added estimates after randomly assigning students to schools. I call these value added estimates permuted value added, as I permute the school assignment vector within a grade by year cell. Figures B.1a, B.1b, and B.1c show the distributions of permuted value added by subgroup for the base sample with base controls, and I plot the distributions on the same axes as Figures 3a, 4a, and 5a so that the variability can be directly compared. Unlike in Naven (2022), there remains a distribution of estimated school by subgroup value added even after randomly assigning students to schools. While Naven (2022) found a variance of essentially zero after randomly assigning students to schools, I find standard deviations of school value added as large as 0.016. Assuming that this variation is purely due to measurement error, then 16% $\left(=\frac{0.014}{0.087}\right)$, 18% $\left(=\frac{0.016}{0.087}\right)$, and 14% $\left(=\frac{0.012}{0.096}\right)$ of the standard deviation in estimated school by subgroup value added is due to noise for SES, ethnicity, and sex, respectively.[22]

It is intuitive that I will find noisier estimates when estimating school by subgroup value added as opposed to school value added as in Naven (2022) because the cell sizes here are smaller. Because each school is effectively cut in half, so that there are twice as many schools but with fewer students, the law of large numbers regarding $\mathbf{E}[\varepsilon_{isgdt}|s,d,t]=0$ is less likely to kick in. In simpler terms, there is more likely to be noise in an average calculated using a subsample than in an average calculated using the entire sample. Nevertheless, the vast majority of the variation in school value added is still due to signal as opposed to noise, so these results alleviate concerns that the value added estimates are purely an artifact of noisy test score measures or small sample variability.

[22]These calculations come from dividing the standard deviation of school by subgroup permuted value added from Figures B.1a, B.1b, and B.1c by the standard deviation of school by subgroup value added from Figures 3a, 4a, and 5a.

Figures B.1a, B.1b, and B.1c also show differences in the average permuted value added between subgroups. This, however, is expected given that we observe differences in the actual value added that schools provide to subgroups. For example, let the subscript p denote a student's permuted school (i.e. the school to which they are randomly assigned). If the distribution of school value added to high-type students has mean μ and the distribution of school value added for low-type students has mean $\mu - x$, then we have

$$\begin{aligned}\mathbf{E}[r_{isgHt}|p,H,t] &= \mathbf{E}[\mu_{sHt}|p,H,t] + \mathbf{E}[\theta_{st}|p,H,t] + \mathbf{E}[\varepsilon_{isgHt}|p,H,t] \\ &= \mathbf{E}[\mu_{sHt}|H] + \mathbf{E}[\theta_{st}] + \mathbf{E}[\varepsilon_{isgHt}] \\ &= \mu\end{aligned} \tag{11}$$

$$\begin{aligned}\mathbf{E}[r_{isgLt}|p,L,t] &= \mathbf{E}[\mu_{sLt}|p,L,t] + \mathbf{E}[\theta_{st}|p,L,t] + \mathbf{E}[\varepsilon_{isgLt}|p,L,t] \\ &= \mathbf{E}[\mu_{sLt}|L] + \mathbf{E}[\theta_{st}] + \mathbf{E}[\varepsilon_{isgLt}] \\ &= \mu - x\end{aligned} \tag{12}$$

where the second lines in equations (11) and (12) follow because μ_{sdt}, θ_{st}, and ε_{isgdt} are independent of p by definition of random assignment and independent of t by assumption of stationarity, and θ_{st}, and ε_{isgdt} are independent of d by assumption. Therefore even after randomly assigning students to schools we would expect to see a difference in the average permuted value added provided to each subgroup, although in theory we would expect no variation in permuted value added.

Another concern is that the value added estimates are the incorrect magnitude. Specifically, the issue is whether a one unit increase in school value added actually is associated with a one standard deviation increase in student test scores. In order to test this, I run a bivariate regression of residualized test scores r_{isgdt} on the school value added estimates $\hat{\mu}_{sdt}$, where the residualized test scores are calculated using equations (4) and (5). This follows the procedure used in Chetty, Friedman and Rockoff (2014) and Rothstein (2017) and calculates by how much a school's estimated value added actually increases the test scores of its students. We expect the coefficient to equal one, which would indicate that a one unit increase in school value added increases student test scores by one standard deviation on average.

The first row of Table B.1 provides this estimate for the base sample with base controls along with its 95% confidence interval. The coefficient estimates range from 1.005 to 1.008, which are statistically different than one but economically close to the expected coefficient. Chetty, Friedman and Rockoff (2014) obtain a coefficient estimate of 0.998. This gives evidence that the school by subgroup value added estimates have the correctly-sized effect on student test scores. Furthermore, Figures B.2a, B.2b, and B.2c graph the relationship between r_{isgdt} and $\hat{\mu}_{sdt}$ in 20 equally sized bins by subgroup for the base sample and base controls. Results show that the value added estimates and test score residuals have an almost perfectly linear relationship throughout the value added distribution and that there are essentially no differences by subgroup. The third row of Table B.1 provides this estimate for the subsample of the test score value added sample that has at least one older sibling and can be matched to the American Community Survey after controlling for neighborhood characteristics, average older-sibling test scores, and peer characteristics. The coefficient estimates are slightly larger in magnitude than for the base sample with base controls, which makes it somewhat ambiguous whether including these additional controls provides less biased value added estimates.

The final concern, and potentially most problematic, involves the potential sorting of students to schools based on unobserved ability. If students with high unobserved ability sort to specific schools, such that $cov(\varepsilon_{isgdt}, \mu_{sdt}) \neq 0$, then these schools' estimated value added will be higher than their true value added. However, this is only an issue if the sorting occurs on *unobserved* ability. Hence, there is no issue if students sort to schools on observed ability, because this will be controlled for with the inclusion of prior test scores and demographic controls. For example, if students with high test scores tend to attend the same schools, as occurs in practice, then we can still obtain unbiased estimates of school by subgroup value added as long as prior test scores are included in the control vector so that $\mathbf{E}[\varepsilon_{isgdt}|s, d, t] = 0$. In fact, research comparing test score value added estimates to estimates obtained using random assignment to schools (Deming, 2014; Angrist et al., 2017) or teachers (Kane and Staiger, 2008) shows that after controlling for prior test scores even the inclusion of demographic characteristics in the control vector is essentially irrelevant because prior test scores are a sufficient statistic for student ability.

The primary threat to this assumption would be if students or parents *changed* their level of input into academic preparation between the student's prior grade and current grade and students sorted to schools based on this change in behavior. For example, if all students of parents who received an increase in income between grades, where the extra income was used to purchase academic assistance, attended the same school, then the estimated value added of this school would be positively biased. This is due to the fact that the prior test scores and demographic controls of those students would not control for this change in academic assistance, so $\mathbf{E}[\varepsilon_{isgdt}|s, d, t] > 0$. If students whose parents *consistently* have high income sort to the same schools there would not be the same issue, because the students' prior test scores would also reflect their high SES.

The issue in determining to what degree students sort to schools on unobserved ability is that, by definition, we have no measures of unobserved ability. However we can approximate unobserved ability using variables in our data that likely would be correlated with ability but that were not included as a control variable in equation (4). Given the available data, the best possible measure of unobserved student ability is an additional prior test score. Under the assumption that this omitted variable is the only component of ε_{isgdt} that is correlated with student test scores, we can then obtain an estimate of $\frac{cov(\varepsilon_{isgdt},\hat{\mu}_{sdt})}{var(\hat{\mu}_{sdt})}$. Chetty, Friedman and Rockoff (2014) call this value forecast bias, which gives an estimate of what proportion of the variation in school by subgroup value added is due to sorting on unobserved ability.[23]

The second row of Table B.1 provides the estimate of forecast bias along with its 95% confidence interval. Here we expect an estimate of zero, which would give evidence that there is no sorting of students to schools on unobservable characteristics. The estimates suggest that between 0.8% (SES and ethnicity) and 1% (sex) of the variance in school value added is due to sorting on unobserved ability, thus selection on unobservables does not appear to be a large issue. Chetty, Friedman and Rockoff (2014) estimate forecast bias of 2.2%. Given that the forecast bias estimates are all negative, this would suggest that students who are unobservably worse tend to attend schools with higher estimated value added. This would result in value added estimates that are closer to zero, thus, if anything, the value added estimates are slightly conservative. Figure B.2 shows that this relationship holds throughout the distribution of school value added and that there are no differences in the relationship by subgroup. The fourth row of Table B.1 provides this estimate for the subsample of the test score value added sample that has at least one older sibling and can be matched to the American Community Survey after controlling for neighborhood characteristics, average older-sibling test scores, and peer characteristics. The coefficient estimates are positive instead of negative and slightly larger in magnitude than for the base sample with base controls, with sorting on unobservables accounting for up to 1.6% of the variance in school value added. While the estimates that control for neighborhoods, older siblings, and peers should theoretically be less biased than those that include only the base controls, forecast bias tests suggest slightly more bias (with respect to three-grade-prior test scores) after including the additional controls.

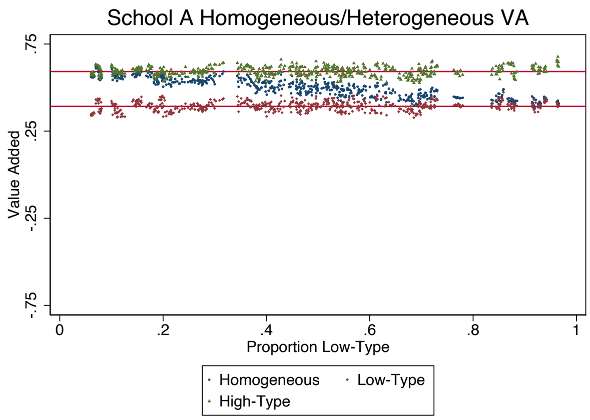
Table B.2 provides specification and forecast bias tests for the postsecondary enrollment value added estimates. For the base sample with base controls, the specification test coefficients range from 1.009 to 1.021 and the forecast bias coefficients range from 0.004 to 0.005. For the subsample of the postsecondary enrollment value added sample that controls for neighborhood characteristics, the proportion of older siblings that attended a 2-year or 4-year university, and peer characteristics, the specification test coefficients range from 1.012 to 1.022 and the forecast bias test coefficients range from -0.001 to -0.004. Thus sorting on unobservables accounts for at most 0.5% of the variance in value added on postsecondary enrollment.

[23] Similar to Chetty, Friedman and Rockoff (2014), I estimate this using the following steps. First I obtain the portion of contemporaneous test scores that projects onto three-grade prior test scores by adding three-grade prior test scores to equation (4). The projection is equal to the predicted value using only the test score from three grades prior. I then regress this projection on school by subgroup value added.

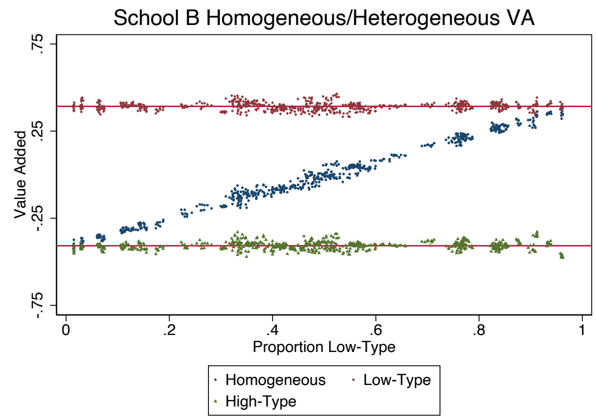
C School Characteristics

Table C.1 shows the correlations between school by subgroup value added and school characteristics. The value added estimates used in this table control for the base controls in equation (4). Columns one, three, and five use test score value added by subgroup as the dependent variable, while columns two, four, and six use postsecondary enrollment value added by subgroup as the dependent variable. Columns one and two give value added by SES, columns three and four give value added by ethnicity, and columns five and six give value added by sex. Each column pertains to a distinct set of value added estimates. Every two rows within each column, which are separated by solid black horizontal lines, are a separate regression.

