The Impact of School Quality on Postsecondary Success: Evidence in the Era of Common Core*

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

[1]University of California, Davis

[2]Washington and Lee University

March 23, 2021

Abstract

This paper explores high school quality in California after the transition to Common Core State Standards (CCSS). Using a longitudinal panel of students' standardized test scores, we estimate high school test score value added in English and mathematics for the 2015-2017 cohorts of 11th grade students. We then link these student-level data to college enrollment records to estimate college enrollment value added. We decompose the college enrollment value added into two components: the persistence of test score value added and non-test score factors (e.g., college counseling services) that influence college enrollment. Results show that there is substantial variation in school quality as measured by both test scores and college enrollment. A one-standard deviation increase in school quality is associated with a 0.15 standard deviation increase in standardized test scores and an 8-percentage point increase college enrollment. Importantly, our results show that both the persistence in test score value added and other non-test score factors within a school are important determinants of college-going value added.

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

1 Introduction

Assessing school performance is an important challenge confronting parents, policymakers, and school officials. A related difficulty lies in creating incentives that encourage schools to maintain and improve educational quality. Test-based accountability programs such as the No Child Left Behind Act of 2002 have been the primary mechanism policymakers have used both to measure and incentivise school effectiveness. These accountability systems involve universal standardized testing of students, making the results of these tests publicly available, and tying sanctions and rewards for schools to student test performance.

Two key assumptions underlie accountability policies. The first is that schools make significant contributions to student test performance. Second, a school's impact on test scores reflects students acquiring productive skills. Both of these assumptions might not hold. For instance, variation in test performance across schools might be driven by differences in student background. Similarly, any effect schools have on test scores might simply reflect "gaming" by schools (e.g., focusing on test preparation or manipulating the set of students taking the tests).[1] Although the usefulness of accountability policies and extensive standardized testing rests in large part on whether these assumptions hold, it remains unclear how much schools contribute to student test scores and whether these contributions translate into socially-valuable learning.

In this paper, we use rich statewide data from California that links K-12 student records to information on college enrollment to examine schools' contribution to test performance, post-secondary schooling outcomes, and the relationship between the two. We begin by estimating the effect of high schools on 11th grade math and English tests. The key empirical challenge for this analysis is distinguishing a school's impact on test scores from the influence of student composition or statistical noise. To address these problems, we estimate "value-added" models by adapting the procedure Chetty, Friedman and Rockoff (2014) use to estimate teacher effects. We show that the resulting estimates of school effectiveness have minimal correlation with lagged test scores that are not used as controls in the value-added estimation, providing support for the strong "selection on observables" assumption required for the validity of the value-added estimates. This is important given recent studies suggesting that estimates of school quality from observational value-added models are biased

[1]Studies examining evidence of gaming as a response to school accountability policies includes Figlio and Winicki, 2005; Jacob, 2005; Dee and Jacob, 2010; Dee, Jacob and Schwartz, 2013; Neal and Schazenbach, 2010; Figlio and Rouse, 2006; Reback, 2008; Chiang, 2009; and Rockoff and Turner, 2010. See Figlio and Loeb (2011) for a review of the literature on school accountability.

(Angrist et al., 2017; Deming, 2014).[2]

We then examine the link between a school's impact on test scores and its effects on longer-run outcomes, specifically postsecondary schooling. First, we estimate the relationship between college outcomes and a school's test score impact. This association is informative about the extent to which test score gains attributable to a student's high school "persist" to longer-run outcomes. We also estimate value-added models using postsecondary schooling as the outcome to measure a school's effect on college-going. We then decompose the total effect on postsecondary schooling into a portion related to the school's contribution to test score gains and a residual component. This decomposition tells us what share of the variation in school impacts on college-going is "explained" by variation in school-level test score impacts. This share is informative about the desirability of a heavy reliance on student test scores for evaluating school effectiveness. In particular, if this share is small, it would suggest that it might be beneficial to incorporate other indicators of school performance in an accountability system.

The findings suggest that schools make important contributions to both test scores and college going. A one standard deviation increase in estimated school effectiveness is associated with a 0.153 standard deviation increase in math test scores and a 8.7 percentage point increase in the likelihood of attending a four-year college. We also find that test score impacts are strongly related to improved college going. An increase of one-standard deviation in a school's math test score impact is associated with a 2.4 percentage point increase in four-year college attendance. Nonetheless, much of the variation in school effectiveness as measured by college going is not accounted for by a school's test score impact. In particular, we estimate that less than one-third of the total variance in college-going value-added is explainable by a school's contribution to test score gains.

Our study contributes to a growing literature on estimating school quality using value-added methods. Angrist et al. (2017), Deming (2014), and Dobbie and Fryer (2013) compare estimates of school effectiveness from conventional value-added models similar to those we use to estimates based on randomized lotteries to oversubscribed charter schools. They find that conventional estimates are useful for evaluating school performance despite having some bias. Other studies have considered the relationship between school impacts on test scores and longer-run outcomes. Dobbie and Fryer (2017) find that Texas charter schools that reduce test scores also reduce labor market earnings,

[2] Value-added models were developed by researchers attempting to estimate teacher effectiveness, and a large literature examines the validity of value-added models in this context (see for example Rothstein, 2010; Rothstein, 2017; Chetty, Friedman and Rockoff, 2017; Bacher-Hicks, Kane, and Staiger, 2016).

while those that increase test scores have no detectable effect on earnings. Abdulkadiroglu et al. (2017) find that school impacts on high school test scores are strongly correlated with school impacts on college outcomes in New York City. Hubbard (2018) shows that attending a school with high test score value-added is associated with better performance in college.

Our work is also related to work on the long-run impact of school accountability policies. Deming et al. (2016) find that students subjected to strong accountability pressure in Texas experienced significant test score gains as well as improved longer-run outcomes, while lower-scoring students suffered negative effects. This paper shows that policy-generated test score gains are associated with better long-run outcomes. We build on that work by examining how much of the overall variation in schools' contributions to test scores can be explained by school effects on test scores.

The paper is organized as follows. In Section 2, we describe the data used for the study. Section 3 presents the value-added methodology we use and the main results. Section 4 discusses mechanisms for our findings. Section 5 concludes.

2 Data

Common Core State Standards were first implemented in California during the 2013-14 academic year, with all students first taking the Smarter Balanced Assessment Consortium (SBAC) tests during the spring of 2015. To measure school quality in the era of Common Core, our dataset consists of three 11th grade cohorts of students from 2014-15 through 2016-17 who attended California public high schools and took the SBAC tests[3]. In total there are 1,431,423 students in our sample who attended one of California's 1,339 public high schools.

We place a series of restrictions on the value added sample which are similar to those found in other work on value added. First, we require that high schools serve only grades 9 through 12, in order to eliminate K-12, 6-12, and a variety of other schools that teach grades outside of traditional high school grades. Second, we use only one observation per student per grade. For students who repeated a grade, we use only the test score from the earliest instance of that grade, and for students who have multiple observations per year we select the observation with non-missing test score data that has the highest number of non-missing demographic variables, breaking ties by randomly selecting a single observation. Third, we only keep students attending conventional high

[3] We keep only students who meet the characteristic "student was enrolled during active testing window, completed the test, and has met completion/attemptedness."

schools.[4] Fourth, we require that the high school enroll at least ten 11th graders. Finally, we keep only students with a valid 11th grade test score, as this is necessary in order to estimate school value added. These restrictions leave us with 1,122,759 students.

