More Gains Than Score Gains? High School Quality and College Success

Daniel Hubbard*

January 16, 2018
This is a working paper! The latest version is available at
www.bit.ly/HubbardJMP.

Abstract

Test-score value-added models have become very popular metrics to determine school quality, but they focus solely on how students perform while they attend the school being evaluated, rather than how that school prepares them to succeed after graduation. A narrow focus on improving test scores may crowd out investments in student learning that may have more persistent effects. I measure the test-score value added of all public high schools in Michigan, then match the results onto student transcripts from public colleges in Michigan to determine the relationship between schools' ability to improve student test scores and the college achievement of their alumni. I find that students who attend high schools with higher value added perform better in college, both in tested and untested subjects; a student who attends a high school one standard deviation above the mean level of value added will have first-year grades about 0.09 grade points higher than the grades of an identical student in an average high school. The effect remains positive and highly significant after a variety of adjustments to deal with selection into college and into high school. This result implies that schools with high value added are not earning those scores by teaching to the test or by reallocating resources toward tested subjects, but instead by preparing students effectively to perform well in the standardized test and beyond.

*E-mail: danieldh@umich.edu. Thanks to the Institute of Education Sciences, U.S. Department of Education for providing support through Grant R305E100008 to the University of Michigan. Thanks to our partners at the Michigan Department of Education (MDE) and Michigan's Center for Educational Performance and Information (CEPI). This research result used data structured and maintained by the Michigan Consortium for Education Research (MCER). MCER data are modified for analysis purposes using rules governed by MCER and are not identical to those data collected and maintained by MDE and CEPI. Results, information and opinions are the author's and are not endorsed by or reflect the views or positions of MDE or CEPI. Special thanks to Charlie Brown, Anne Clark, Sue Dynarski, Julian Hsu, Brian Jacob, Elizabeth Mann, Nicolás Morales, Alexander Persaud, Kevin Stange, Lesley Turner, Gonzalo Vázquez, Pieter de Vlieger, and seminar participants at the University of Michigan for helpful discussions and valuable suggestions.

1 Introduction

Most measures of school quality focus only on how a school's students perform while they attend that school. Accountability measures such as average test scores, school value-added measures, graduation rates, and many other metrics focus on outcomes that occur before (or as) a school's students graduate. While this allows the metrics to focus on the things that the schools being evaluated influence most directly, it also treats success at the given level of education as an end goal rather than a stepping stone. In reality, earning a high score on a standardized test does not, or should not, mean anything on its own. Test scores and other such metrics are valuable as signals of what students have learned and, perhaps more importantly, how much knowledge they have accumulated to support them in further education and in their careers.

Test-score value-added models are the current methodological gold standard, but even these models generally stop at or before graduation. The skills used to perform well on a test may not transfer well into other contexts, and the knowledge accumulated may fade out before students can continue to apply it. Even well-designed exams can have unique types of questions that test students' exam-preparation skills more than their content knowledge. When faced with pressure to have high test scores, either in average scores or in value added, teachers can "teach to the test", drilling students on specific aspects of the exams in lieu of maximizing their content knowledge or providing transferrable skills. Students who are "taught to the test" will score highly on that particular exam, but will not have knowledge of the subject matter that will persist into other contexts.

I measure the persistent value of going to a "good high school", defined here as a high school that raises students' test scores. Particularly, I evaluate the relationship between high schools' contributions to test-score gains and their alumni's achievement in college. In this study, I develop a theoretical model in which schools allocate their resources between test-specific preparation and teaching of content, subject to an endowment budget constraint and accountability measures of varying strictness. Schools with high endowments (generous funding, for instance, or talented students) are unconstrained by the test score requirement and can allocate their resources as they please, while schools with lower endowments may need to allocate resources away from teaching and toward test preparation in a way that they would not if there was no accountability. Some schools may not be able to satisfy high-stakes accountability constraints regardless of their resource allocation, and they go out of business.

I then test the findings using administrative data from public high schools and public colleges in Michigan. I develop value-added scores for each high school in Michigan using test scores in math and reading (the two most highly-emphasized subjects in school accountability), then match high school students to their college transcripts and examine the effects of test-score value added on their grades in their first-year college courses in tested subjects and other subjects. I show that attending a school with high test-score value added predicts higher first-year grades across the board, in both tested and untested subjects, controlling for students' middle-school test scores and an extensive set of covariates. I include a number of adjustments for selection into college and into high school, and the result is robust to all of them. Effects are similar across racial groups, socioeconomic groups, and college settings; the benefits are not limited to more-privileged students or students in four-year colleges.

2 Literature Review

This study fits into two principal strands of literature. The first strand deals with the long-term effects of attending a "good" school on outcomes such as college graduation, earnings, or disciplinary incidents. The second focuses on how schools respond to high-stakes testing and other forms of accountability.