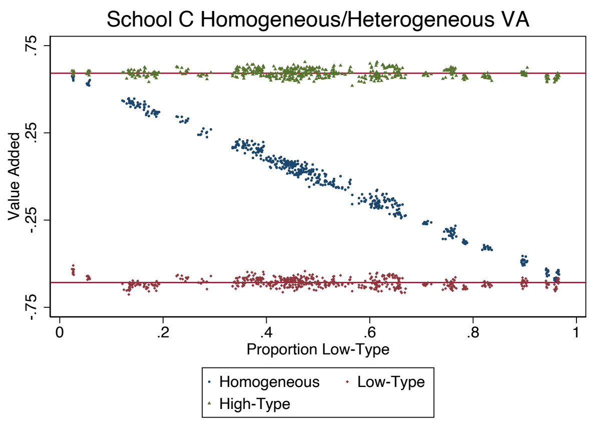
While the coefficient estimates change somewhat relative to table 5, overall the general patterns are relatively similar. These regressions are only correlational and not causal, but the similarities between both sets of value added estimates provides further evidence that the correlations may be worth exploring in future work.



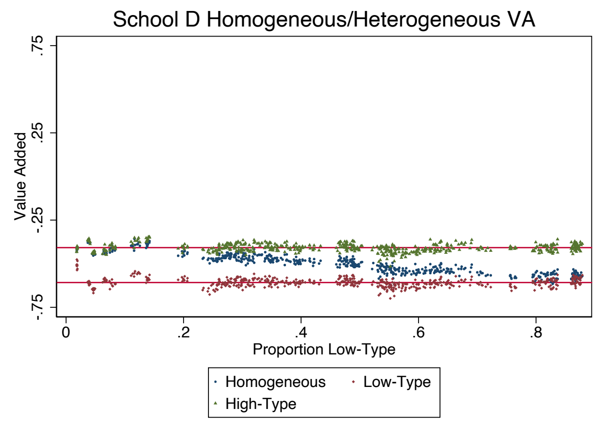
(a) $\mu_{AL} = 0.41$, $\mu_{AH} = 0.61$



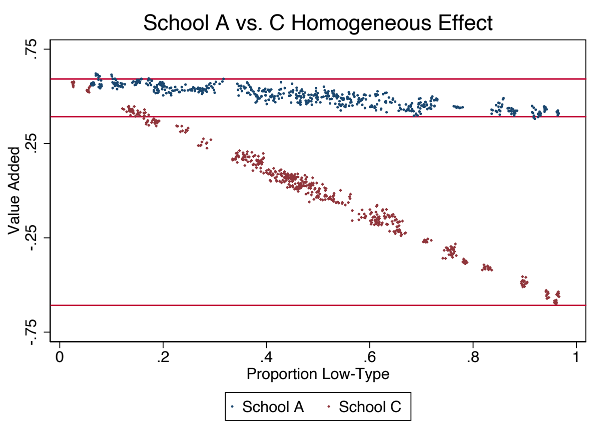
(b) $\mu_{BL} = 0.41$, $\mu_{BH} = -0.39$



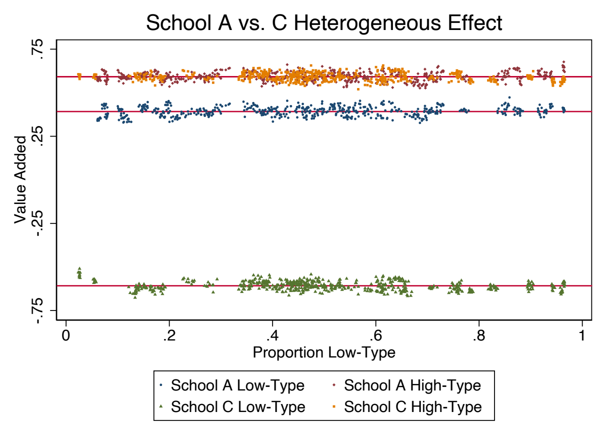
(c) $\mu_{CL} = -0.59$, $\mu_{CH} = 0.61$



(d) $\mu_{DL} = -0.59$, $\mu_{DH} = -0.39$



(e) $\mu_{AL} = 0.41$, $\mu_{CL} = -0.59$, $\mu_{AH} = \mu_{CH} = 0.61$



(f) $\mu_{AL} = 0.41$, $\mu_{CL} = -0.59$, $\mu_{AH} = \mu_{CH} = 0.61$

Figure 1: Value Added Heterogeneity Simulation

Figures 1a, 1b, 1c, and 1d give school (homogeneous) value added and school by subgroup (low-type and high-type) value added estimates for schools of type A, B, C, and D, respectively. Figure 1e gives school value added for schools A and C. Figure 1f gives school by subgroup value added for schools A and C. The horizontal axis plots the proportion of students who are low-type within the school. The vertical axis plots the value added estimate. The horizontal red lines mark the true value added for each school. Each point is a value added estimate from a separate school and year.

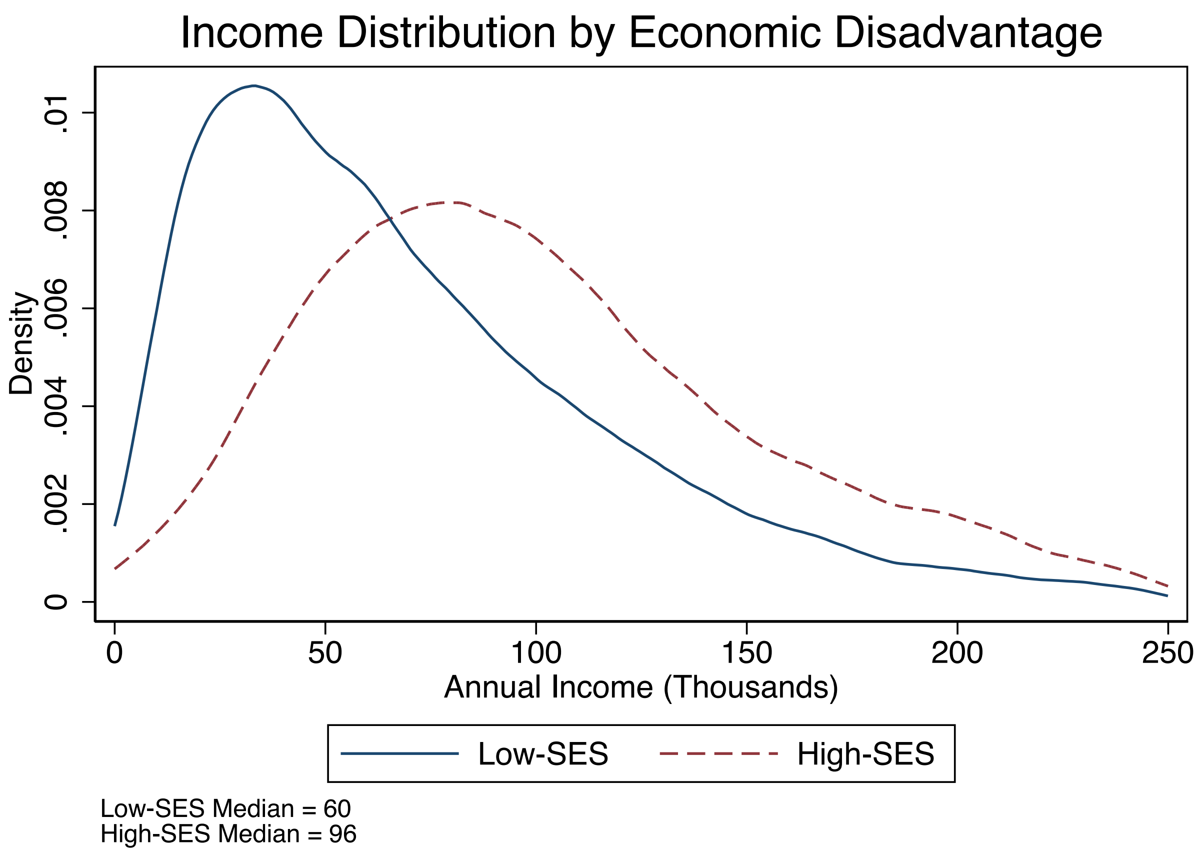
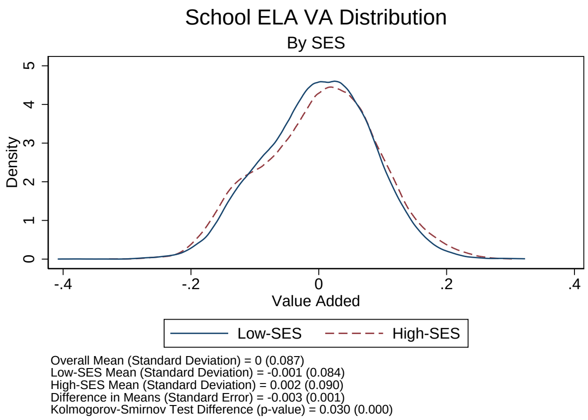
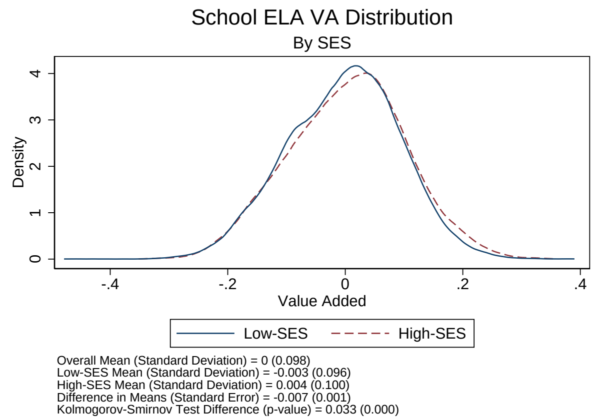


Figure 2: Income Distribution by Economic Disadvantage

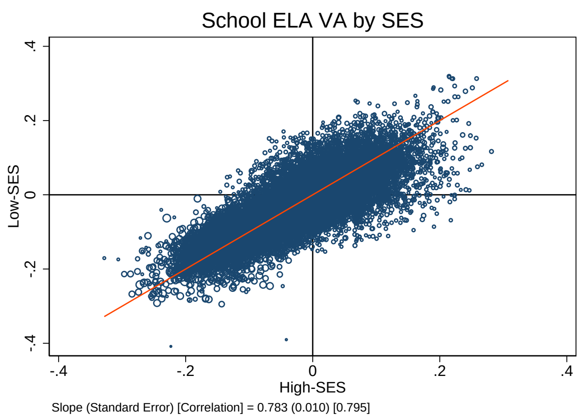
Figure 2 gives the distribution of total household income in 2017 dollars by socioeconomic disadvantage status from the Survey of Income and Program Participation (SIPP). Observations with a total household income of greater than $250,000 are excluded from the figure but not the calculation of the median income.