To these data, we link student by year level data from the California Standards Tests taken in previous years. We drop students without a valid state student ID, as we cannot match them to prior test scores, and only use prior scores for which the lagged birth date is equivalent to current birth date in order to ensure that we are tracking the same student over time. In total we were able to match 900,464 students to pre-high school test scores in English Language Arts (ELA) and mathematics with a complete set of demographic characteristics. For ELA, we use 8th grade scores. For mathematics, we have to use 6th grade scores because that is the last grade in which all students take the same mathematics exam.[5] The demographic characteristics in our dataset include race/ethnicity, gender, economic disadvantage[6], limited English proficiency, and disability, and a linear age variable. Our value added sample places the final restriction that schools must have at least seven 11th graders who meet all the prior restrictions regarding school characteristics and nonmissing test score and demographic information. This leaves us with our final value added sample of 900,252 students. Table 1 gives the number of students and average test score for the sample after imposing each of these restrictions.

We then match these student-level data to college enrollment records from the National Student Clearinghouse and supplement these with application, enrollment, and degree receipt records from the California State University and the California Community College systems to estimate each high school's value added on college enrollment.[7] Since college enrollment records are not available for the 2016-17 11th grade cohort, this match only covers the first two cohorts of 11th graders in our data.

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

[5]Students in California take statewide standardized tests in grade 2-8 and 11. We use the most recent ELA test taken in 8th grade. However, under the CST, starting in 7th grade, mathematics tests were not standardized, rather they were specific to student placement into math courses. Therefore, we use 6th grade mathematics test scores since these were the most recent common mathematics test taken by students in our sample.

[6]Economic disadvantage is measured by eligibility for free/reduced price lunch or neither parent obtained a high school degree.

[7]Because we don't have a universal identifier for students in both the K-12 and postsecondary data, we match on first, middle, and last name, birth date, and gender. We allow for fuzzy matching on names to account for nicknames or misspellings in the data. We use a sieve procedure to perform the match so that the first round of names are matched on first, middle, and last name, birth date, and gender, the second round matches on first and last name, birth date, and gender, and the final round matches on first and last name and birth date.

Table 1: Sample Counts

	ELA		Math	
	# of Students	Z-Score Mean	# of Students	Z-Score Mean
All Students	1,431,423	6.05e-07	1,431,423	-2.43e-07
+ 9-12 School	1,342,420	.0027	1,342,420	.00775
+ First Test Score for Grade	1,327,469	.00567	1,327,469	.0106
+ Conventional School	1,198,999	.081	1,198,999	.0851
+ 11th Graders per School > 10	1,198,441	.0811	1,198,441	.0853
+ Nonmissing Subject Test Score	1,122,759	.0811	1,119,723	.0853
+ Nonmissing Demographic Controls	1,117,769	.0815	1,114,638	.0858
+ Nonmissing Prior Test Scores	900,464	.199	896,279	.183
+ School VA Sample Size ≥ 7	900,275	.199	896,080	.183

Values are counts of the number of observations in each sample along with the average test score for the sample. Each row is additive, so the restrictions from all prior rows are also present in the current row. Data comes from public schools in the state of California between the 2014-2015 and 2016-2017 school years.

Finally, we augment these student-level data with publicly-available high school statistics provided by the California Department of Education. These data contain information on school-level inputs such as instructional and non-instructional spending as well as teacher characteristics such as race, gender, and education.

Table 2, Panel A shows summary statistics for 11th grade test takers from 2014-15 through 2016-17 in our value added estimating sample, which includes only those students who meet the restrictions mentioned above, as well as for 11th grade students who were excluded from our value added sample due to missing demographic characteristics or not meeting other restrictions. California students are highly diverse. Among the 11th grade value added sample, 52 percent are Hispanic/Latino, 26 percent are white, 14 percent are Asian, and 5 percent are black. Over half of all students are economically disadvantaged, while 5 percent have limited English proficiency and 3 percent have a disability. Test scores have been standardized to have mean zero and standard deviation one within each grade by year by subject cell for all test takers in that year, thus our value added estimating sample shows that the value added sample is positively selected, with test scores on average nearly 0.2 standard deviations above the mean. Additionally, students in the estimating sample are less likely to be an underrepresented minority, economically disadvantaged, LEP, or disabled. The fact that our value added sample is positively selected is not surprising and is similar to the large teacher value added literature that also relies on matching students to prior test scores Raj Chetty, John N

Friedman and Jonah E Rockoff (2014).

Table 2, Panel B shows college going summary statistics for the 2014-2015 and 2015-2016 11th grade cohorts. College going rates in California are quite high, with 60% of all 11th-grade test takers enrolling in college the year after scheduled high school graduation. College-going rates among students we use to estimate value-added are even higher (72%), which reflects the positive selection into the estimation sample. A higher proportion of students in our value added sample enroll in a 2-year college relative to a 4-year college, although this gap is smaller than it is for the full sample. The vast majority of college-goers attend a public institution in California.

3 Methods and Main Results

In the subsections below, we describe how we use a value-added framework to estimate school effectiveness. We use school impacts on contemporaneous 11th-grade ELA and mathematics test scores and on college enrollment outcomes as measures of school effectiveness. Value-added on test scores is important since test scores are the primary metrics used in school accountability systems, while value-added on college-going provides an indication of how a school affects longer-run well-being. We then examine the relationship between test score and college-going value-added. First, we examine the correlation between college enrollment value-added and test score value-added. This association is informative about whether a school's contribution to test scores "persists" to longer-run outcomes. We then decompose the estimated college enrollment value added into two components: the persistence of test score value added and factors that are orthogonal to test score value added. This decomposition allows us to see how important test score value-added is for explaining total variation in college-going value-added relative to other factors such as the quality of college counseling services that might influence college enrollment but have little effect on test scores.

Table 2: Summary Statistics

	VA Sample		Excluded	
	Mean	Standard Deviation	Mean	Standard Deviation
Panel A: 11th Grade Characteristics				
11th Graders per School	473	[199]	393	[567]
Age in Years	16.7	[0.399]	17.0	[1.452]
Male	0.490		0.548	
Asian	0.141		0.101	
Hispanic or Latino	0.523		0.520	
Black or African American	0.051		0.083	
Other Race	0.030		0.062	
Economic Disadvantage	0.536		0.585	
Limited English Proficiency Status	0.050		0.189	
Disabled	0.030		0.200	
White	0.255		0.259	
ELA Z-Score	0.199	[0.893]	-0.454	[1.080]
Math Z-Score	0.183	[0.932]	-0.416	[1.025]
Prior ELA Z-Score	0.125	[0.955]	-0.343	[1.019]
Prior Math Z-Score	0.125	[0.978]	-0.376	[0.951]
Observations	900,275		531,148	
Panel B: Postsecondary Outcomes				
Enrolled at a Postsecondary Institution	0.715		0.426	
Enrolled at a 2-Year College	0.384		0.271	
Enrolled at a 4-Year University	0.331		0.155	
Enrolled at a Public Institution	0.653		0.384	
Enrolled at a Private Institution	0.062		0.042	
Enrolled at a CA Institution	0.645		0.373	
Enrolled at an Out-of-State Institution	0.070		0.053	
Observations	594,007		371,750	

Data comes from public schools in the state of California between the 2014-2015 and 2016-2017 school years. Panel A contains the sample used to estimate ELA test score value added with the exception of "math z-score", which comes from the math test score value added sample. Panel B contains the subset of panel A students who could be linked to the NSC data and only includes the 2014-2015 and 2015-2016 cohorts. The summary statistics are essentially identical between the ELA and math samples. The math test score value added sample contains 896,080 observations and 561,975 of these were linked to the NSC data. "VA Sample" refers to the ELA test score value added sample. "Excluded" refers to the students excluded from the ELA test score value added sample because they did not meet all of the requirements for inclusion.