The first strand largely takes advantage of excellent integrated state data systems in states such as Texas and North Carolina. Deming et al. (2014) take advantage of a school choice lottery in Charlotte, finding that students who attend their first-choice (and presumably higher-quality) school are more likely to complete college, with effects concentrated among female students. Deming et al. (2016) find that Texas students who attend schools that raise high-stakes test scores in response to school accountability are more likely to attend and graduate from a four-year college, and their earnings at age 25 are higher. Jennings et al. (2015) find that high school quality explains more of the difference in college attainment than in test scores, and that high school quality can reduce racial gaps in student outcomes but exacerbates income gaps. Jackson (working) uses data from North Carolina to show that teachers can have larger and more-persistent effects on behavior, grades, and on-time completion that surpass their impacts on test scores, particularly for English teachers.

Other analyses in this strand make use of charter-school lotteries and other randomized experiments. Dynarski, Hyman, and Schanzenbach (2013) revisit the Tennessee STAR experiment and find that assignment to a small class increased students' probability of attending and completing college and their probability of studying a high-earning field such as science, engineering, or business; the effects were particularly large among black students, and they were well-predicted by the shorter-term effects on standardized test scores. Angrist et al. (2016) find significant effects of charter attendance on exit exam scores, SAT scores, and AP scores; the effects on college attendance are more modest and mostly involve movement away from community colleges and toward four-year colleges. Dobbie and Fryer (2015) find wide-ranging effects of assignment to the Harlem Children's Zone, ranging from increased test scores to reductions in the probability of teen pregnancy and incarceration. Allensworth et al. (2017) find improvements on academic outcomes from attending higher-performing non-selective schools in Chicago, though the effects do not extend to selective schools.

A number of studies have weighed in on the effects of accountability on teaching practices and student learning. Cohodes (2016) finds reason for optimism in Boston's charter schools, as these schools manage to raise students' test scores without placing disproportionate weight on higher-stakes subjects or common question types. Merseth (2010) views the high-performing Boston charters with more skepticism, noting their students' more modest gains on college entrance exams.

Other schools' test score gains may owe more to behaviors that have less to do with sustainable learning. Jennings and Bearak (2014) find that most of the score gains in several large states come from the most common question types, implying that such questions are particularly emphasized in the test preparation as well as in the testing. Jacob (2005) examines a new accountability policy in the Chicago public school system and finds that schools that raise high-stakes exam scores often do not raise low-stakes exam scores, as the schools focus heavily on test preparation, retention of underperforming students, and careful selection of the set of students to be tested. McNeil and Valenzuela (2000) find an almost single-minded focus on test score gains in Texas schools, crowding out many other valuable school functions. Both Jacob (2005) and Neal and Schanzenbach (2010) note the reallocation of resources toward students who are on the margin between passing and failing high-stakes exams in the Chicago public schools. Ahn (2016) and Muralidharan and Sundararaman (2011) outline valuable theoretical models of behavior under accountability; Ahn (2016) focuses on school-level investments, proposing that schools invest in test preparation when they are in danger of sanction but this detracts from teaching, while Muralidharan and Sundararaman (2011) focus on teacher merit pay and show that teachers may similarly be tempted to move away from curriculum teaching and toward test preparation when their bonuses depend on test scores.

3 Theoretical Framework

3.1 Setup

Schools exist for the purpose of helping students learn, and in the absence of any testing, schools would expend as much effort as they saw fit on student learning. However, once low-stakes testing is put into place, if test scores are less than perfectly correlated with student learning, schools may adjust their practices to maximize some function of test scores and student learning. This does not impose any constraints; it merely adds another variable to the schools' objective functions. In practical terms, even if there is no formal accountability system in place, families may still be hesitant to send children to a school with low standardized test scores, giving schools an incentive to consider them in their resource allocation decisions. In turn, this incentive may induce schools to increase their effort.

High-stakes accountability imposes formal penalties for low performance. A simple high-stakes accountability system states that a school's average test score must be above some threshold score θ, or else the school will be closed. High-stakes accountability may also induce a further increase in effort above the level under low-stakes accountability.

Let each school j consume two types of resources: short-term resources S, which only affect standardized test scores T, and long-term resources L, which simultaneously[1] affect both test scores and student learning G. Short-term resources are more effective in producing test scores than are long-term resources. In other words, $T_j = f(S_j, L_j); G_j = g(L_j); \frac{\partial T}{\partial S} > \frac{\partial T}{\partial L} > 0$.