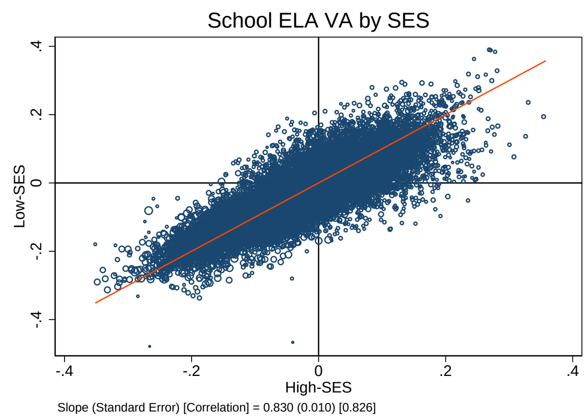
(a) Base Controls



(b) Base + Neighborhood + Sibling + Peer Controls



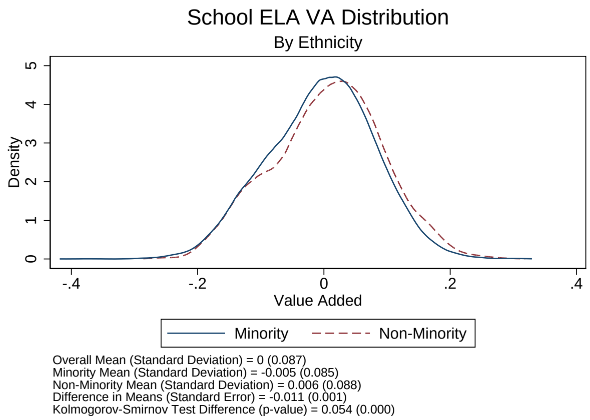
(c) Base Controls



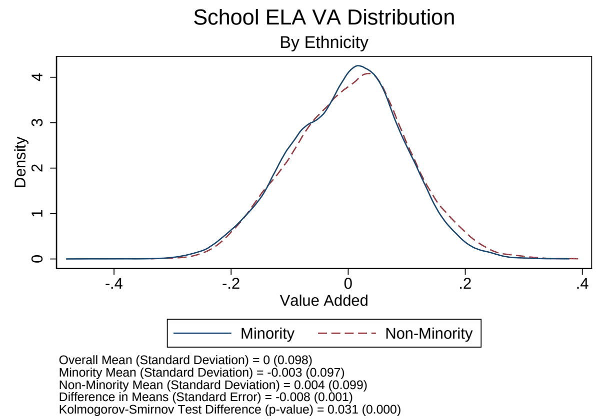
(d) Base + Neighborhood + Sibling + Peer Controls

Figure 3: School Test Score Value Added by SES

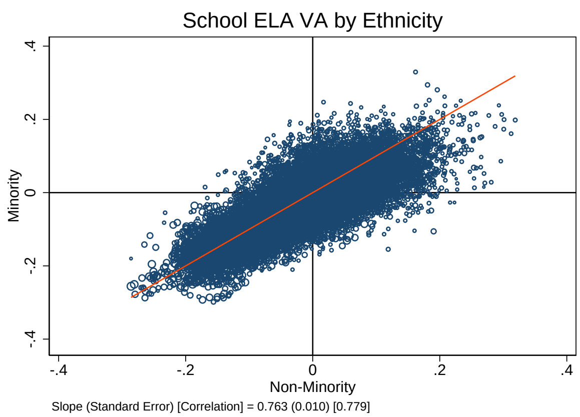
Figures 3a and 3b give the kernel density of school by SES test score value added, $\hat{\mu}_{sdt}$. The horizontal axis gives the school by SES value added estimate and the vertical axis gives the probability density of that estimated value. The two distributions in each figure represent value added for low- and high-SES students. Each figure gives the mean and standard deviation of overall school by SES value added (combining both low- and high-SES value added estimates) as well as school by SES value added for low- and high-SES students individually. Each figure reports the difference in mean school by SES value added between low- and high-SES students as well as the standard error of this difference. Each figure reports the Kolmogorov-Smirnov test statistic for the equality of probability distributions between school by SES value added for low- and high-SES students. The p-value for the Kolmogorov-Smirnov test is estimated under the null hypothesis that school by SES value added comes from the same distribution for both low- and high-SES students. Figures 3c and 3d give scatter plots of school by SES test score value added, $\hat{\mu}_{sdt}$, for each school. Each point represents a single school-year observation, with school value added for high-SES students on the horizontal axis and school value added for low-SES students on the vertical axis. The red diagonal line is the y = x line which represents a school providing equal value added to low- and high-SES students. The slope, standard error of the slope, and correlation between value added for low- and high-SES students is presented at the bottom of each figure. Figures 3a and 3c give school by SES value added for the base value added sample using the base value added controls. Figures 3b and 3d give school by SES value added for the subsample of students that have at least one older sibling and can be matched to the American Community Survey and controls for ACS Census tract characteristics, average older-sibling test scores, and peer (jackknife) averages of the independent variables used in the estimation of value added.



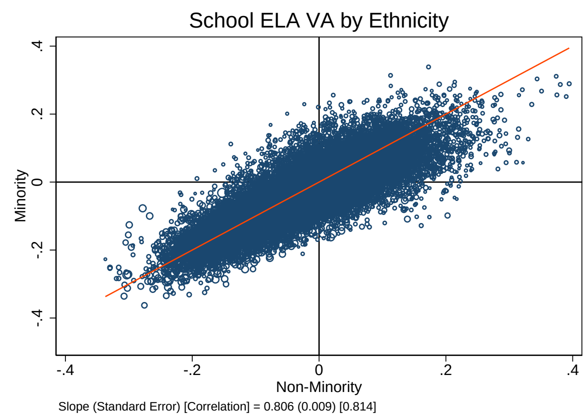
(a) Base Controls



(b) Base + Neighborhood + Sibling + Peer Controls



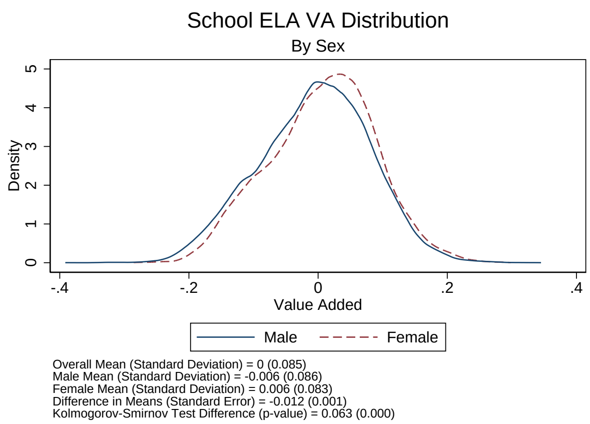
(c) Base Controls



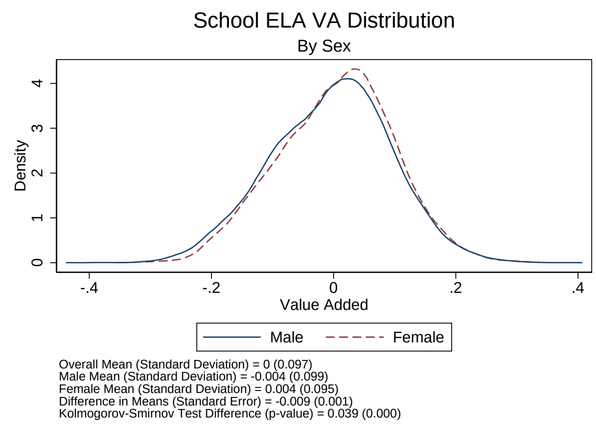
(d) Base + Neighborhood + Sibling + Peer Controls

Figure 4: School Test Score Value Added by Ethnicity

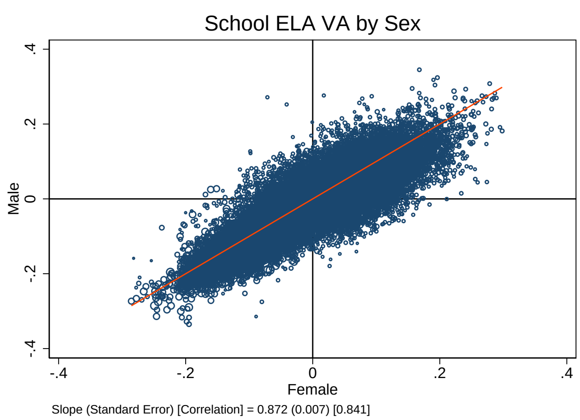
Figures 4a and 4b give the kernel density of school by ethnicity test score value added, $\hat{\mu}_{sdt}$. The horizontal axis gives the school by ethnicity value added estimate and the vertical axis gives the probability density of that estimated value. The two distributions in each figure represent value added for minority and non-minority students. Each figure gives the mean and standard deviation of overall school by ethnicity value added (combining both minority and non-minority value added estimates) as well as school by ethnicity value added for minority and non-minority students individually. Each figure reports the difference in mean school by ethnicity value added between minority and non-minority students as well as the standard error of this difference. Each figure reports the Kolmogorov-Smirnov test statistic for the equality of probability distributions between school by ethnicity value added for minority and non-minority students. The p-value for the Kolmogorov-Smirnov test is estimated under the null hypothesis that school by ethnicity value added comes from the same distribution for both minority and non-minority students. Figures 4c and 4d give scatter plots of school by ethnicity test score value added, $\hat{\mu}_{sdt}$, for each school. Each point represents a single school-year observation, with school value added for minority students on the horizontal axis and school value added for non-minority students on the vertical axis. The red diagonal line is the y = x line which represents a school providing equal value added to minority and non-minority students. The slope, standard error of the slope, and correlation between value added for minority and non-minority students is presented at the bottom of each figure. Figures 4a and 4c give school by ethnicity value added for the base value added sample using the base value added controls. Figures 4b and 4d give school by ethnicity value added for the subsample of students that have at least one older sibling and can be matched to the American Community Survey and controls for ACS Census tract characteristics, average older-sibling test scores, and peer (jackknife) averages of the independent variables used in the estimation of value added.



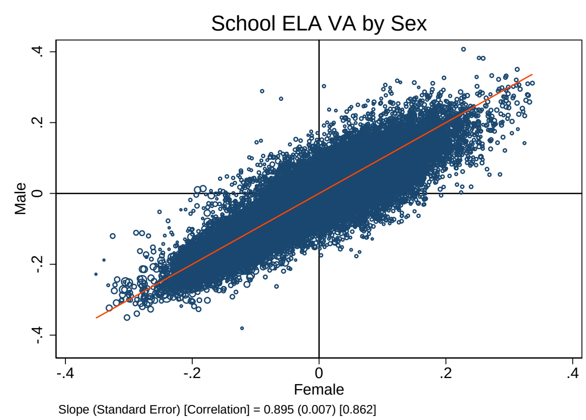
(a) Base Controls



(b) Base + Neighborhood + Sibling + Peer Controls



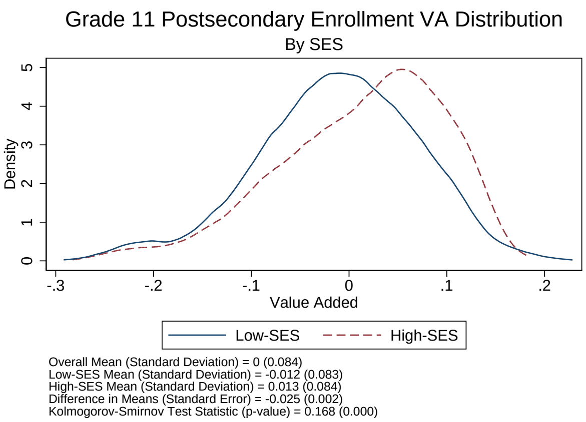
(c) Base Controls



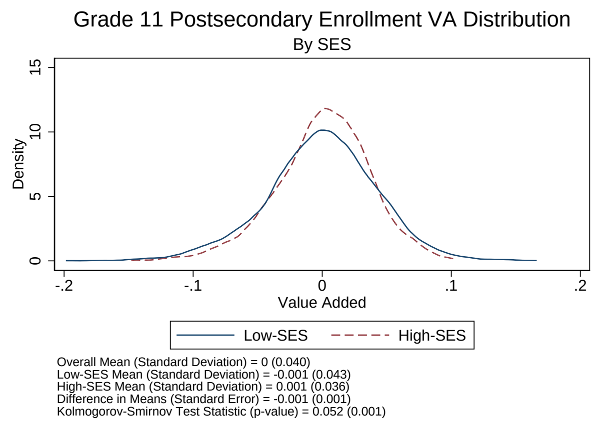
(d) Base + Neighborhood + Sibling + Peer Controls

Figure 5: School Test Score Value Added by Sex

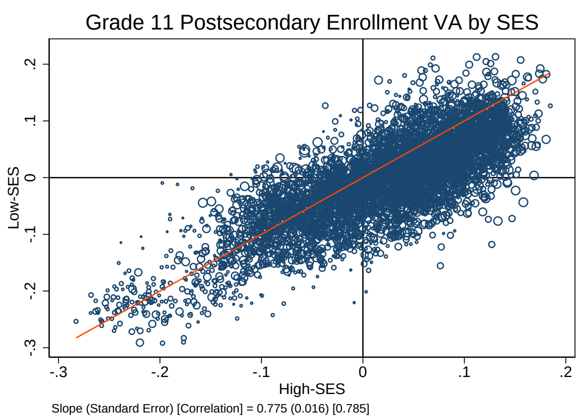
Figures 5a and 5b give the kernel density of school by sex test score value added, $\hat{\mu}_{sdt}$. The horizontal axis gives the school by sex value added estimate and the vertical axis gives the probability density of that estimated value. The two distributions in each figure represent value added for male and female students. Each figure gives the mean and standard deviation of overall school by sex value added (combining both male and female value added estimates) as well as school by sex value added for male and female students individually. Each figure reports the difference in mean school by sex value added between male and female students as well as the standard error of this difference. Each figure reports the Kolmogorov-Smirnov test statistic for the equality of probability distributions between school by sex value added for male and female students. The p-value for the Kolmogorov-Smirnov test is estimated under the null hypothesis that school by sex value added comes from the same distribution for both male and female students. Figures 5c and 5d give scatter plots of school by sex test score value added, $\hat{\mu}_{sdt}$, for each school. Each point represents a single school-year observation, with school value added for female students on the horizontal axis and school value added for male students on the vertical axis. The red diagonal line is the y = x line which represents a school providing equal value added to male and male students. The slope, standard error of the slope, and correlation between value added for male and female students is presented at the bottom of each figure. Figures 5a and 5c give school by sex value added for the base value added sample using the base value added controls. Figures 5b and 5d give school by sex value added for the subsample of students that have at least one older sibling and can be matched to the American Community Survey and controls for ACS Census tract characteristics, average older-sibling test scores, and peer (jackknife) averages of the independent variables used in the estimation of value added.