3.1 Test Score Value Added

To measure each high school's value added on student's 11th grade test scores, we consider the following model in equation (1):

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

where Y_{ist} is student i's 11th grade test score in school s and year t. X_{ist} is a vector of controls including the number of students in the school's 11th grade cohort, prior test scores, student age, and indicators for gender, race and ethnicity (Hispanic/Latino, white, Asian, black, "other"), economic disadvantage, limited English proficiency, and disability. For prior test scores, we use cubic polynomials in 8th-grade ELA scores and 6th-grade math scores. γ_t are year fixed effects. The error term u_{ist} is comprised of three components: school by year value added λ_{st}, a school by year common shock ξ_{st}, and a student-specific random term ϵ_{ist}.

The goal of the estimation is to isolate λ_{st}, which provides a measure of a school's contribution to test scores beyond what would be expected given student demographic characteristics and prior test scores. To do so, we use the value added with drift procedure developed in Chetty, Friedman and Rockoff (2014) for estimating teacher effectiveness.[8] The first step involves estimating Equation 1 and computing the residuals (which we refer to as "student performance residuals"). We then collapse the student performance residuals u_{ist} to the school by year level:

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

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

[8] In practice, Chetty, Friedman and Rockoff (2014) is a reweighted version of Carrell and West (2010). Appendix section A of this paper discusses the robustness of our main results to using the alternative methodology in Carrell and West (2010).

$\bar{\epsilon}_{st}$ will converge to zero.

In order to reduce the variation due to the common shocks ξ_{st}, our value added estimates are the predicted value from a regression of u_{st} on $\boldsymbol{u}_{st'}$, where $\boldsymbol{u}_{st'}$ is the vector of $u_{st'}$ for all $t' \neq t$. In the ideal case that the common shocks are uncorrelated with school value added ($cov(\lambda_{st}, \xi_{st'}) = 0$) and the common shocks are uncorrelated across time ($cov(\xi_{st}, \xi_{st'}) = 0$), this functions as the first stage of an instrumental variables regression in which we isolate variation in λ_{st} while eliminating variation in ξ_{st}. The variation in λ_{st} under this methodology comes from the assumption that school value added is correlated from year to year ($cov(\lambda_{st}, \lambda_{st'} \neq 0)$), which is likely given that most schools will not experience complete faculty and staff turnover between years. Chetty, Friedman and Rockoff (2014) and Naven (2019) provide additional methodological details.[9]

Figure 1 shows the distribution of estimated school test score value-added. The results show that schools vary substantially in their contribution to ELA and mathematics value-added. The standard deviation in test score value-added (σ_λ) is 0.142 for ELA and 0.153 for mathematics. This indicates that a one-standard deviation increase in school effectiveness on test scores is associated with an increase in average student test scores of about 0.15 standard deviations of the student standardized test score distribution. These magnitudes are similar to estimates of school effectiveness from other settings.[10] They are also strikingly similar to those found in the teacher quality literature which show that a one-standard deviation increase in teacher quality is associated with between a 0.10 and 0.20 increase in student achievement (Kane, Rockoff and Staiger, 2008; Chetty, Friedman and Rockoff, 2014).

The validity of these value-added measures depends on the strong assumption that, after controlling for pre-high school test scores and student demographics, variation in high school test scores across schools reflects the influence of schools and not omitted variables. This assumption has been criticized in the teacher effectiveness literature by Rothstein (2009, 2010, 2017). Similarly, while recent studies show that school value-added estimates based on observational methods are useful for policymakers, some studies show that these estimates have some bias Angrist et al. (2017).

[9] Our value added estimates differ slightly from Chetty, Friedman and Rockoff (2014) and Naven (2019) because our estimates do not include a school fixed effect when estimating equation (1), which would account for potential correlation between school value added and the demographic characteristics of students. This is because we want to use across-school comparisons when controlling for test score value added when estimating school value added on long-run outcomes that operates independently of test scores. Test score value added estimates that include a school fixed effect in equation (1) have a correlation of 0.99 with the value added estimates used in this paper.

[10] Hubbard (2017) finds that a standard deviation in school value-added corresponds to 0.23 standard deviations in student test scores, Deming (2014) reports standard deviations ranging from about 0.05 to 0.1, and Angrist et al. (2017) report standard deviations ranging from 0.15 to 0.25.