Schools get some utility[2] U_j from student learning and, if there is a testing regime in place, test scores. However, they gain no utility if they are shut down because they do not meet the high-stakes accountability threshold. We can phrase this as $U_j = p(G_j)$ under no accountability, $U_j^* = p(T_j, G_j)$ under low-stakes testing, and $U_j^{**} = \mathbf{1}(T_j \geq \theta)p(T_j, G_j)$ under high-stakes testing. Under any accountability regime, $\frac{\partial U_j}{\partial T_j} > 0, \frac{\partial U_j}{\partial G_j} > 0$.

The resources have costs c_S and c_L, respectively. Schools' effort E is an increasing function of the strictness of accountability A; for convenience, I normalize $E(0) = 1$, to phrase all effort levels as relative to the baseline of no accountability. Each school has an endowment Ω_j; schools face the effort-weighted budget constraint $\Omega_j E(A) = c_S S_j + c_L L_j$. To emphasize the shortcut nature of teaching to the test, let $c_S < c_L$. High-stakes accountability is more stringent than low-stakes accountability, which in turn is stricter than no accountability; $A^{**} > A^* > 0$.

3.2 An Example With Cobb-Douglas and Linear Functions

Let both the utility function and the production functions be Cobb-Douglas, and let the effort function be linear, as follows. All j subscripts are removed for ease of reading.

[1]This is extremely important. While most traditional production equations require resources to be allocated toward producing one good or another, if a school invests in long-term resources, those resources increase the output of student learning and test scores at the same time. This models the true usefulness of those resources more effectively, in addition to making the model much easier to solve.

[2]I am modeling schools as firms here, but I use "utility" in place of "profit" in order to prevent confusion related to for-profit and non-profit schools. The schools in this model do not have a financial motive.

$$U = T^{\alpha} G^{\beta} \tag{1}$$

$$T = K S^{\gamma} L^{\delta}; \gamma > \delta \tag{2}$$

$$G = B L^{\zeta} \tag{3}$$

$$E = 1 + A \tag{4}$$

K, B, and all lower-case Greek letters are non-negative constants. Different schools may have different values of these parameters.

I begin with the case of no accountability, in which $A = 0$ and thus $E = 1$. The school spends its entire endowment on long-term resources:

$$L_0 = \frac{\Omega}{c_L} \tag{5}$$

Under low-stakes testing $A^* > 0$, the school's effort increases to $1 + A^*$. Substitute the production functions into the profit function in order to create a utility function over the consumption of the teaching resources. I remove j subscripts for readability.

$$U = (K S^{\gamma} L^{\delta})^{\alpha} (B L^{\zeta})^{\beta} \tag{6}$$

Rearrange and collect terms, and define a new constant $D \equiv K^{\alpha} B^{\beta}$.

$$U = D S^{\alpha\gamma} L^{\alpha\delta + \beta\zeta} \tag{7}$$

The marginal rate of substitution between S and L is:

$$MRS_{S,L} = \frac{\alpha\gamma L}{(\alpha\delta + \beta\zeta) S} \tag{8}$$

Set this equal to the price ratio $\frac{c_S}{c_L}$, substitute in from the budget equation, and get:

$$S^* = \frac{\Omega(1+A^*)}{c_S}\left(\frac{\alpha\gamma}{\alpha\gamma+\alpha\delta+\beta\zeta}\right) \tag{9}$$

$$L^* = \frac{\Omega(1+A^*)}{c_L}\left(\frac{\alpha\delta+\beta\zeta}{\alpha\gamma+\alpha\delta+\beta\zeta}\right) \tag{10}$$

$L^* > L_0$ if $1 + A^* > \frac{\alpha\gamma+\alpha\delta+\beta\zeta}{\alpha\delta+\beta\zeta}$; in other words, long-term resource consumption (and, by extension, learning) is higher under low-stakes testing than under no testing if the increase in effort is greater than the relative importance of long-term resources in the school's utility function.

Moving to a high-stakes testing regime $A^{**} > A^*$ is unequivocally beneficial in schools that meet the threshold score with their unconstrained optimal resource bundle; the math is the same except A^{**} replaces A^* in the respective formulas. Even if the school would not have met the threshold at its low-stakes optimum, if the additional effort induced by the high-stakes testing puts the school over the threshold, there will be a commensurate increase in long-term resource consumption and therefore student learning.

The problem is not analytically tractable under high-stakes testing if the unconstrained optimum bundle does not meet the constraint; the school will alter its its resource bundle so that the constraint is just satisfied and consume the solution to the following system:

$$KS^\gamma L^\delta = \theta \tag{11}$$

$$\Omega(1+A^{**}) = c_S S + c_L L \tag{12}$$

In lieu of determining the exact levels of S and L that a school will consume under high-stakes accountability in order to meet the threshold, I show that schools will respond to falling short of the threshold score by moving toward short-term resources if $\frac{\alpha\delta+\beta\zeta}{c_L} > \frac{\alpha\delta}{c_S}$. If the costs of the short-term and long-term resources are equal, it is always more profitable to move toward short-term resource use from the unconstrained optimum. Details are in the Mathematical Appendix.