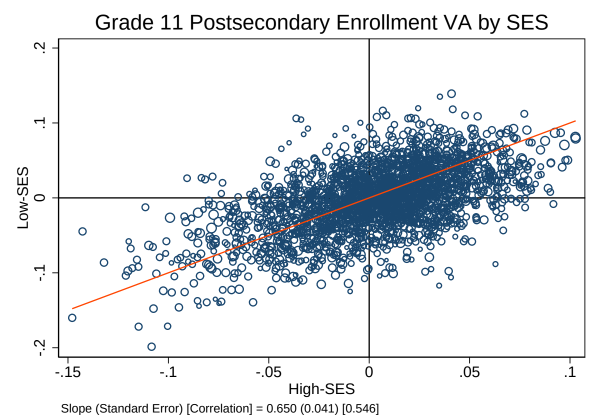
(a) Base Controls



(b) Base + Neighborhood + Sibling + Peer Controls



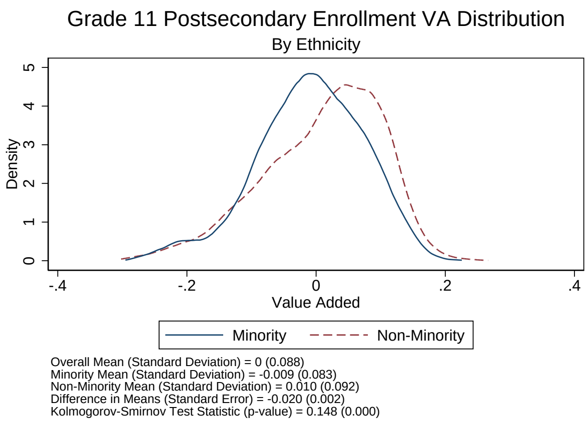
(c) Base Controls



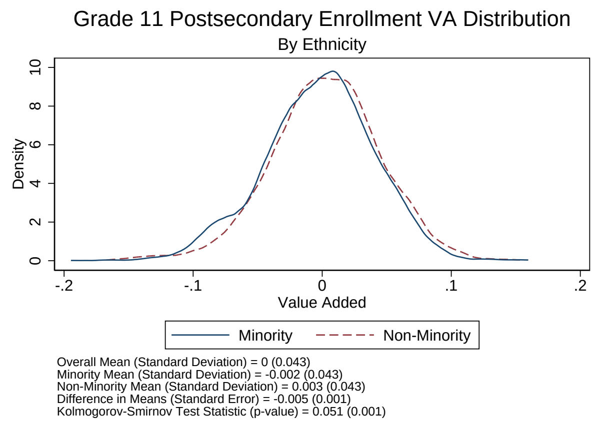
(d) Base + Neighborhood + Sibling + Peer Controls

Figure 6: School Postsecondary Enrollment Value Added by SES

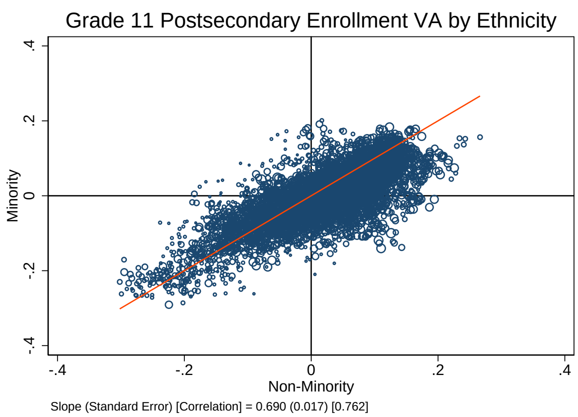
Figures 6a and 6b give the kernel density of school by SES postsecondary enrollment value added, $\hat{\mu}_{sdt}$. The horizontal axis gives the school by SES value added estimate and the vertical axis gives the probability density of that estimated value. The two distributions in each figure represent value added for low- and high-SES students. Each figure gives the mean and standard deviation of overall school by SES value added (combining both low- and high-SES value added estimates) as well as school by SES value added for low- and high-SES students individually. Each figure reports the difference in mean school by SES value added between low- and high-SES students as well as the standard error of this difference. Each figure reports the Kolmogorov-Smirnov test statistic for the equality of probability distributions between school by SES value added for low- and high-SES students. The p-value for the Kolmogorov-Smirnov test is estimated under the null hypothesis that school by SES value added comes from the same distribution for both low- and high-SES students. Figures 6c and 6d give scatter plots of school by SES postsecondary enrollment value added, $\hat{\mu}_{sdt}$, for each school. Each point represents a single school-year observation, with school value added for high-SES students on the horizontal axis and school value added for low-SES students on the vertical axis. The red diagonal line is the y = x line which represents a school providing equal value added to low- and high-SES students. The slope, standard error of the slope, and correlation between value added for low- and high-SES students is presented at the bottom of each figure. Figures 6a and 6c give school by SES value added for the base value added sample using the base value added controls. Figures 6b and 6d give school by SES value added for the subsample of students that have at least one older sibling and can be matched to the American Community Survey and controls for ACS Census tract characteristics, the proportion of older siblings that attended a 2-year or 4-year university, and peer (jackknife) averages of the independent variables used in the estimation of value added. Because each student's enrollment outcome is invariant across grades, the school by SES value added estimates are calculated using only observations from the 11th grade.



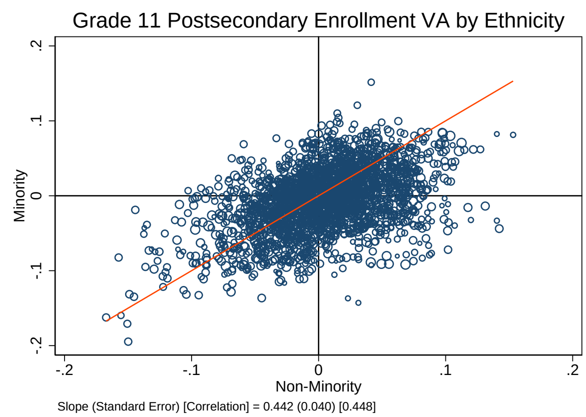
(a) Base Controls



(b) Base + Neighborhood + Sibling + Peer Controls



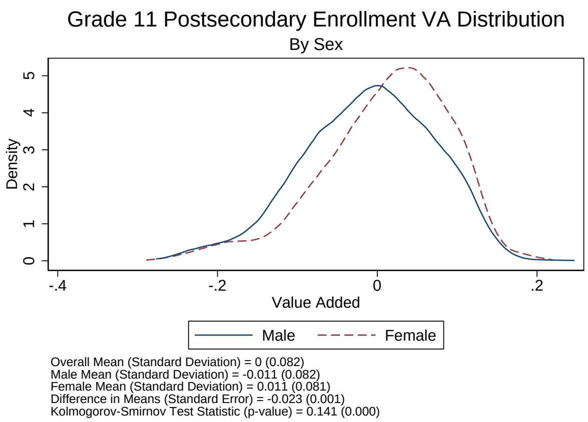
(c) Base Controls



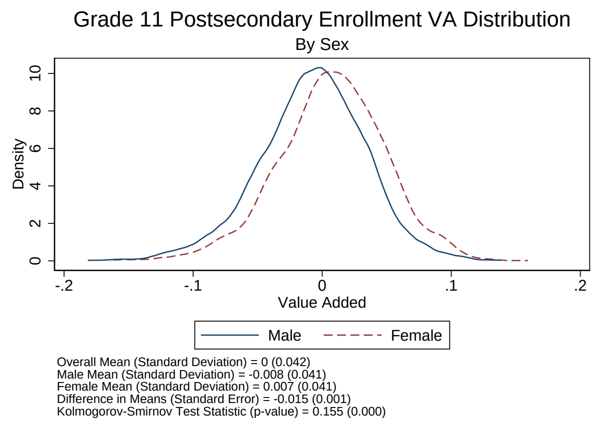
(d) Base + Neighborhood + Sibling + Peer Controls

Figure 7: School Postsecondary Enrollment Value Added by Ethnicity

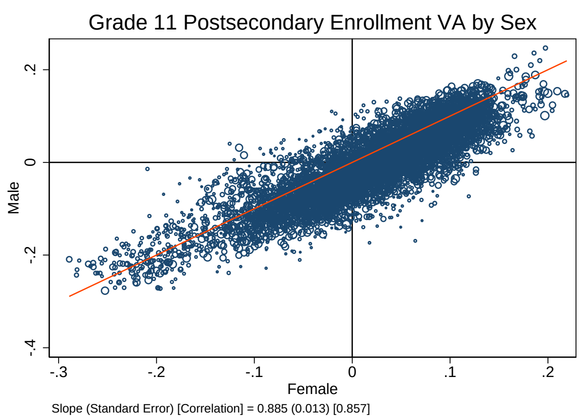
Figures 7a and 7b give the kernel density of school by ethnicity postsecondary enrollment value added, $\hat{\mu}_{sdt}$. The horizontal axis gives the school by ethnicity value added estimate and the vertical axis gives the probability density of that estimated value. The two distributions in each figure represent value added for minority and non-minority students. Each figure gives the mean and standard deviation of overall school by ethnicity value added (combining both minority and non-minority value added estimates) as well as school by ethnicity value added for minority and non-minority students individually. Each figure reports the difference in mean school by ethnicity value added between minority and non-minority students as well as the standard error of this difference. Each figure reports the Kolmogorov-Smirnov test statistic for the equality of probability distributions between school by ethnicity value added for minority and non-minority students. The p-value for the Kolmogorov-Smirnov test is estimated under the null hypothesis that school by ethnicity value added comes from the same distribution for both minority and non-minority students. Figures 7c and 7d give scatter plots of school by ethnicity postsecondary enrollment value added, $\hat{\mu}_{sdt}$, for each school. Each point represents a single school-year observation, with school value added for minority students on the horizontal axis and school value added for non-minority students on the vertical axis. The red diagonal line is the y = x line which represents a school providing equal value added to minority and non-minority students. The slope, standard error of the slope, and correlation between value added for minority and non-minority students is presented at the bottom of each figure. Figures 7a and 7c give school by ethnicity value added for the base value added sample using the base value added controls. Figures 7b and 7d give school by ethnicity value added for the subsample of students that have at least one older sibling and can be matched to the American Community Survey and controls for ACS Census tract characteristics, the proportion of older siblings that attended a 2-year or 4-year university, and peer (jackknife) averages of the independent variables used in the estimation of value added. Because each student's enrollment outcome is invariant across grades, the school by ethnicity value added estimates are calculated using only observations from the 11th grade.