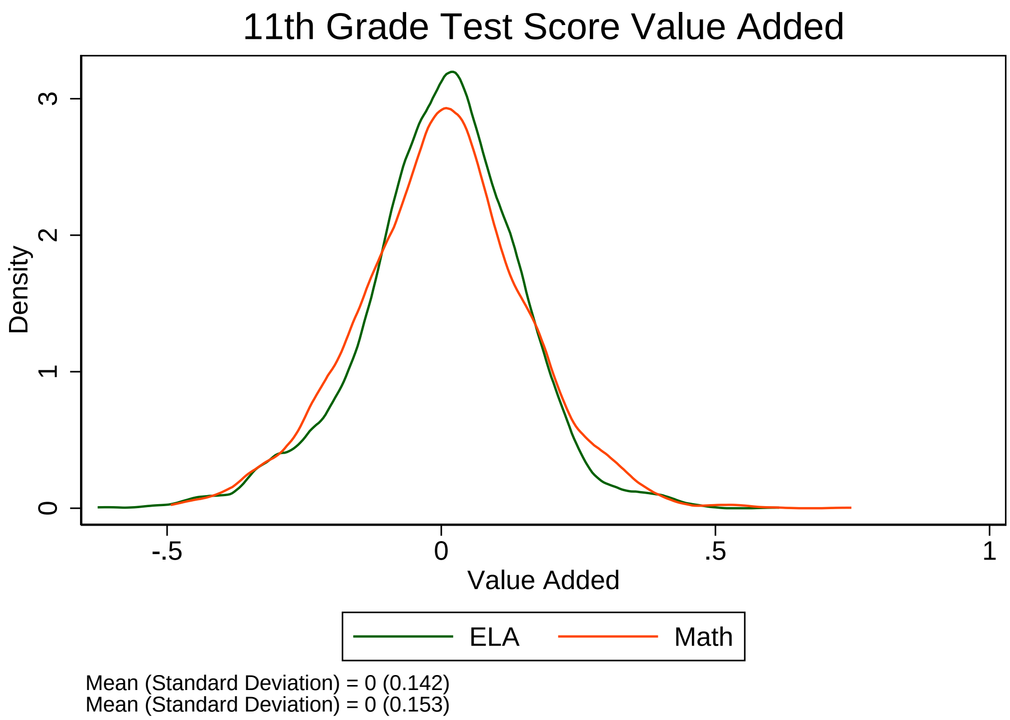


Figure 1: 11th Grade Test Score Value Added Distribution

Table 3: 11th Grade Test Score Value Added Specification/Forecast Bias Test

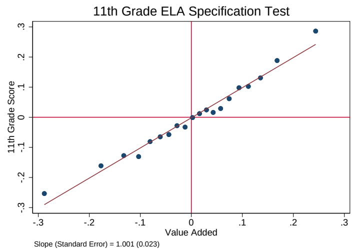
	ELA	Math
VA Specification Test: 11th Grade Score	1.001	1.030
	(0.023)	(0.011)
	[0.956,1.045]	[1.010,1.051]
VA Forecast Bias Test: Predicted Score using 7th Grade ELA Score	0.004	0.007
	(0.003)	(0.001)
	[-0.001,0.009]	[0.005,0.009]
VA Forecast Bias Test: Predicted Score using ACS Census Tract Demographics	0.029	0.154
	(0.005)	(0.015)
	[0.019,0.040]	[0.124,0.183]

Each cell represents a separate regression. The first row contains the coefficient for a bivariate regression of test score residuals u_{ist} on school value added $\hat{\lambda}_{st}$. Statistical inference is conducted under the null hypothesis that the coefficient equals 1. The second row contains the coefficient for a bivariate regression of test scores as predicted by residualized excluded observables $\hat{u}_{ist}$ on school value added $\hat{\lambda}_{st}$. Statistical inference is conducted under the null hypothesis that the coefficient equals 0. Standard errors clustered at the school level are presented in parenthesis. The 95% confidence intervals are presented in brackets.

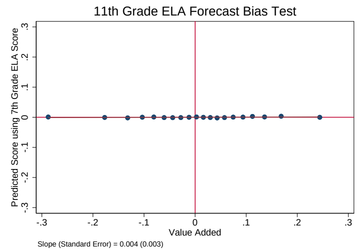
To assess the validity of our school value-added estimates, we follow Chetty, Friedman and Rockoff (2014) and Rothstein (2017) and perform specification and forecast bias tests on our value added estimates. Specifically, our specification test examines whether a change in estimated school value-added corresponds to a one-for-one change in 11th-grade student achievement. If instead there is a larger or smaller than one-for-one change in student achievement associated with a change in school value-added it would suggest that the value-added estimates were biased. The forecast bias test examines whether our value-added estimates are correlated with pre-high school test scores that are not included when estimating the student performance residuals (specifically 7th grade ELA scores).[11] If the school value-added estimates are uncorrelated with the excluded prior test scores, it would suggest that any bias from omitted variables would have to be from factors that are themselves uncorrelated with the "hold out" test scores. In both the specification and forecast bias tests we include all of the controls included in the value added specification, as we are interested in measuring the degree to which our value added estimates are correlated with *unobservable* characteristics.

Figure 2 and Table 3 show the results of these tests. Reassuringly, for both ELA and mathematics we find that our value added estimates accurately predict 11th grade test scores. A one-standard deviation increase in value added is associated with a 1.002 and 1.030 standard deviation increase in 11th grade ELA and math test scores, respectively. Likewise, our value added estimates are also shown to be virtually uncorrelated with prior test scores. The results suggest that only 0.4% and 0.7% of the variation in school value added for ELA and math respectively is due to students sorting to schools on unobservable ability as measured by 7th grade ELA test scores.

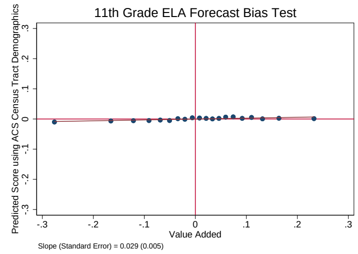
[11] The coefficient from the forecast bias test is an estimate of the value of $\frac{cov(\epsilon_{ist},\hat{\lambda}_{st})}{var(\hat{\lambda}_{st})}$.



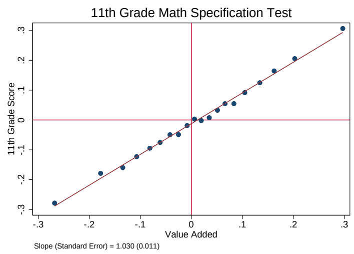
(a) ELA Specification Test



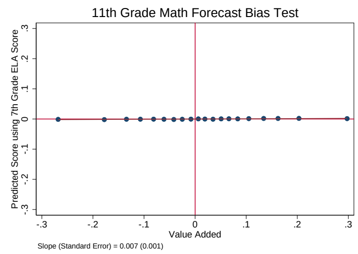
(b) ELA Forecast Bias Test



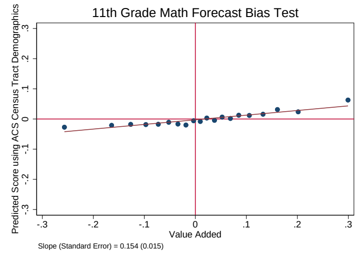
(c) ELA Forecast Bias Test



(d) Math Specification Test



(e) Math Forecast Bias Test



(f) Math Forecast Bias Test

Figure 2: 11th Grade Test Score Value Added Specification/Forecast Bias Tests

3.2 Value Added on Longer-run Outcomes

We next seek to examine whether high schools affect students' longer-run outcomes in a meaningful way. To do so, we estimate each high school's college enrollment value added and then decompose this into two components: the persistence of test score value added and other non-test score factors (e.g., college counseling services) that influence college enrollment. To do so, we consider the following model in equation (3):

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

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

The primary parameters of interest are ρ and β_{st}. The parameter ρ measures the relationship between a school's contribution to 11th grade test score gains and postsecondary schooling outcomes. In other words, ρ reflects the extent to which test score value-added "persists" to long-run outcomes; when ρ is large, schools that generate sizable test score gains also tend to induce significant numbers of students to enroll in college. Likewise, high values of β_{st} indicate that there are other factors within the school, not measured by test score gains, that improve students' college enrollment outcomes. An example of a factor that would go into β_{st} is a particularly effective college guidance counseling staff that has little role in classroom performance and hence does not affect test scores.

We start with estimating each school's total effect on college going π_{st}, which contains both the persistence of test score value added and other non-test score factors. Similar to the our estimates of test score value added, we estimate σ_π using the techniques described by Chetty, Friedman and Rockoff (2014).[12] We first compute the average performance residual for school s in year t as:

$$\begin{aligned} \nu_{st} &= \frac{1}{N_{st}} \sum_{i=1}^{N_{st}} [\rho\lambda_{st} + \beta_{st} + \theta_{st} + e_{ist}] \\ &= \rho\lambda_{st} + \beta_{st} + \theta_{st} + \frac{1}{N_{st}} \sum_{i=1}^{N_{st}} e_{ist} \\ &= \pi_{st} + \theta_{st} + \bar{\epsilon}_{st} \end{aligned} \tag{4}$$

In order to reduce the variation due to the common shocks θ_{st}, our value added estimates are the predicted value from a regression of ν_{st} on $\boldsymbol{\nu}_{st'}$, where $\boldsymbol{\nu}_{st'}$ is the vector of $\nu_{st'}$ for all $t' \neq t$.