Some schools have such low endowments that they cannot meet the constraint; there does not exist a pair (S, L) in their budget set such that $KS^\gamma L^\delta \geq \theta$. In this case, the school shuts down.

3.3 Interpretation

The main implication of this model is that test score gains under high-stakes accountability are more likely to reflect increased content knowledge in wealthier schools or schools with higher-performing students, while schools without these luxuries may make more use of test preparation methods to improve scores. As such, these schools with greater endowments will have a stronger relationship between test score improvements and long-term learning than their poorer counterparts. Schools with high endowments are more likely to be able to choose their utility-maximizing levels of resources without being bound by the constraint. Because the constraint requires only a certain level of test scores, constrained schools will be forced to sacrifice student learning to meet the test score minimum, and will often do this through trading long-term resources for short-term ones. The endowment can be thought of as the school's financial budget, but this is not the only interpretation; a school could also have a high endowment because its teachers are effective or its students are talented. If its teachers and/or its students are especially skilled, a school does not need to devote many resources to test preparation and can focus as much as it wishes on student learning[3].

In the Cobb-Douglas example, schools will consume more short-term resources S if their endowment Ω increases, the cost of short-term resources c_S decreases, the productivity of short-term resources γ increases, or the importance of test scores in the utility function α increases. Schools will consume more long-term resources L if Ω increases, the cost of long-term resources c_L decreases, the productivity of long-term resources in producing test scores δ increases, the productivity of long-term resources in student learning ζ increases, or the importance of student learning in the utility function β increases.

I am intentionally agnostic about the proper values of α and β. Some schools may be philosophically opposed to testing and have an α of 0; that is, test scores have no role in their utility function other than through the constraint. Others may place heavy weight on test scores and have very high values of α; one could imagine a for-profit charter school in a poor and densely-populated area, for instance. There are many schools to choose from in the area, but they are generally seen as being low-quality. The charter must attract enough students to make a profit, and the easiest signifier of quality when the competing schools are low-performing is higher test scores.

Note that the threshold score θ does not enter into the expressions for the unconstrained maximizing resource consumption bundle. If the threshold is met by the actions that the school would take anyway, the numeric value of the threshold score does not matter. However, when the constraint binds, θ does enter into the expressions for the maximizing bundle.

[3]Alternatively, these qualities could be seen as lowering the costs of the resources; better teachers can provide content knowledge more easily, for instance.

Schools must adjust their consumption and consume the right amounts of resources to just meet the threshold.

This model assumes a single representative student in a school; it does not account for how resources could be targeted within a school. In reality, of course, schools are composed of wide varieties of students with abilities all over the distribution. If schools have an idea of how their students would perform on a standardized test at the moment, they can allocate their resources among the students in a more targeted fashion. Under an accountability regime in which schools are rated based on the fraction of students exceeding a proficiency threshold, for instance, schools may apply short-term resources to students near the threshold to ensure that they pass the test, while students in the upper part of the distribution may receive long-term resources and students at the very bottom might receive nothing at all. A regime like the one outlined here, in which schools are rated based on average test scores, might leave more room for students in most of the distribution to receive short-term resources, although if one assumes the returns to short-term resources are low for students near the top of the distribution (both because they cannot score much higher and because the last few concepts are the most difficult), the highest-ranking students are still less likely than their lower-achieving peers to receive short-term resources.

The Cobb-Douglas functional form is a convenient illustration, but some of the most basic conclusions hold when the functional form assumption is relaxed. Specifically, as the cost of a resource decreases, schools will demand more of it; as endowments increase, schools will demand more of both resources. As schools place more emphasis on test scores (via increasing $\frac{\partial \Pi_j}{\partial T_j}$), their demand for short-term resources will go up. Effects of some other parameters are more ambiguous.

4 Data and Methodology

4.1 K-12 Data

This project draws from several different administrative data sources. I begin by using student-year level test scores and demographic data from public middle and high schools in Michigan to estimate school value added, then merge the K-12 data with several sources of college data to create the variables necessary to measure postsecondary outcomes.

My base sample is all students in Michigan public[4] schools, who first sit for the 8$^{\text{th}}$-grade math and reading Michigan Educational Assessment Program ("MEAP") test between the

[4]For the purposes of this paper, "public" schools include both traditional public schools run by local school districts and charter schools.