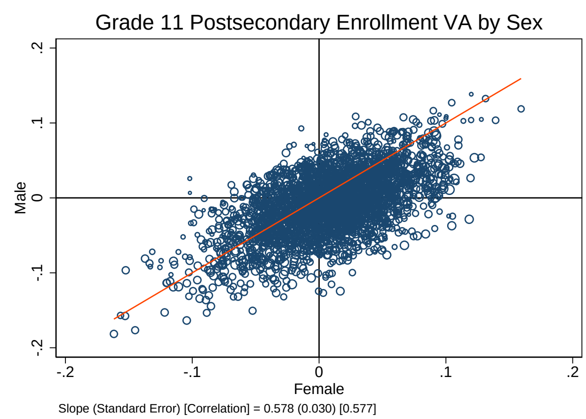
(a) Base Controls



(b) Base + Neighborhood + Sibling + Peer Controls



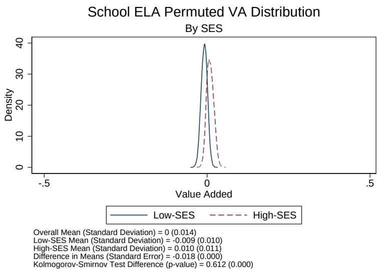
(c) Base Controls



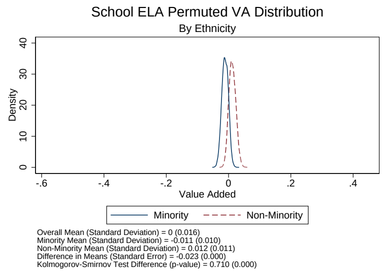
(d) Base + Neighborhood + Sibling + Peer Controls

Figure 8: School Postsecondary Enrollment Value Added by Sex

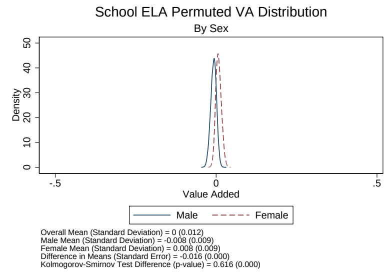
Figures 8a and 8b give the kernel density of school by sex postsecondary enrollment value added, $\hat{\mu}_{sdt}$. The horizontal axis gives the school by sex value added estimate and the vertical axis gives the probability density of that estimated value. The two distributions in each figure represent value added for male and female students. Each figure gives the mean and standard deviation of overall school by sex value added (combining both male and female value added estimates) as well as school by sex value added for male and female students individually. Each figure reports the difference in mean school by sex value added between male and female students as well as the standard error of this difference. Each figure reports the Kolmogorov-Smirnov test statistic for the equality of probability distributions between school by sex value added for male and female students. The p-value for the Kolmogorov-Smirnov test is estimated under the null hypothesis that school by sex value added comes from the same distribution for both male and female students. Figures 8c and 8d give scatter plots of school by sex postsecondary enrollment value added, $\hat{\mu}_{sdt}$, for each school. Each point represents a single school-year observation, with school value added for female students on the horizontal axis and school value added for male students on the vertical axis. The red diagonal line is the y = x line which represents a school providing equal value added to male and male students. The slope, standard error of the slope, and correlation between value added for male and female students is presented at the bottom of each figure. Figures 8a and 8c give school by sex value added for the base value added sample using the base value added controls. Figures 8b and 8d give school by sex value added for the subsample of students that have at least one older sibling and can be matched to the American Community Survey and controls for ACS Census tract characteristics, the proportion of older siblings that attended a 2-year or 4-year university, and peer (jackknife) averages of the independent variables used in the estimation of value added. Because each student's enrollment outcome is invariant across grades, the school by sex value added estimates are calculated using only observations from the 11th grade.



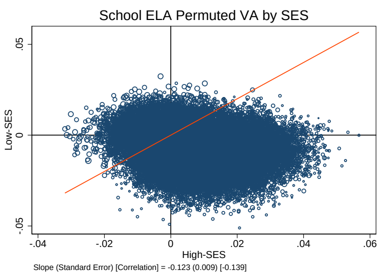
(a) SES



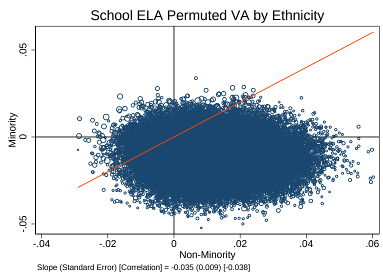
(b) Ethnicity



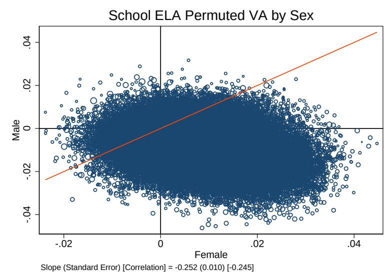
(c) Sex



(d) SES



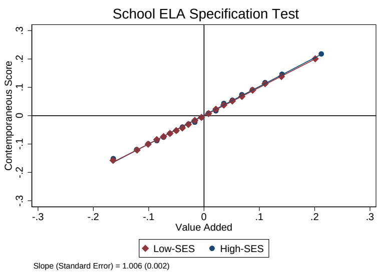
(e) Ethnicity



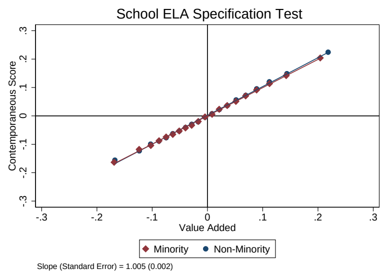
(f) Sex

Figure B.1: School Test Score Permuted Value Added by Subgroup

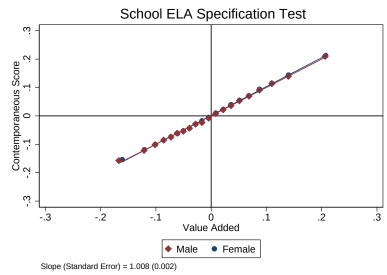
Figures B.1a, B.1b, and B.1c give the kernel density of school by subgroup test score permuted value added, $\hat{\mu}_{sdt}$. The horizontal axis gives the school by subgroup permuted value added estimate and the vertical axis gives the probability density of that estimated value. The two distributions in each figure represent the two subgroups in each analysis. Each figure gives the mean and standard deviation of overall school by subgroup permuted value added (combining both subgroups) as well as school by subgroup permuted value added for each subgroup individually. Each figure reports the difference in mean school by subgroup permuted value added between the two subgroups as well as the standard error of this difference. Each figure reports the Kolmogorov-Smirnov test statistic for the equality of probability distributions between school by subgroup permuted value added for each subgroup. The p-value for the Kolmogorov-Smirnov test is estimated under the null hypothesis that school by subgroup permuted value added comes from the same distribution for both subgroups. Figures B.1d, B.1e, and B.1f give scatter plots of school by subgroup test score permuted value added, $\hat{\mu}_{sdt}$, for each school. Each point represents a single school-year observation, with the horizontal and vertical axes representing permuted value added for the two subgroups in each analysis. The red diagonal line is the y = x line which represents a school providing equal permuted value added to each subgroup. The slope, standard error of the slope, and correlation between permuted value added for the two subgroups is presented at the bottom of each figure. All value added estimates are calculated using the base value added sample and base value added controls. Figures B.1a and B.1d give school by SES permuted value added. Figures B.1b and B.1e give school by ethnicity permuted value added. Figures B.1c and B.1f give school by sex permuted value added. Permuted value added is estimated by first permuting the school assignment vector within a grade by year cell before applying the value added methodology outlined in section 4.2.



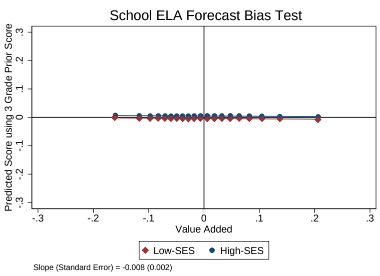
(a) SES



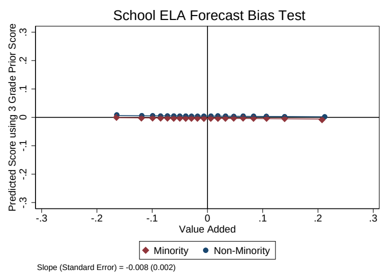
(b) Ethnicity



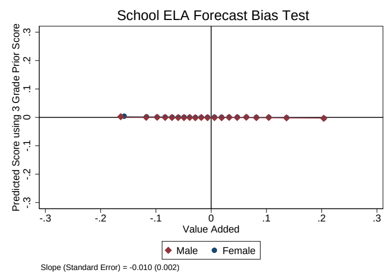
(c) Sex



(d) SES



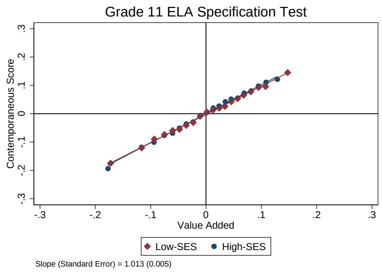
(e) Ethnicity



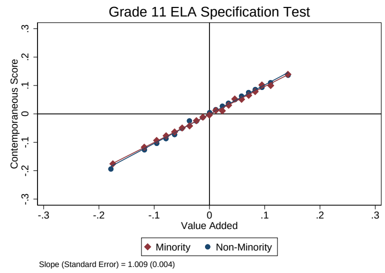
(f) Sex

Figure B.2: School Test Score Value Added Specification and Forecast Bias Tests

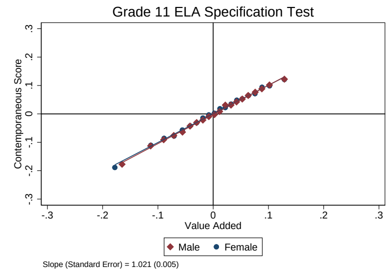
Figures B.2a, B.2b, and B.2c give binned scatter plots of test score residuals r_{isgdt} and school by subgroup value added $\hat{\mu}_{sdt}$. Each point represents the average test score residual for a given school by subgroup value added estimate, with the horizontal axis giving the school by subgroup value added vingtile and the vertical axis giving the average test score residual in that vingtile. Average test score residuals conditional on school by subgroup value added vingtile are calculated at the student level. Each figure reports the coefficient estimate and standard error from the bivariate regression of test score residuals r_{isgdt} on school by subgroup value added $\hat{\mu}_{sdt}$ run at the student level (equivalent to the coefficient estimate in table B.1). Figures B.2d, B.2e, and B.2f give binned scatter plots of the projection of contemporaneous test scores onto three grade prior test scores and school by subgroup value added $\hat{\mu}_{sdt}$. Each point represents the average projection of contemporaneous test scores onto three grade prior test scores for a given school by subgroup value added estimate, with the horizontal axis giving the school by subgroup value added vingtile and the vertical axis giving the average projection of contemporaneous test scores onto three grade prior test scores in that vingtile. Projections of test scores onto three grade prior test scores include all of the controls included in the estimation of school by subgroup value added. Average projections of test scores onto three grade prior test scores conditional on school by subgroup value added vingtile are calculated at the student level. Each figure reports the coefficient estimate and standard error from the regression of the projection of contemporaneous test scores onto three grade prior test scores on school by subgroup value added $\hat{\mu}_{sdt}$ run at the student level (equivalent to the coefficient estimate in table B.1). All value added estimates are calculated using the base value added sample and base value added controls. Figures B.2a and B.2d give school by SES value added. Figures B.2b and B.2e give school by ethnicity value added. Figures B.2c and B.2f give school by sex value added.