Results are shown in Figure 3 and indicate there is substantial variation in college-going value added across schools. The estimates of σ_π (in percentage points) are 7.2 for any college enrollment, 7.7 for 2-year college enrollment, and 8.7 for 4-year college enrollment.[13]

As we did for test score value-added, we perform specification and forecast bias tests using 7th grade ELA scores. Figure 4 shows the results for any college enrollment, two-year college enrollment, and four-year enrollment. For all three outcomes, the specification tests indicate that a one-unit change in long-run value-added is associated with close to a one-to-one change in the corresponding college outcome. Similarly, long-run value-added has very little correlation with excluded pre-high

[12] Results using methods in Carrell and West (2010) NBER working paper version are shown in appendix section A.

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

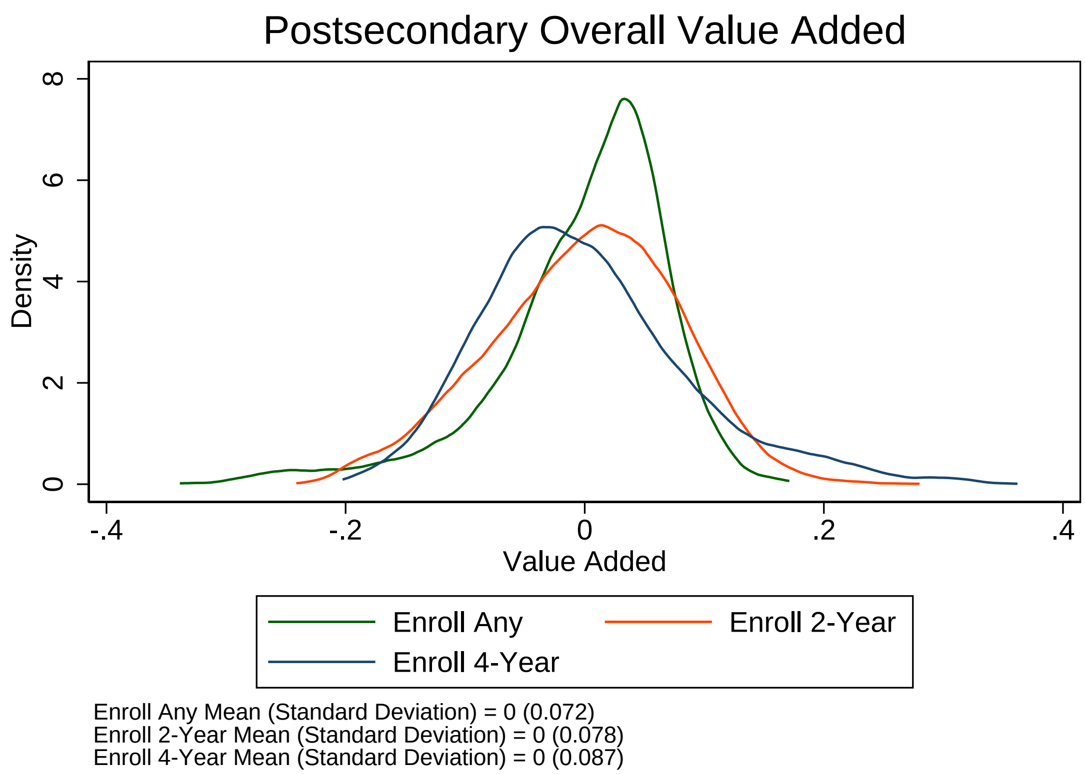
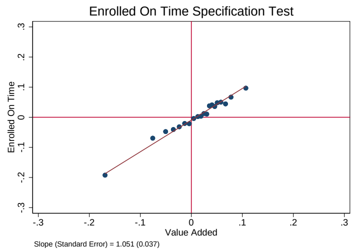
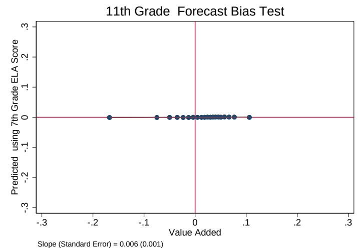


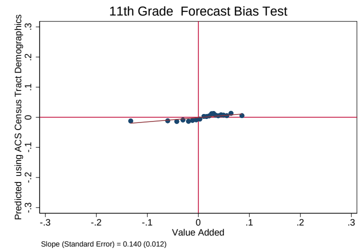
Figure 3: Overall



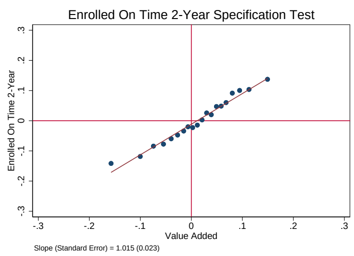
(a) Postsecondary Enrollment Specification Test



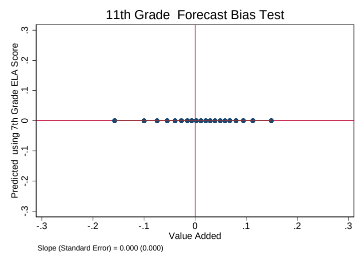
(b) Postsecondary Enrollment Forecast Bias Test



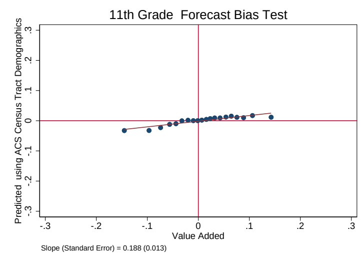
(c) Postsecondary Enrollment Forecast Bias Test



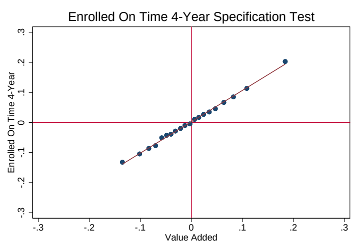
(d) 2-Year Enrollment Specification Test



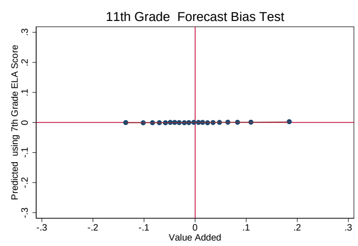
(e) 2-Year Enrollment Forecast Bias Test



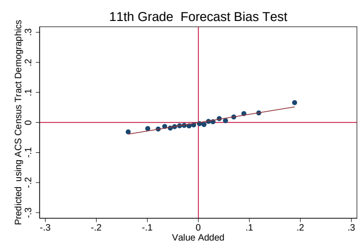
(f) 2-Year Enrollment Forecast Bias Test



(g) 4-Year Enrollment Specification Test



(h) 4-Year Enrollment Forecast Bias Test



(i) 4-Year Enrollment Forecast Bias Test

Figure 4: Postsecondary Enrollment Overall Value Added Specification/Forecast Bias Tests

school test scores.