2005-06 and 2007-08 school years. Students must take the 11$^{\text{th}}$-grade standardized exam (which includes the ACT) to be included in deriving the value-added model; other outcomes do not condition on having any data past eighth grade. Students who take the 8$^{\text{th}}$-grade MI-Access exam for special-education students are dropped from both samples.

I merge in some student characteristics (race, gender, age, limited English proficiency, economic disadvantage[5], special education status) measured each year, and the student's home district and the school that the student attends measured three times per year[6], as well as the student's ZIP code and Census block group. The latter two variables allow me to merge in neighborhood household income from the American Community Survey (along with a missing indicator if such data are unavailable for the given student).

Because students may change schools within an academic year, I briefly reshape the data to the student-collection period level, so that I can determine the fraction of periods between the middle-school exam and the high-school exam that a student attends each school. I use this fraction to assign each student to the school that the student attends for the most collection periods; students who do not attend any school for four collection periods or more are not used to derive the VAM.[7] I keep the values of economic disadvantage, limited English proficiency, and special education enrollment from the student's eighth-grade year as to avoid any manipulation by their high schools. I bring in a few school-level aggregate variables, keeping the most common values for indicator variables and the means for continuous variables. Finally, I reduce the sample to one observation per student, keeping the student's first home district and Census block group while attending their longest-tenured school.

4.2 Value-Added Estimation

To calculate schools' value added, I follow the procedure outlined in Chetty, Friedman, and Rockoff (2014; henceforth "CFR"). This process starts by regressing students' 11$^{\text{th}}$-grade test scores Y_{ijnt} (an average of math and reading) on their 8$^{\text{th}}$-grade scores $Y_{i,t-3}$ (math, reading, and the interaction of the two), a variety of student, school, and neighborhood demographics[8] (X_i, $\mathbf{\bar{X}_{j,t-1}}$, and $\mathbf{\bar{Z}_n}$, respectively), cohort dummies τ_t and a high school fixed effect σ_j.

[5] During the sample period, economic disadvantage is measured by a student's eligibility for subsidized school lunch.

[6] Following the Center for Educational Performance and Information's terminology, I refer to the three measurement dates per year (once each in the fall, spring, and end of year) as "collection periods".

[7] Previous versions of this paper have weighted the contributions of each student+school combination according to the fraction of time spent in that school; the results do not change. The current specification is simpler to explain.

[8] See Appendix B.1 for the full list.

(13) $$Y_{ijnt} = \lambda + \psi_1 Y_{i,t-3} + \mathbf{\Psi_2 X_i} + \mathbf{\Psi_3 \bar{X}_{j,t-1}} + \mathbf{\Psi_4 \bar{Z}_n} + \tau_t + \sigma_j + \epsilon_{ijnt}$$

I take a "residualized score" for each student, consisting of the school fixed effect and the error term, and collapse the data to leave one observation per school-by-year combination, keeping an average residualized score ρ_{jt} for school j in year t.

(14) $$\rho_{jt} = \frac{1}{N_{jt}} \sum_{i=1}^{N_{jt}} \sigma_j + \epsilon_{ijnt}$$

I then regress the average residualized score on the same school's average residualized scores from each of the preceding and following two years, plus the relevant missing indicators. The predicted value $\hat{\rho}_{jt}$ from this regression is the value added for the given school in the given year.

(15) $$\rho_{jt} = \sum_{y=-2, y\neq 0}^{2} \kappa_y \rho_{j,t+y} + \xi_y Missing_{j,t+y} + \o_{jt}$$

The main advantage of the CFR model is that it is robust to noise, measurement error, and cohort-specific shocks, through its Bayesian shrinkage "leave-one-out" framework. Results do not change significantly if I use a Bayesian shrinkage estimate of the school value added without the leave-one-out specification (see Koedel, Mihaly, and Rockoff 2015; Herrmann, Walsh, and Isenberg 2016), a simpler one-step VAM, or a two-step model (as in Ehlert et al. 2014).

Figure 1 presents the distribution of value added across schools, weighted by the number of students. Most estimates are between -0.3 and 0.3 standard deviations; there are fewer positive estimates than negative estimates. Figure 2 repeats the exercise but with one observation per school instead of per student; there is more dispersion in these estimates, but the outlier schools tend to be small. A school that is one standard deviation better than average improves students' test scores by 0.234 student-level standard deviations over their 8^{th}-grade baseline scores, corresponding to about 1.1 points out of 36 on the ACT composite.

Table 1 provides detail about how students are distributed across levels of school value added. Even as the measure focuses on student improvement rather than raw scores, the students in higher-value added schools are more privileged and higher achieving than their counterparts in lower-value added schools. 58% of students in these schools are economically disadvantaged, and 35% are black; these figures are 19% and 13%, respectively, in the highest-quartile schools. The average 8^{th}-grade exam score in the lowest-quartile schools is 0.341 standard deviations below the statewide mean, while the average in the highest-quartile schools is 0.384 standard deviations above the statewide mean. I also include several intermediate outcome measures separated by school VAM quartile; students in schools with higher value added perform better on their 11^{th}-grade exams and are more likely to graduate from high school and enroll in college.