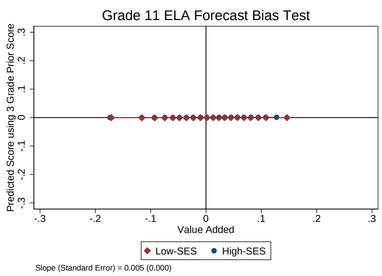
(a) SES



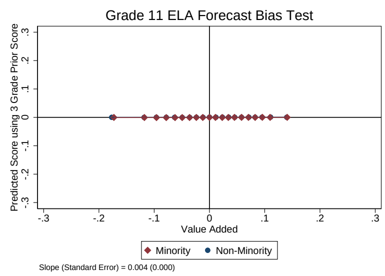
(b) Ethnicity



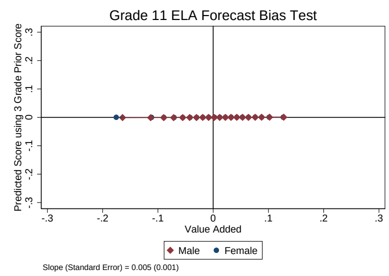
(c) Sex



(d) SES



(e) Ethnicity



(f) Sex

Figure B.3: School Postsecondary Enrollment Value Added Specification and Forecast Bias Tests

Figures B.3a, B.3b, and B.3c give binned scatter plots of postsecondary enrollment residuals r_{isgdt} and school by subgroup value added $\hat{\mu}_{sdt}$ in grade 11. Each point represents the average postsecondary enrollment residual for a given school by subgroup value added estimate, with the horizontal axis giving the school by subgroup value added vingtile and the vertical axis giving the average postsecondary enrollment residual in that vingtile. Average postsecondary enrollment residuals conditional on school by subgroup value added vingtile are calculated at the student level. Each figure reports the coefficient estimate and standard error from the bivariate regression of postsecondary enrollment residuals r_{isgdt} on school by subgroup value added $\hat{\mu}_{sdt}$ run at the student level (equivalent to the coefficient estimate in table B.2). Figures B.3d, B.3e, and B.3f give binned scatter plots of the projection of postsecondary enrollment onto three grade prior test scores and school by subgroup value added $\hat{\mu}_{sdt}$. Each point represents the average projection of postsecondary enrollment onto three grade prior test scores for a given school by subgroup value added estimate, with the horizontal axis giving the school by subgroup value added vingtile and the vertical axis giving the average projection of postsecondary enrollment onto three grade prior test scores in that vingtile. Projections of postsecondary enrollment onto three grade prior test scores include all of the controls included in the estimation of school by subgroup value added. Average projections of postsecondary enrollment onto three grade prior test scores conditional on school by subgroup value added vingtile are calculated at the student level. Each figure reports the coefficient estimate and standard error from the regression of the projection of postsecondary enrollment onto three grade prior test scores on school by subgroup value added $\hat{\mu}_{sdt}$ run at the student level (equivalent to the coefficient estimate in table B.1). All value added estimates are calculated using the base value added sample and base value added controls. Figures B.3a and B.3d give school by SES value added. Figures B.3b and B.3e give school by ethnicity value added. Figures B.3c and B.3f give school by sex value added.

s	μ_{sLt}	μ_{sHt}	$p(s = S)$
A	0.41	0.61	0.25
B	0.41	-0.39	0.25
C	-0.59	0.61	0.25
D	-0.59	-0.39	0.25

Table 1 gives the true school by subgroup value added of four types of schools: A, B, C, and D. Column one gives the type of school. Column two gives the value added the school provides to their low-type students. Column three gives the value added the school provides to their high-type students. Column four gives the proportion of total schools that are that type of school. School A and school D are unambiguously the best and worst schools, respectively. School B is particularly effective with low-type students while school C is effective with high-type students.

Table 1: Rank Example Value Added

$\theta_{st} \sim \mathcal{N}(0, 0.05)$
$\varepsilon_{isdt} \sim \mathcal{N}(0, 0.5)$
$N_{sLt}, N_{sHt} \sim U[7, 1007]$
$j_{sLt}, j_{sHt} \sim U[1, 1.1]$
of Schools = 250
of Years = 9

Table 2 gives parameters and their distributions for the simulation of school by subgroup value added. θ_{st} are school by year common shocks. ε_{isdt} is a noise term. N_{sLt} and N_{sHt} are the number of students of low- and high-type, respectively. j_{sLt} and j_{sHt} are inflation terms that, when multiplied by N_{sLt} and N_{sHt}, allow a school's population to deviate year to year from a baseline average.

Table 2: Rank Example Parameters

	Low-SES	High-SES	Minority	Non-Minority	Male	Female
Demographic Controls						
Age in Years	12.97	13.35	13.05	13.27	13.17	13.11
	[2.294]	[2.327]	[2.303]	[2.330]	[2.320]	[2.315]
Male	0.494	0.501	0.491	0.505	1	0
	[0.500]	[0.500]	[0.500]	[0.500]	[0]	[0]
Hispanic or Latino	0.718	0.236	0.860	0	0.496	0.504
	[0.450]	[0.425]	[0.347]	[0]	[0.500]	[0.500]
White	0.110	0.509	0	0.694	0.294	0.287
	[0.313]	[0.500]	[0]	[0.461]	[0.455]	[0.452]
Asian	0.0865	0.179	0	0.306	0.131	0.125
	[0.281]	[0.383]	[0]	[0.461]	[0.338]	[0.331]
Black or African American	0.0714	0.0514	0.107	0	0.0603	0.0644
	[0.257]	[0.221]	[0.309]	[0]	[0.238]	[0.245]
Other Race	0.0140	0.0253	0.0328	0	0.0190	0.0192
	[0.117]	[0.157]	[0.178]	[0]	[0.136]	[0.137]
Economic Disadvantage	1	0	0.757	0.257	0.544	0.551
	[0]	[0]	[0.429]	[0.437]	[0.498]	[0.497]
Limited English Proficient Status	0.276	0.0414	0.256	0.0498	0.182	0.157
	[0.447]	[0.199]	[0.436]	[0.217]	[0.386]	[0.364]
Disabled	0.0445	0.0412	0.0441	0.0415	0.0572	0.0289
	[0.206]	[0.199]	[0.205]	[0.199]	[0.232]	[0.168]
Test Scores						
Current Test Score	-0.245	0.495	-0.218	0.516	0.000199	0.177
	[0.868]	[0.944]	[0.875]	[0.947]	[0.990]	[0.953]
1 Grade Prior Test Score	-0.237	0.520	-0.206	0.538	0.0218	0.188
	[0.859]	[0.934]	[0.867]	[0.938]	[0.980]	[0.952]
2 Grade Prior Test Score	-0.233	0.531	-0.199	0.545	0.0330	0.190
	[0.858]	[0.932]	[0.867]	[0.939]	[0.976]	[0.956]
Observations	12,204,773	10,067,455	12,957,621	9,310,725	11,073,636	11,213,388

Values are means and standard deviations [in brackets] of the dependent and independent variables used in the test score value added estimation. Only students included in the test score value added sample are included in this table. Columns two and three are the summary statistics for low- and high-SES students, respectively. Columns four and five are the summary statistics for minority and non-minority students, respectively. Columns six and seven are the summary statistics for male and female students, respectively. Data comes from grades 4-11 in public schools in the state of California between the 2004-2005 and 2012-2013 school years.

Table 3: K–12 Summary Statistics

	Low-SES	High-SES	Minority	Non-Minority	Male	Female
Enrolled at a Postsecondary Institution	.517	.692	.527	.7	.57	.641
	[.5]	[.462]	[.499]	[.458]	[.495]	[.48]
Enrolled at a 2-Year College	.326	.335	.335	.326	.335	.327
	[.469]	[.472]	[.472]	[.469]	[.472]	[.469]
Enrolled at a 4-Year University	.191	.357	.191	.375	.235	.314
	[.393]	[.479]	[.393]	[.484]	[.424]	[.464]
Enrolled at a Public Institution	.482	.597	.483	.608	.518	.561
	[.5]	[.491]	[.5]	[.488]	[.5]	[.496]
Enrolled at a Private Institution	.0352	.0952	.0435	.0919	.0515	.0791
	[.184]	[.293]	[.204]	[.289]	[.221]	[.27]
Enrolled at a CA Institution	.493	.592	.493	.603	.513	.572
	[.5]	[.492]	[.5]	[.489]	[.5]	[.495]
Enrolled at an Out-of-State Institution	.0238	.1	.0337	.0971	.0561	.0687
	[.153]	[.301]	[.18]	[.296]	[.23]	[.253]
Observations	879,399	896,272	969,990	805,459	878,589	898,537

Values are means and standard deviations [in brackets]. Only students included in the postsecondary value added sample are included in this table. Columns two and three are the summary statistics for low- and high-SES students, respectively. Columns four and five are the summary statistics for minority and non-minority students, respectively. Columns six and seven are the summary statistics for male and female students, respectively. Data comes from grade 11 in public schools in the state of California between the 2008-2009 and 2012-2013 school years.

Table 4: Postsecondary Summary Statistics

	SES		Ethnicity		Sex	
	Test Score	Enrollment	Test Score	Enrollment	Test Score	Enrollment
FTE Teachers per Student	42.2***	-14.6**	31.5***	1.68	22.9***	.0379
	(3.34)	(7.19)	(3.43)	(8.72)	(2.57)	(6.33)
Low-SES/Minority/Male × FTE Teachers per Student	-26.6***	32.2***	-11.7***	6.83	-.56	10.2*
	(3.49)	(8.61)	(3.54)	(10.1)	(.93)	(5.27)
FTE Pupil Services per Student	-115***	-80.3***	-112***	-48.2*	-136***	21.5
	(12.3)	(25.8)	(11.5)	(29.2)	(8.27)	(22.1)
Low-SES/Minority/Male × FTE Pupil Services per Student	-31.7**	166***	-40.6***	116***	3.69	-11.9
	(13.7)	(31.7)	(12.9)	(35.2)	(2.99)	(21.8)
English Learner Staff Per Student	21.3***	4.68	20.1***	7.2**	22.4***	2.12
	(1.46)	(3.35)	(1.44)	(3.66)	(.98)	(3.28)
Low-SES/Minority/Male × English Learner Staff Per Student	6.38***	-9.43**	6.92***	-14.6***	-1.15***	-6.13**
	(1.55)	(4.58)	(1.46)	(4.84)	(.393)	(2.76)
Proportion $\leq$ 3 Years Experience Teachers	-1.65***	.925**	-1.73***	1.13**	-1.83***	1.11***
	(.167)	(.417)	(.17)	(.459)	(.121)	(.357)
Low-SES/Minority/Male × Proportion $\leq$ 3 Years Experience Teachers	-.256	.592	-.129	.187	-.0597	.304
	(.166)	(.502)	(.169)	(.539)	(.0457)	(.3)
Proportion Full Credential Teachers	3.72***	-.934	3.83***	-.86	3.28***	-1.2
	(.311)	(.713)	(.329)	(.746)	(.176)	(.757)
Low-SES/Minority/Male × Proportion Full Credential Teachers	-.53*	-.406	-.681**	-.527	.0711	.564
	(.308)	(.956)	(.325)	(1.01)	(.0749)	(.659)
Proportion Male Teachers	-4.52***	2.35***	-4.43***	2.95***	-4.21***	2.67***
	(.0952)	(.528)	(.0988)	(.608)	(.0766)	(.586)
Low-SES/Minority/Male × Proportion Male Teachers	.559***	-.453	.403***	-1.44*	.00306	-1.63***
	(.103)	(.873)	(.103)	(.864)	(.0376)	(.467)
Proportion Minority Teachers	-1.51***	.551*	-1.63***	.905**	-.317***	1.43***
	(.184)	(.322)	(.193)	(.376)	(.102)	(.248)
Low-SES/Minority/Male × Proportion Minority Teachers	1.75***	.765**	1.8***	.455	-.0455	-1.02***
	(.204)	(.375)	(.215)	(.429)	(.0368)	(.25)
Instruction Expenditures ($1,000s) per Student	.00878	-.0233**	.01	-.0145	-.00115	.00544
	(.00666)	(.00989)	(.00638)	(.0111)	(.00316)	(.00552)
Low-SES/Minority/Male × Instruction Expenditures ($1,000s) per Student	-.0111*	.0376***	-.0129**	.0266**	.00116	.00318
	(.00605)	(.00973)	(.0056)	(.0103)	(.00102)	(.00387)
Pupil Services Expenditures ($1,000s) per Student	-.191***	-.113*	-.165***	-.0107	-.0784***	.0547
	(.0493)	(.0641)	(.0479)	(.085)	(.0217)	(.0351)
Low-SES/Minority/Male × Pupil Services Expenditures ($1,000s) per Student	.164***	.184***	.129***	.059	-.0068	-.0497
	(.0461)	(.0641)	(.0444)	(.0839)	(.00694)	(.0304)
Ancillary Services Expenditures ($1,000s) per Student	-.889***	.1	-.816***	.00422	-.91***	.834
	(.31)	(.495)	(.301)	(.519)	(.228)	(.525)
Low-SES/Minority/Male × Ancillary Services Expenditures ($1,000s) per Student	.152	2.45***	-.0258	2.37***	.208**	.731**
	(.327)	(.815)	(.291)	(.778)	(.0834)	(.352)
Other Expenditures ($1,000s) per Student	-.00892	-.00757	-.00722	-.00015	-.0228***	.00436
	(.00597)	(.00827)	(.00564)	(.00914)	(.00454)	(.00839)
Low-SES/Minority/Male × Other Expenditures ($1,000s) per Student	-.0177***	.0422***	-.0208***	.0267**	.0055***	.0197***
	(.00649)	(.013)	(.00599)	(.0121)	(.0016)	(.00741)
General Administration Expenditures ($1,000s) per Student	-.0106	-.0343*	-.00412	-.0298	-.00418	-.000344
	(.0135)	(.0185)	(.0134)	(.0198)	(.0111)	(.019)
Low-SES/Minority/Male × General Administration Expenditures ($1,000s) per Student	.00885	.0564***	.0012	.0507***	.000215	.00849
	(.0124)	(.0216)	(.0108)	(.0195)	(.00351)	(.0134)