We next decompose this effect into the persistence of test score value added and non-test score factors within the school. Estimates of ρ are shown in Table 7. To estimate ρ we estimate equation 3 by regressing Y_{ist} on X_{ist}, γ_t, and the estimated test score value added $\hat{\lambda}_{st}$, obtained using the procedure in Chetty, Friedman and Rockoff (2014). The estimates in Table 7 are the coefficients on $\hat{\lambda}_{st}$.[14] Columns 1-3 show results for the outcome of enrollment in any college. The estimated ρ of 0.201 for ELA in column 1 indicates that schools which are one-standard deviation above the mean in ELA test score value added (i.e., test scores are improved by 0.142 standard deviations) increase

[14] This approach is a reweighted equivalent of the methodologies employed by Jacob, Lefgren and Sims (2010) and Carrell and West (2010)'s NBER working paper. Appendix section A gives additional details on the similarities between the various methodologies.

Table 4: Postsecondary Enrollment Value Added Specification/Forecast Bias Test

	Overall	ELA VA	Math VA	ELA & Math VA
VA Specification Test: Enrolled On Time	1.051	1.037	1.034	1.034
	(0.037)	(0.037)	(0.036)	(0.037)
	[0.979,1.123]	[0.964,1.110]	[0.962,1.105]	[0.962,1.106]
VA Forecast Bias Test: Predicted Score using 7th Grade ELA Score	0.006	0.005	0.003	0.003
	(0.001)	(0.001)	(0.001)	(0.001)
	[0.004,0.008]	[0.004,0.007]	[0.001,0.005]	[0.001,0.005]
VA Forecast Bias Test: Predicted Score using ACS Census Tract Demographics	0.140	0.143	0.090	0.102
	(0.012)	(0.013)	(0.012)	(0.012)
	[0.117,0.164]	[0.118,0.168]	[0.067,0.113]	[0.079,0.125]

Each cell represents a separate regression. The first row contains the coefficient for a bivariate regression of postsecondary enrollment residuals u_{ist} on school value added $\hat{\lambda}_{st}$. Statistical inference is conducted under the null hypothesis that the coefficient equals 1. The second row contains the coefficient for a bivariate regression of postsecondary enrollment as predicted by residualized excluded observables $\hat{u}_{ist}$ on school value added $\hat{\lambda}_{st}$. Statistical inference is conducted under the null hypothesis that the coefficient equals 0. Standard errors clustered at the school level are presented in parenthesis. The 95% confidence intervals are presented in brackets.

Table 5: 2-Year Enrollment Value Added Specification/Forecast Bias Test

	Overall	ELA VA	Math VA	ELA & Math VA
VA Specification Test: Enrolled On Time 2-Year	1.015	1.014	1.014	1.015
	(0.023)	(0.023)	(0.024)	(0.024)
	[0.969,1.061]	[0.969,1.059]	[0.967,1.060]	[0.968,1.061]
VA Forecast Bias Test: Predicted Score using 7th Grade ELA Score	0.000	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
	[-0.000,0.001]	[-0.000,0.001]	[-0.001,0.001]	[-0.001,0.001]
VA Forecast Bias Test: Predicted Score using ACS Census Tract Demographics	0.188	0.188	0.183	0.179
	(0.013)	(0.013)	(0.013)	(0.012)
	[0.162,0.213]	[0.162,0.213]	[0.158,0.208]	[0.155,0.203]

Each cell represents a separate regression. The first row contains the coefficient for a bivariate regression of 2-year enrollment residuals u_{ist} on school value added $\hat{\lambda}_{st}$. Statistical inference is conducted under the null hypothesis that the coefficient equals 1. The second row contains the coefficient for a bivariate regression of 2-year enrollment as predicted by residualized excluded observables $\hat{u}_{ist}$ on school value added $\hat{\lambda}_{st}$. Statistical inference is conducted under the null hypothesis that the coefficient equals 0. Standard errors clustered at the school level are presented in parenthesis. The 95% confidence intervals are presented in brackets.

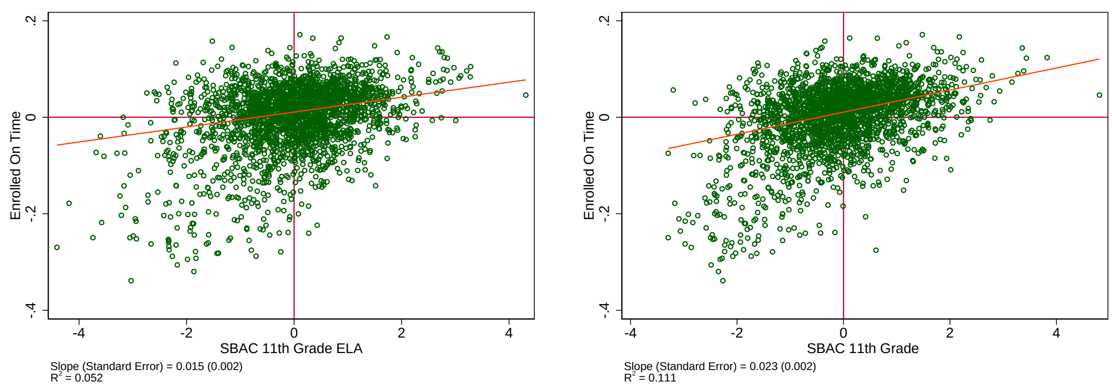
Table 6: 4-Year Enrollment Value Added Specification/Forecast Bias Test

	Overall	ELA VA	Math VA	ELA & Math VA
VA Specification Test: Enrolled On Time 4-Year	1.042	1.029	1.028	1.029
	(0.009)	(0.011)	(0.013)	(0.013)
	[1.024,1.060]	[1.008,1.050]	[1.002,1.054]	[1.003,1.055]
VA Forecast Bias Test: Predicted Score using 7th Grade ELA Score	0.009	0.009	0.005	0.005
	(0.001)	(0.001)	(0.002)	(0.002)
	[0.006,0.012]	[0.006,0.012]	[0.002,0.008]	[0.002,0.008]
VA Forecast Bias Test: Predicted Score using ACS Census Tract Demographics	0.278	0.293	0.241	0.241
	(0.018)	(0.019)	(0.018)	(0.017)
	[0.242,0.314]	[0.256,0.329]	[0.207,0.276]	[0.208,0.275]

Each cell represents a separate regression. The first row contains the coefficient for a bivariate regression of 4-year enrollment residuals u_{ist} on school value added $\hat{\lambda}_{st}$. Statistical inference is conducted under the null hypothesis that the coefficient equals 1. The second row contains the coefficient for a bivariate regression of 4-year enrollment as predicted by residualized excluded observables $\hat{u}_{ist}$ on school value added $\hat{\lambda}_{st}$. Statistical inference is conducted under the null hypothesis that the coefficient equals 0. Standard errors clustered at the school level are presented in parenthesis. The 95% confidence intervals are presented in brackets.

Table 7: Persistence of Test Score Value Added to Postsecondary Enrollment

	Enrolled			Enrolled 2-Year			Enrolled 4-Year		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ELA VA	0.194		0.026	0.020		0.053	0.174		-0.027
	(0.004)		(0.006)	(0.005)		(0.006)	(0.004)		(0.005)
Math VA		0.257	0.241		-0.015	-0.048		0.272	0.289
		(0.004)	(0.005)		(0.004)	(0.006)		(0.004)	(0.005)
Observations	622,669	622,669	622,669	622,669	622,669	622,669	622,669	622,669	622,669
R^2	.121	.124	.124	.0489	.0489	.049	.235	.238	.238



(a) ELA

(b) Math

Figure 5: Postsecondary Enrollment Value Added vs. Test Score Value Added

college enrollment by 2.85 percentage points (0.201 * 0.142 = 0.0285). This effect size corresponds to a modest 3.99 percent increase in enrollment. Similarly, the estimated ρ of 0.264 in column 2 for math indicates that a one-standard deviation increase in math value added is associated with a 4.04 percentage point (5.65 percent) increase in any enrollment (0.264 * 0.153 = 0.0404). Results in column 3 show that when we estimate persistence of both math and ELA value-added in the same regression, the persistence of math value added dominates ELA value added, as the coefficient on ELA value added is small and not statistically different than zero. The graphical evidence in Figure 5 also shows the positive association between test score and college enrollment value-added.