4.3 College Data

Most of the outcome data used in this paper come from a data set called STARR. The STARR data consist of student-course level records for all public colleges in Michigan, starting with students who attended college in 2009. Each student would have a separate observation for each course that the student has taken at a Michigan public college (including both community colleges and four-year colleges), containing information about the student, the course, and the student's grade in the course. Unless stated otherwise, grades are expressed on a 4.0 scale (3.7 for an A-, 3.3 for a B+, etc.) in this study.

I keep only credit-bearing courses from a student's first year in a Michigan public college[9]. I drop courses titled "Departmental Credit" (which tend to be credits for Advanced Placement or International Baccalaureate scores rather than for college coursework), drop observations from students who take the subject in question at multiple institutions in the same year, and restrict math and English courses[10] to be the first course taken in the given subject; if a student takes multiple math courses or multiple English courses at once in the student's first semester taking a course in that subject, the course with the lower course number is kept (for instance, "MATH 215" over "MATH 217"). I keep all first-year courses in subjects other than math and English.

The final data preparation step is to merge the data sets together. I merge the value-added measures onto the 8^{th}-11^{th} grade student observations, and then merge the resulting file into the college data. What remains is one observation per student-course combination, containing the student's demographics, high-school value added, and course performance. In order to reduce the impact of events that happen between high school and college, college observations are dropped if a student does not start college "on time", meaning five years after taking the 8^{th}-grade exam.

In order to have at least some information about students who do not attend college at a

[9] Because STARR only contains data from Michigan public colleges, students technically could have taken courses at other colleges first, but this is unlikely given later restrictions on time since graduation.

[10] I identify math and English courses using the course codes listed in Appendix B.2.

Michigan public institution, I also merge in enrollment and graduation information from the National Student Clearinghouse. These data are available for colleges attended by about 90% of Michigan public-school students. The focus of the paper is on the course outcomes, but examining the effects of high school quality on other outcomes such as college attendance and completion informs the calculations made to ensure that the course grade results are robust.

Table 2 outlines the changes in college attendance across the distribution of test-score value added. While every quartile of schools in the distribution sends around 60% or more of its students to college, students who attend schools with higher value added are more likely to attend college, a gain driven mostly by increased probability of attending a public four-year college in Michigan. The fraction of students attending a private college in Michigan and the fraction attending a community college is higher in the middle of the distribution than at the ends but stays in a fairly narrow band. Of particular concern for the identification of this paper, however, is the steady increase in the probability of being in the college grade sample as high school VAM increases. I explore this more in Section 4.5.

4.4 Empirical Specification

To determine the effect of high-school value added on college performance, I build up to the following specification:

$$Grade_{cijknt} = \iota + \phi_1 Y_{i,t-4} + \phi_2 ValueAdded_{j,t-1} + \mathbf{\Phi_3 X_i} + \mathbf{\Phi_4 \bar{X}_{j,t-1}} + \mathbf{\Phi_5 \bar{Z}_n} + \chi_{ckt} + \nu_{ijkct} \quad (16)$$

The course grade earned by student i from neighborhood n, who attended high school j before enrolling at college k and taking course c in semester t, is a function of the student's middle-school test score $Y_{i,t-4i}$; student characteristics $\mathbf{X_i}$; school-level average characteristics $\mathbf{\bar{X}_{j,t-1}}$; neighborhood characteristics $\mathbf{\bar{Z}_n}$; a course fixed effect χ_{ckt}; and the high-school[11] value added.

Empirically, the course grade is measured on a 4.0 scale, and the student characteristics are the same ones used in the value-added model (except with the 8^{th}-grade score included as an average of math and reading, up to a fourth-order polynomial, as opposed to separating them and including an interaction). I scale the value added in terms of its school-level standard deviation; the coefficient ϕ_2 represents the effect of raising a school's test-score value added from the statewide average to one standard deviation above it. Standard errors are

[11] I use this term for convenience to describe the school that a student attends most frequently between the exam taken in grade 8 and the exam taken in grade 11. It could be a junior high school, which has grades 7-9; it could be a school that contains grades K-12. It does not need to only contain grades 9-12.

clustered by school; this accounts for serial correlation and is generally more conservative than clustering by school and year. To account for the generated regressor in the value-added term, I present bootstrapped standard errors in the full specifications, following Bastian (working), among others. The fixed effects χ_{ckt} are for each combination of college, year, semester, subject code, and course number (for instance, University of Michigan-Ann Arbor, fall 2011, ENGLISH 125). Only students who take a course in a Michigan public college within five years of the year in which they take their 8$^{\text{th}}$-grade test are included. Observations in the "all subjects" and "other subjects" specifications are weighted by their fraction of the student's relevant credits in the non-bootstrapped specifications.