Every two rows within each column represent a separate trivariate regression of school by subgroup value added on a school characteristic, a subgroup fixed effect, and the school characteristic interacted with the subgroup fixed effect. Standard errors clustered at the school level are presented in parentheses. In columns two and three the subgroup is SES, and the indicator variable on "low-SES/minority/male" is low-SES. In columns four and five the subgroup is ethnicity, and the indicator variable on "low-SES/minority/male" is minority. In columns six and seven the subgroup is sex, and the indicator variable on "low-SES/minority/male" is male. In columns two, four, and six the dependent variable is school by subgroup value added on ELA test scores. In columns three, five, and seven the dependent variable is school by subgroup value added on postsecondary enrollment. Data comes from grades 4-11 in public schools in the state of California between the 2004-2005 and 2012-2013 school years. The top and bottom 2.5% of each school characteristic are dropped in order to account for outliers and potential errors in the data that schools report.

Table 5: School Value Added Characteristics

	Low-SES	High-SES	Minority	Non-Minority	Male	Female
All Students	19,074,518	15,611,689	20,311,714	14,211,005	17,805,845	16,917,298
+ Nonmissing Test Score	18,382,026	15,340,434	19,619,420	13,939,815	17,177,068	16,581,166
+ First Test Score for Grade	17,578,554	14,731,328	18,709,351	13,450,869	16,397,313	15,931,325
+ Conventional School	16,782,019	14,154,031	17,841,509	12,957,591	15,622,334	15,330,020
+ School Size > 10	16,780,942	14,152,992	17,840,758	12,956,273	15,621,206	15,329,012
+ Nonmissing Subject Test Score	16,414,230	13,927,755	17,454,740	12,741,995	15,266,721	15,089,651
+ Nonmissing Demographic Controls	15,829,264	13,444,067	16,877,981	12,389,262	14,595,926	14,689,166
+ Nonmissing 1 Grade Prior Test Score	13,981,890	11,690,273	14,846,874	10,820,356	12,773,264	12,908,342
+ Nonmissing 2 Grade Prior Test Score	12,218,694	10,084,057	12,970,545	9,328,377	11,085,886	11,225,027
+ Nonmissing Peer Controls	12,217,342	10,082,631	12,969,315	9,326,854	11,084,463	11,223,746
+ School VA Sample Size ≥ 7	12,204,773	10,067,455	12,957,621	9,310,725	11,073,636	11,213,388

Values are counts of the number of observations in each sample. Each row is additive, so the restrictions from all prior rows are also present in the current row. The first row denotes the total number of observations in the CST dataset. The second row keeps students who have information on test scores, as opposed to just demographic characteristics. 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. I impose this restriction because students repeating a grade are tested on material for which they have already been tested at least once. The fourth row keeps only students at "conventional" schools. This includes schools in the following categories defined by the 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). The fifth row drops any schools that enroll 10 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 and ninth rows drop students who are missing test scores from one grade and two grades prior, respectively. This restriction is the cause of the vast majority of observations which are excluded from the value added estimates. The tenth row drops observations for which peer averages of the control variables could not be calculated. The eleventh row drops students if fewer than seven observations can be used to estimate value added for their school by subgroup by year cell, which insures that all value added estimates are based on at least seven observations. Data comes from grades 4-11 in public schools in the state of California between the 2004-2005 and 2012-2013 school years.

Table A.1: K–12 Counts

	Low-SES	High-SES	Minority	Non-Minority	Male	Female
Test Scores						
Current Test Score	-0.185	0.690	-0.164	0.674	0.136	0.298
	[0.873]	[0.928]	[0.883]	[0.936]	[1.012]	[0.976]
ACS Controls						
Prop. Hispanic	0.565	0.241	0.574	0.227	0.411	0.421
	[0.257]	[0.191]	[0.256]	[0.168]	[0.279]	[0.281]
Prop. White	0.224	0.500	0.232	0.493	0.356	0.345
	[0.202]	[0.227]	[0.206]	[0.231]	[0.255]	[0.253]
Prop. Asian	0.124	0.190	0.109	0.209	0.154	0.154
	[0.150]	[0.178]	[0.128]	[0.190]	[0.166]	[0.167]
Prop. Black	0.0671	0.0364	0.0667	0.0367	0.0522	0.0534
	[0.0887]	[0.0550]	[0.0912]	[0.0501]	[0.0753]	[0.0769]
Prop. Other	0.265	0.127	0.267	0.124	0.200	0.204
	[0.149]	[0.0931]	[0.150]	[0.0851]	[0.143]	[0.144]
Prop. H.S. Dropout	0.300	0.102	0.293	0.108	0.206	0.212
	[0.161]	[0.0993]	[0.167]	[0.101]	[0.167]	[0.169]
Prop. H.S. Degree	0.248	0.174	0.244	0.178	0.213	0.215
	[0.0612]	[0.0781]	[0.0639]	[0.0792]	[0.0788]	[0.0782]
Prop. Assoc. Degree	0.0684	0.0858	0.0689	0.0855	0.0768	0.0761
	[0.0320]	[0.0310]	[0.0323]	[0.0309]	[0.0327]	[0.0327]
Prop. $\geq$ Bach. Degree	0.178	0.422	0.188	0.413	0.294	0.287
	[0.132]	[0.196]	[0.145]	[0.198]	[0.205]	[0.204]
Prop. Families in Poverty	0.171	0.0618	0.163	0.0701	0.119	0.123
	[0.111]	[0.0618]	[0.113]	[0.0712]	[0.106]	[0.107]
Med. HH Income	55.05	96.74	58.09	93.56	74.81	73.58
	[22.00]	[37.36]	[25.14]	[38.57]	[36.67]	[36.27]
Sibling Controls						
Avg. Older-Sibling Test Score	-0.243	0.568	-0.223	0.552	0.132	0.127
	[0.852]	[0.908]	[0.861]	[0.915]	[0.965]	[0.966]
Observations	1,597,942	1,347,745	1,612,444	1,330,210	1,476,087	1,476,038

Values are means and standard deviations [in brackets] of the dependent variable and a subset of the independent variables used in the test score value added estimation. Only students included in the subsample of students that have at least one older sibling and can be matched to the American Community Survey are included in this table. Columns two and three are the summary statistics for low- and high-SES students, respectively. Columns four and five are the summary statistics for minority and non-minority students, respectively. Columns six and seven are the summary statistics for male and female students, respectively. Data comes from grades 4-11 in public schools in the state of California between the 2004-2005 and 2012-2013 school years.

Table A.2: ACS and Sibling Subsample Summary Statistics

	Low-SES	High-SES	Minority	Non-Minority	Male	Female
Peer Controls						
Peer Avg. Age in Years	13.48	13.71	13.53	13.66	13.57	13.60
	[2.168]	[2.157]	[2.162]	[2.165]	[2.165]	[2.164]
Peer Avg. Male	0.510	0.510	0.510	0.511	0.511	0.509
	[0.0404]	[0.0379]	[0.0394]	[0.0389]	[0.0391]	[0.0391]
Peer Avg. Hispanic or Latino	0.630	0.298	0.637	0.284	0.473	0.482
	[0.256]	[0.212]	[0.254]	[0.195]	[0.288]	[0.289]
Peer Avg. White	0.165	0.436	0.175	0.427	0.294	0.285
	[0.184]	[0.232]	[0.189]	[0.238]	[0.248]	[0.246]
Peer Avg. Asian	0.113	0.192	0.0964	0.212	0.149	0.149
	[0.145]	[0.183]	[0.120]	[0.194]	[0.168]	[0.169]
Peer Avg. Black or African American	0.0768	0.0508	0.0753	0.0524	0.0642	0.0650
	[0.0925]	[0.0625]	[0.0930]	[0.0627]	[0.0798]	[0.0810]
Peer Avg. Other Race	0.0182	0.0289	0.0186	0.0286	0.0234	0.0229
	[0.0277]	[0.0345]	[0.0308]	[0.0314]	[0.0316]	[0.0313]
Peer Avg. Economic Disadvantage	0.707	0.302	0.676	0.335	0.516	0.526
	[0.230]	[0.225]	[0.254]	[0.250]	[0.304]	[0.304]
Peer Avg. Limited English Proficient Status	0.245	0.0977	0.234	0.108	0.175	0.179
	[0.153]	[0.0917]	[0.155]	[0.103]	[0.147]	[0.148]
Peer Avg. Disabled	0.0862	0.0735	0.0852	0.0746	0.0809	0.0800
	[0.0529]	[0.0493]	[0.0524]	[0.0501]	[0.0514]	[0.0517]
Peer Avg. 1 Grade Prior Test Score	-0.130	0.412	-0.106	0.390	0.124	0.114
	[0.348]	[0.391]	[0.368]	[0.403]	[0.457]	[0.457]
Peer Avg. 2 Grade Prior Test Score	-0.119	0.429	-0.0951	0.408	0.139	0.128
	[0.348]	[0.396]	[0.369]	[0.406]	[0.461]	[0.460]
Observations	1,597,942	1,347,745	1,612,444	1,330,210	1,476,087	1,476,038

Values are means and standard deviations [in brackets] of the peer (jackknife) averages of the independent variables used in the test score value added estimation. Peer averages are calculated at the school by grade by year level and include students omitted from the value added sample. Only students included in the subsample of students that have at least one older sibling and can be matched to the American Community Survey are included in this table. Columns two and three are the peer summary statistics for low- and high-SES students, respectively. Columns four and five are the peer summary statistics for minority and non-minority students, respectively. Columns six and seven are the peer summary statistics for male and female students, respectively. Data comes from grades 4-11 in public schools in the state of California between the 2004-2005 and 2012-2013 school years.