Columns 4-9 repeat the analysis by examining 2-year and 4-year enrollment separately. For 2-year enrollment, estimates of ρ are economically small for both ELA (0.027) and math (-0.009) value added. In contrast, the effect sizes of ρ are quite large for 4-year enrollment. Results in columns 7 and 8 indicate that a one-standard deviation increase in test score value added is associated with a 2.49 and 4.18 percentage point (7.56 and 12.70 percent) increase in 4-year enrollment for ELA and math value added, respectively. Again, the results in column 9 show that math value added dominates ELA value added when estimating the persistence terms simultaneously for four-year college enrollment.

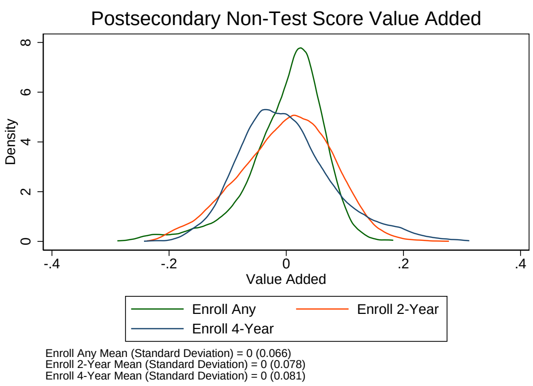
Estimates of Equation 4 with controls for test score value-added (i.e., $\hat{\lambda}_{st}$) can be used to examine the importance of non-test score factors within high schools for affecting college going. The average performance residual for school s in year t is now:

$$\begin{aligned} \nu_{st} &= \frac{1}{N_{st}} \sum_{i=1}^{N_{st}} [\beta_{st} + \theta_{st} + e_{ist}] \\ &= \beta_{st} + \theta_{st} + \frac{1}{N_{st}} \sum_{i=1}^{N_{st}} e_{ist} \\ &= \beta_{st} + \theta_{st} + \bar{e}_{st} \end{aligned} \tag{5}$$

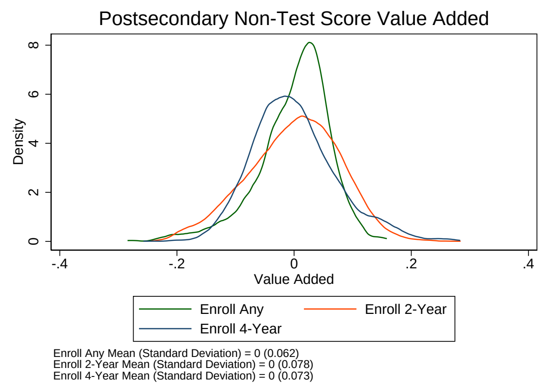
We again employ the techniques described by Chetty, Friedman and Rockoff (2014) to estimate β_{st} and its variance. Results are shown in Figure 6 and indicate there is substantial variation in non-test score college-going value added across schools. A one-standard deviation increase in non-test score school quality is associated with a 6.1 percentage point (8.5 percent) increase in any college enrollment, a 7.7 percentage point (21.9 percent) increase in 2-year enrollment and an 7.2 percentage point (20.0 percent) increase in 4-year enrollment when controlling for both ELA and math value added. Not surprisingly, test score value added has very little relationship with our estimates of college enrollment value-added that include controls for test score value-added (Figure 7).

These results indicate that schools vary substantially in school quality as measured by both standardized test scores and college enrollment, and that test score impacts persist in the sense that schools that generate larger test score gains also tend to generate better longer-run outcomes. An important question is how important is the persistence of test score value-added for explaining the total variation in long-run school value-added.

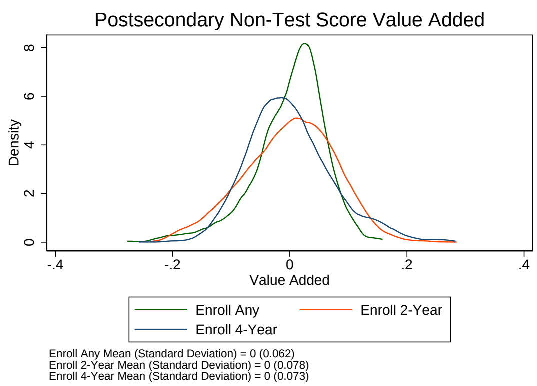
To answer this question, note that, by definition of π_{st}, Equation (6) shows that the share of total variance in long-run school effectiveness (measured by σ_{π}^2) accounted for by test score value-added



(a) Other (Non-Test Score) Factors, Controlling for ELA

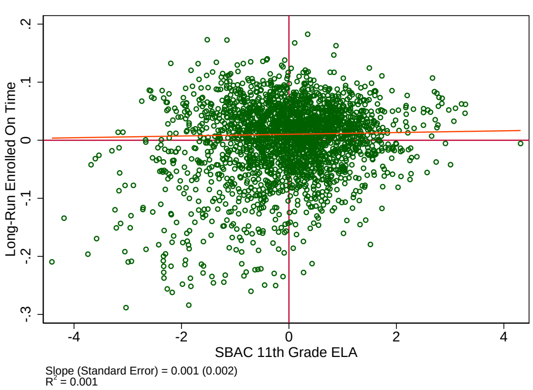


(b) Other (Non-Test Score) Factors, Controlling for Math

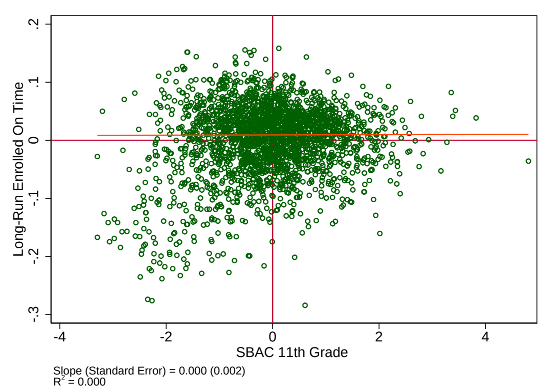


(c) Other (Non-Test Score) Factors, Controlling for ELA and Math

Figure 6: Postsecondary Enrollment Value Added Distribution



(a) ELA



(b) Math

Figure 7: Postsecondary Enrollment Other (Non-Test Score) Factors Value Added vs. Test Score Value Added

persistence is equal to $1-\frac{\sigma_\beta^2}{\sigma_\pi^2}$, where σ_β^2 is the variance of value-added due to non-test score factors.[15]

$$\sigma_\pi^2 = \rho^2\sigma_\lambda^2 + \sigma_\beta^2 \tag{6}$$

The values of σ_π and σ_β are reported in Figure 6. Focusing on the results when we control for both math and ELA test score value added persistence, we see that test score persistence accounts for 28 percent of total college enrollment value added and 31 percent of total 4-year college enrollment. Test score persistence accounts for very little of total variance for 2-year college enrollment. These results indicate that, despite a significant contribution of test score persistence to schools' effects on postsecondary enrollment, more than two-thirds of the variation in school effects on postsecondary enrollment is due to factors orthogonal to schools' effects on test scores.[16]

4 Robustness and Mechanisms

4.1 Correlates with Neighborhood Characteristics

ACS "falsification tests" - don't necessarily have to be in a table (per Scott).