The specifications for outcomes other than college course grades are similar, but with a few important modifications. There is one observation per student, rather than one observation per student-course combination; there is no course fixed effect, because there is no course being measured; and the sample no longer consists only of students who enroll in college, or even students who take the 11$^{\text{th}}$-grade exam. Instead, anyone who has an 8$^{\text{th}}$-grade exam score that is not from the MI-Access special education exam is included in the sample.

4.5 Threats to Identification

Potential biases lurk throughout the empirical analysis process. First, the 8$^{\text{th}}$-grade exam scores may be measured with error, stemming from anything from poorly-filled bubbles to a malfunctioning Scantron machine. If the measurement error is classical, this would introduce attenuation bias into the value-added estimates, and in turn the coefficients in the outcome regressions would be biased upward as they measure the effect of an attenuated regressor. The measurement error is not precisely classical, as scores are necessarily bounded between 0% and 100%, but as 0.02% of students receive the minimum or maximum score on the math exam and 0.01% receive the minimum or maximum on the reading exam, the bounds are so rarely reached that I treat the measurement error as classical.

The next threat comes from selection into high schools. One of the most important issues in the value-added literature is that attendance in "better" schools is not random. Even though they cannot observe value added explicitly at the time that students enroll, schools' reputation for quality is presumably at least somewhat positively correlated with value added. Students then sort across districts across two dimensions that reinforce each other. Families sort across home districts based on preferences for education and available resources, among other things; students can then, conditioning on where they live, take advantage of policies that allow them to attend schools in other districts or in specially-designed settings such as charter schools and magnet schools. In both cases, students best equipped to succeed in college will be sorting into higher-quality high schools, biasing my estimates upward.

I present a handful of falsification tests in Table 4 to quantify the degree of the sorting. If there was no sorting, then high-school value added should not predict 7^{th}-grade test scores, 7^{th}-grade attendance, Census block group poverty rates, or Census block group education levels. However, all of those variables except for poverty rates are indeed predicted by high-school VAM. For instance, students who attend high schools with one standard deviation higher value added measures have 7^{th}-grade scores that are about 0.043 standard deviations higher, even after controlling for 8^{th}-grade scores and the other typical covariates. This is an economically modest but statistically significant bias.

Finally, students in higher-VAM schools are more likely to attend college, most notably at the in-state public institutions that collect the transcript data used in this study, as shown in Table 3. Table 3 contains probit marginal effects for various attainment outcomes: taking the 11^{th}-grade state exam, graduating high school, attending college, and being in the sample for the course grade specifications. All of these are predicted very well by the test-score value added of a student's high school. The college outcomes track closely with the results shown in Table 2, showing that the unconditional results in Table 2 are not driven solely by covariates that are controlled for in the probits. However, they raise concerns about selection into the college grades sample.

There are two different interpretations of the selection. On one hand, if students are more likely to get into college if they attend a good high school, then perhaps the students who still manage to get there despite attending a low-performing high school must be particularly resilient, which makes them likely to perform better in college. This would argue against finding a positive effect of high-school value added on college performance. However, an alternative interpretation is that students are sorted into high schools by some unobserved quality; this quality makes the students perform better in high school, raising their value added, and in college, raising their grades, but not due to anything that their high schools contributed. This would bias toward finding a positive effect.

5 Course Results

5.1 Main Specifications

Table 5 presents results for the full sample, adding more covariates with each column. The specification in column 5 weights the contribution of each observation by its fraction of the student's total credits, but does not bootstrap the standard errors; column 6 includes bootstrapping but weights 1-credit classes equally to 4-credit classes. Regardless of specification, there is a positive and significant relationship between high-school value added and college course grades. The effect of attending a school with one standard deviation higher VAM is about 0.089 grade points, almost one third of the difference between a B and a B+. This result provides evidence against the hypotheses that schools are teaching to the test; there appear to be long-term benefits to attending a school that raises high-stakes test scores. The effect size does fall short of the 0.1-standard deviation threshold that denotes a large impact in the education literature, however; the standard deviation of course grades is about 1.34 grade points in this sample, meaning that the effect size is closer to 0.07 standard deviations.