Table A.3: K–12 Peer Summary Statistics

	Low-SES	High-SES	Minority	Non-Minority	Male	Female
Postsecondary Enrollment Outcomes						
Enrolled at a Postsecondary Institution	.536 [.499]	.744 [.436]	.537 [.499]	.749 [.434]	.601 [.49]	.668 [.471]
Enrolled at a 2-Year College	.327 [.469]	.323 [.468]	.332 [.471]	.318 [.466]	.335 [.472]	.316 [.465]
Enrolled at a 4-Year University	.209 [.407]	.421 [.494]	.205 [.404]	.431 [.495]	.266 [.442]	.352 [.477]
Enrolled at a Public Institution	.497 [.5]	.628 [.483]	.49 [.5]	.641 [.48]	.541 [.498]	.578 [.494]
Enrolled at a Private Institution	.0388 [.193]	.116 [.32]	.0473 [.212]	.108 [.31]	.06 [.238]	.0902 [.286]
Enrolled at a CA Institution	.515 [.5]	.627 [.484]	.506 [.5]	.64 [.48]	.54 [.498]	.596 [.491]
Enrolled at an Out-of-State Institution	.0212 [.144]	.117 [.321]	.0307 [.173]	.109 [.311]	.0609 [.239]	.072 [.258]
Prop. Older Siblings Enrolled 2-Year	.286 [.421]	.289 [.436]	.292 [.426]	.283 [.431]	.292 [.43]	.285 [.427]
Prop. Older Siblings Enrolled 4-Year	.184 [.364]	.387 [.47]	.182 [.365]	.394 [.47]	.28 [.43]	.279 [.429]
Observations	168,349	150,972	172,743	146,419	158,598	161,250

Values are means and standard deviations [in brackets]. Only students included in the subsample of students that have at least one older sibling and can be matched to the American Community Survey are included in this table. Columns two and three are the summary statistics for low- and high-SES students, respectively. Columns four and five are the summary statistics for minority and non-minority students, respectively. Columns six and seven are the summary statistics for male and female students, respectively. Data comes from grade 11 in public schools in the state of California between the 2008-2009 and 2012-2013 school years.

Table A.4: ACS and Sibling Subsample Postsecondary Summary Statistics

	SES	Ethnicity	Sex
Base Controls			
VA Specification Test: Contemporaneous Score	1.006***	1.005***	1.008***
	(0.002)	(0.002)	(0.002)
	[1.002,1.009]	[1.002,1.009]	[1.005,1.012]
VA Forecast Bias Test: Prior Score	-0.008***	-0.008***	-0.010***
	(0.002)	(0.002)	(0.002)
	[-0.012,-0.005]	[-0.011,-0.005]	[-0.013,-0.006]
Base + Neighborhood + Sibling + Peer Controls			
VA Specification Test: Contemporaneous Score	1.020***	1.021***	1.027***
	(0.002)	(0.002)	(0.003)
	[1.016,1.025]	[1.016,1.025]	[1.022,1.032]
VA Forecast Bias Test: Prior Score	0.014***	0.016***	0.016***
	(0.004)	(0.004)	(0.004)
	[0.007,0.021]	[0.009,0.023]	[0.008,0.023]

Each cell represents a separate regression. The first and third rows contain the coefficient for a bivariate regression of test score residuals r_{isgdt} on school by subgroup value added $\hat{\mu}_{sdt}$. Statistical inference is conducted under the null hypothesis that the coefficient equals 1. The second and fourth rows contain the coefficient for a regression of the projection of contemporaneous test scores onto three grade prior test scores on school value added $\hat{\mu}_{sdt}$. Statistical inference is conducted under the null hypothesis that the coefficient equals 0. Rows one and two give results for the base value added sample using the base value added controls. Rows three and four give results for the subsample of students that have at least one older sibling and can be matched to the American Community Survey and controls for ACS Census tract characteristics, average older-sibling test scores, and peer (jackknife) averages of the independent variables used in the estimation of value added. Standard errors clustered at the school level are presented in parentheses. The 95% confidence intervals are presented in brackets. Data comes from grades 4-11 in public schools in the state of California between the 2004-2005 and 2012-2013 school years.

Table B.1: School Test Score Value Added Specification/Forecast Bias Tests

	SES	Ethnicity	Sex
Base Controls			
VA Specification Test: Contemporaneous Score	1.013***	1.009**	1.021***
	(0.005)	(0.004)	(0.005)
	[1.004,1.022]	[1.001,1.018]	[1.011,1.030]
VA Forecast Bias Test: Prior Score	0.005***	0.004***	0.005***
	(0.000)	(0.000)	(0.001)
	[0.004,0.006]	[0.003,0.005]	[0.004,0.006]
Base + Neighborhood + Sibling + Peer Controls			
VA Specification Test: Contemporaneous Score	1.022	1.014	1.012
	(0.021)	(0.019)	(0.021)
	[0.981,1.062]	[0.977,1.050]	[0.971,1.052]
VA Forecast Bias Test: Prior Score	-0.004***	-0.003***	-0.001
	(0.001)	(0.001)	(0.001)
	[-0.006,-0.001]	[-0.005,-0.001]	[-0.004,0.001]

Each cell represents a separate regression. The first and third rows contain the coefficient for a bivariate regression of postsecondary enrollment residuals r_{isgdt} on school by subgroup value added $\hat{\mu}_{sdt}$ in grade 11. Statistical inference is conducted under the null hypothesis that the coefficient equals 1. The second and fourth rows contain the coefficient for a regression of the projection of postsecondary enrollment onto three grade prior test scores on school value added $\hat{\mu}_{sdt}$ in grade 11. Statistical inference is conducted under the null hypothesis that the coefficient equals 0. Rows one and two give results for the base value added sample using the base value added controls. Rows three and four give results for the subsample of students that have at least one older sibling and can be matched to the American Community Survey and controls for ACS Census tract characteristics, the proportion of older siblings that attended a 2-year or 4-year university, and peer (jackknife) averages of the independent variables used in the estimation of value added. Standard errors clustered at the school level are presented in parentheses. The 95% confidence intervals are presented in brackets. Data comes from grade 11 in public schools in the state of California between the 2004-2005 and 2012-2013 school years.

Table B.2: School Postsecondary Enrollment Value Added Specification/Forecast Bias Tests

	SES		Ethnicity		Sex	
	Test Score	Enrollment	Test Score	Enrollment	Test Score	Enrollment
FTE Teachers per Student	30.1***	-11.3***	26.9***	-9.27**	23.8***	-6.13**
	(4.02)	(3.8)	(3.11)	(4.07)	(2.11)	(2.99)
Low-SES/Minority/Male × FTE Teachers per Student	-10.8***	15.9***	-6.6**	9.45**	-1.81**	-.0294
	(3.96)	(4.18)	(2.79)	(3.75)	(.763)	(1.94)
FTE Pupil Services per Student	-29	-47.2***	-27.4	-19.1	-54.7*	-38.2***
	(23)	(15.9)	(21.3)	(16.7)	(31.6)	(12.1)
Low-SES/Minority/Male × FTE Pupil Services per Student	-51.6***	43.5***	-44.1***	-7.54	17.1***	10.7
	(7.39)	(16.1)	(10.8)	(15.2)	(5.6)	(6.61)
English Learner Staff Per Student	3.16	-7.29***	3.38	-4.67***	6.37	-3.88***
	(2.6)	(1.66)	(2.76)	(1.61)	(4.42)	(1.4)
Low-SES/Minority/Male × English Learner Staff Per Student	8.43**	7.18***	6.27**	1.82	-2.03*	-4.23***
	(3.54)	(2.1)	(3.09)	(1.68)	(1.08)	(1.03)
Proportion $\leq$ 3 Years Experience Teachers	-1.53***	-1.04***	-1.24***	-.538**	-1.18***	-.69***
	(.0993)	(.239)	(.106)	(.257)	(.0741)	(.18)
Low-SES/Minority/Male × Proportion $\leq$ 3 Years Experience Teachers	.516***	1.1***	.13	.463*	.00141	.0814
	(.103)	(.239)	(.11)	(.244)	(.0216)	(.095)
Proportion Full Credential Teachers	3.71***	2.14***	3.3***	1.17**	2.79***	1.18***
	(.171)	(.487)	(.19)	(.513)	(.114)	(.364)
Low-SES/Minority/Male × Proportion Full Credential Teachers	-1.15***	-2.26***	-.609***	-1.02**	.0031	.298
	(.17)	(.528)	(.196)	(.512)	(.0408)	(.216)
Proportion Male Teachers	-4.44***	.962**	-4.21***	.937**	-3.97***	1.21***
	(.0635)	(.387)	(.0684)	(.412)	(.049)	(.367)
Low-SES/Minority/Male × Proportion Male Teachers	.9***	-.127	.567***	-.298	.178***	-1.26***
	(.0645)	(.394)	(.0696)	(.365)	(.0193)	(.154)
Proportion Minority Teachers	-1.3***	-2.63***	-.726***	-1.44***	-.226***	-1.07***
	(.107)	(.243)	(.116)	(.287)	(.0682)	(.156)
Low-SES/Minority/Male × Proportion Minority Teachers	1.55***	2.53***	.858***	.902***	.0428**	-.563***
	(.112)	(.234)	(.129)	(.284)	(.0174)	(.0784)
Instruction Expenditures ($1,000s) per Student	.00243	-.0182**	.00559	-.0122	-.00197	-.00717
	(.00415)	(.00726)	(.00381)	(.00771)	(.00214)	(.00497)
Low-SES/Minority/Male × Instruction Expenditures ($1,000s) per Student	-.00592	.0189***	-.01***	.00936	-.000172	.00348
	(.00377)	(.00596)	(.00335)	(.00632)	(.000562)	(.00276)
Pupil Services Expenditures ($1,000s) per Student	-.203***	-.182***	-.173***	-.168***	-.116***	-.0897***
	(.0335)	(.0379)	(.0344)	(.0368)	(.017)	(.0293)
Low-SES/Minority/Male × Pupil Services Expenditures ($1,000s) per Student	.136***	.143***	.0978***	.108***	.00556	-.0431***
	(.0304)	(.04)	(.0325)	(.0409)	(.00349)	(.0141)
Ancillary Services Expenditures ($1,000s) per Student	-.439***	-.0677	-.403***	-.313**	-.435***	-.199*
	(.128)	(.149)	(.131)	(.158)	(.102)	(.119)
Low-SES/Minority/Male × Ancillary Services Expenditures ($1,000s) per Student	.065	-.23	-.0238	.113	.0846***	.0627
	(.0921)	(.148)	(.111)	(.151)	(.0299)	(.0785)
Other Expenditures ($1,000s) per Student	-.00744**	.0022	-.00159	.0116*	-.0124***	-.00895
	(.00356)	(.00618)	(.00328)	(.0062)	(.00289)	(.00603)
Low-SES/Minority/Male × Other Expenditures ($1,000s) per Student	-.00998***	-.00836	-.0196***	-.0219***	.00108	.0147***
	(.0032)	(.00633)	(.00321)	(.00629)	(.000685)	(.00302)
General Administration Expenditures ($1,000s) per Student	-.011*	-.0171	-.0111**	-.0218**	-.00614	-.00893
	(.00603)	(.0105)	(.00486)	(.00875)	(.00475)	(.0102)
Low-SES/Minority/Male × General Administration Expenditures ($1,000s) per Student	.00683	.015*	.00555	.0219**	-.000313	.00694
	(.00635)	(.00865)	(.0058)	(.00892)	(.00145)	(.00484)

Every two rows within each column represent a separate trivariate regression of school by subgroup value added on a school characteristic, a subgroup fixed effect, and the school characteristic interacted with the subgroup fixed effect. Standard errors clustered at the school level are presented in parentheses. In columns two and three the subgroup is SES, and the indicator variable on "low-SES/minority/male" is low-SES. In columns four and five the subgroup is ethnicity, and the indicator variable on "low-SES/minority/male" is minority. In columns six and seven the subgroup is sex, and the indicator variable on "low-SES/minority/male" is male. In columns two, four, and six the dependent variable is school by subgroup value added on ELA test scores. In columns three, five, and seven the dependent variable is school by subgroup value added on postsecondary enrollment. Data comes from grades 4-11 in public schools in the state of California between the 2004-2005 and 2012-2013 school years. The top and bottom 2.5% of each school characteristic are dropped in order to account for outliers and potential errors in the data that schools report.

Table C.1: School Value Added Characteristics