Cite CFR that discusses the interpretation of the falsification test.

Cite papers that correlate school VA with income, etc.

Remove row 3 from table 3? that does the falsification test. Just discuss that in this paragraph.

Write a paragraph of what you did with the Census variables and how to interpret the results. Make the address matching sound like it was a lot of work/innovative.

Scott and Paco write why VA estimates may still be unbiased even with positive and significant coefficient. Better schools may be in better neighborhoods (property taxes, PTA, etc.)

Cite Naven (2020) on the fact that other school/district correlates don't tell much (i.e. district spending). Write a few sentences on what the high school correlates were from Matt's JMP.

4.2 Correlates with Student, Teacher, and Parent Perceptions

Put a table here with the teacher/school survey stuff

[15]Equation (6) follows from the assertion that $\text{cov}(\lambda_{st}, \beta_{st}) = 0$ by definition of λ_{st} and β_{st}. Figure 7 provides empirical evidence that this assertion holds.

[16]We find qualitatively similar results when using methods used in Carrell and West (2010), with test score factors accounting for at most 20-percent of the variation in college-going value added. These results are shown in appendix section A.

5 Conclusion

This paper explores high school quality in California after the transition to Common Core State Standards (CCSS). We use rich data from California that links K-12 student records to information on college enrollment to examine high schools' contribution to test score performance, postsecondary schooling outcomes, and the relationship between the two. We estimate "value-added" models by adapting the procedure Chetty, Friedman and Rockoff (2014) use to estimate teacher effects.

We examine the link between a school's impact on test scores and its effects on post-secondary enrollment to examine the extent to which test score gains "persist" to longer-run outcomes. We also estimate value-added models using postsecondary schooling as the outcome to measure a school's effect on college-going. Finally, we decompose the share of the variation in school impacts on college-going that is "explained" by variation in school-level test score impacts.

Our results suggest that schools make important contributions to both test scores and college enrollment. A one standard deviation increase in estimated school effectiveness is associated with roughly a 0.15 standard deviation increase in test scores. We also find that test score impacts are strongly related to improved college going. An one-standard deviation increase in a school's math test score impact is associated with a 2.4 percentage point increase in four-year college attendance. Nonetheless, much of the variation in school effectiveness as measured by college going is not accounted for by a school's test score impact. Importantly, we show that less than one-third of the total variance in college-going value-added is explainable by a school's contribution to test score gains.

References

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.

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

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

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

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

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

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

Naven, Matthew. 2019. “Human Capital Formation During Childhood and Adolescence: Evidence from School Quality and Postsecondary Success in California.”

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

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

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

A Methods

In section 3 we showed results using the value added with drift methodology as in Chetty, Friedman and Rockoff (2014). In this section we show that the results are robust to using the alternative methodological framework outlined in Carrell and West (2010).

Carrell and West (2010) show that when $\mathbb{E}[\lambda_{st}\xi_{st}] = 0$ and $\mathbb{E}[\xi_{st}\xi_{st'}] = 0$, the covariance between school-level performance residuals within the same school across years yields a consistent estimate of σ^2_λ.[17] When these conditions hold, σ^2_λ is given by the expression in equation (7):

$$\begin{aligned}\mathbb{E}[u_{st}u_{st'}] &= \mathbb{E}\left[\left\{\lambda_{st} + \xi_{st} + \frac{1}{N_{ist}}\sum_{i=1}^{N_{st}}\epsilon_{ist}\right\}\left\{\lambda_{st'} + \xi_{st'} + \frac{1}{N_{st'}}\sum_{i=1}^{N_{st'}}\epsilon_{ist'}\right\}\right] \\ &= \sigma^2_\lambda\end{aligned} \tag{7}$$

The assumptions needed for this to hold amount to assuming that school value-added is uncorrelated with school-by-year common shocks and that the common shocks are not correlated across time.

Row one of table 8 gives the estimated value of σ^2_λ using the methodology in Carrell and West (2010). This equates into a standard deviation of school value added σ_λ of 0.149 and 0.150 for ELA and math respectively. Compared to the values of σ_λ obtained using the methodology in Chetty, Friedman and Rockoff (2014) in section 3.

We can alternatively estimate σ^2_π using the 2SLS methodology employed by Jacob, Lefgren and Sims (2010). These can be obtained when taking the pairwise covariance of performance residuals, ν_{st} and $\nu_{st'}$ within the same school across different years as in Carrell and West (2010)'s NBER working paper version.

It can be shown that when $\mathbb{E}[\lambda_{st}\theta_{st}] = 0$, $\mathbb{E}[\beta_{st}\theta_{st}] = 0$, and $\mathbb{E}[\theta_{st}\theta_{st'}] = 0$

$$\begin{aligned}\mathbb{E}[\nu_{st}\nu_{st'}] &= \mathbb{E}\left[\left\{\rho\lambda_{st} + \beta_{st} + \theta_{st} + \frac{1}{N_{ist}}\sum_{i=1}^{N_{st}}e_{ist}\right\}\left\{\rho\lambda_{st'} + \beta_{st'} + \theta_{st'} + \frac{1}{N_{st'}}\sum_{i=1}^{N_{st'}}e_{ist'}\right\}\right] \\ &= \rho^2\sigma^2_\lambda + \sigma^2_\beta\end{aligned} \tag{8}$$

To estimate ρ, we follow the methodologies of Carrell and West (2010) and Jacob, Lefgren and Sims (2010) who estimate the persistence of teacher value added onto longer-run outcomes. From Carrell and West (2010):

$$\begin{aligned}\mathbb{E}[u_{st}\nu_{st'}] &= \\ &= \rho\sigma^2_\lambda\end{aligned} \tag{9}$$

Table 8 gives the estimates of these values using the methodology in Carrell and West (2010).

[17] For details see the NBER working paper version of Carrell and West (2010). This condition also requires that student-specific shocks are uncorrelated within schools in a given year so that $\mathbb{E}[\epsilon_{ist}|st] = 0$.

Table 8: Carrell and West (2010) Methodology

	Enrolled		Enrolled 2-Year		Enrolled 4-Year	
	ELA	Math	ELA	Math	ELA	Math
σ_λ^2	0.022	0.022	0.022	0.022	0.022	0.022
σ_λ	0.149	0.150	0.149	0.150	0.149	0.150
$\rho^2\sigma_\lambda^2 + \sigma_\beta^2$	0.006	0.006	0.008	0.008	0.007	0.007
σ_β^2	0.006	0.005	0.008	0.008	0.006	0.006
σ_β	0.074	0.072	0.088	0.088	0.080	0.074
$\rho\sigma_\lambda^2$	0.004	0.005	0.000	-0.000	0.004	0.006