Table 6 presents results by subject, using the specification from column 6 of Table 5; the effect is very similar. Controlling for a full set of demographics and bootstrapping standard errors, the effect of going to a school with one standard deviation higher VAM is about 0.122 grade points in math, positive and highly significant. For English, the effect sizes are slightly smaller at about 0.065 grade points, still highly significant when including full controls and bootstrapped standard errors. In subjects other than math and English, the impact of attending a one-standard deviation more-effective school is about 0.087 grade points. This is evidence against the hypothesis that schools are focusing only on subjects that have high-stakes tests, at the expense of overall skills and learning in fields not subject to test score-based accountability. The courses in this category include anything from psychology to business to welding; none of these subjects are included in accountability measures and many of them are not even taught in high schools.[12]

5.2 Accounting For Biases

I begin by accounting for the attenuation bias in the value-added model. Although students' 7^{th}-grade test scores are likely measured with a similar type of error as their 8^{th}-grade scores, if these measurement errors are not correlated with each other, one can be used as an instrument for the other. In this case, I use 7^{th}-grade test scores to instrument for their 8^{th}-grade counterparts in the first step of the VAM. I then proceed with the model as normal, using the residualized scores from the instrumented first step to constructed the predictions used in the value added, and then using the resulting value-added measures as regressors in the outcome specifications.

To reduce the bias from selection across high schools, I construct a "home-district" sample. This sample consists only of students who attend a non-charter, non-magnet high school in their zoned school district. These students often, though not always if their district offers

[12]One reason why there may not be reallocation toward high-stakes subjects in high school is the Michigan Merit Curriculum ("MMC"), which requires that students take biology, either chemistry or physics, three social studies courses, two years of a foreign language, gym, art, and an online course (in addition to its math and English requirements). The first cohort exposed to the MMC entered high school in the fall of 2007; some cohorts in this study preceded the MMC, while others were exposed to it. (Jacob et al. 2017)

multiple high schools, attend the "default" high school associated with their home address. This reduces the upward bias from selection into higher-quality high schools, although it does not eliminate it, as families also sort residentially by income, education, and preferences for school quality. Appendix Table 12 presents the results from Table 4 for this sample; it remains selected on some observable characteristics, even if the home-district sample deals with selection on some unobservable characteristics such as motivation.

I conduct three different adjustments for selection into college. First, I present results estimated for a smaller sample of students who are very likely to go to college. These students have less margin to have their college-going decisions altered by the quality of their high school. To construct this subsample, I take advantage of ACT's cutoff score for college readiness: a student who meets all of ACT's benchmarks for college preparation would have a composite score of 21 (ACT, Inc.). 31% of students in the sample received an ACT composite of 21 or higher. Because ACT scores are, by construction, influenced by high school quality, I instead include the top 31% of scorers on the 8^{th}-grade standardized test. 86% of these students in the lowest-VAM schools attended college; 92.4% of these students in the highest-quartile schools attended college. Other than in the very lowest quartile, there does not seem to be a significant unconditional relationship between high-school value added and college attendance for these "college-ready" students. A relationship remains when I condition on the usual set of covariates, but it is notably smaller than the equivalent relationship in the full sample, and there is no significant relationship between high-school value added and presence in the grade sample for the college-ready students, as shown in Appendix Table 11

Second, Oster (2016) derives a method of accounting for selection on unobservables that incorporates the relative changes in treatment effects and R^2 values as covariates are added to the model, building on work by Altonji, Elder, and Taber (working). To construct this, I start by regressing college grades on high-school value added with no other covariates, saving the R^2 value R_0 and the treatment effect ϕ_0. I then run the fully-specified outcome model, again saving the R^2 value R_{full} and the coefficient on value added ϕ_{full}. Oster (2016) imposes that a specification with all possible controls, observable and unobservable, would explain 30% more of the variation than the model with the full set of covariates[13]; therefore, I let $R_{max} \equiv 1.3R_{full}$. The formula for the bias-adjusted coefficient ϕ^* is:

$$\phi^* = \phi_{full} - (\phi_0 - \phi_{full})\left(\frac{R_{max} - R_{full}}{R_{full} - R_0}\right) \tag{17}$$

[13]This parameter is chosen to allow the results of 90% of randomized experiments to be upheld in the bias-adjusted framework.

The bias-adjusted coefficients that result can provide a bound on the effect size; it is a good sign if the bias-adjusted coefficients are not statistically different from their unadjusted counterparts, while if the bias-adjusted coefficient is actually larger than its unadjusted counterpart, then the bias is skewing the effect toward zero and is of relatively less concern. The fundamental assumption behind the Oster (2016) formulation is that the unobservables bias the result in the same direction that the observables do; if the sample is positively selected on observables, then it must be positively selected on unobservables, and the opposite.

Third, I present results in which I impute grades for students who are not in the grade sample. I geocode all public colleges in the state, and assign all students who did not attend college to their nearest community college, while assigning all students who attended private or out-of-state colleges, or did not take any large enough credit-bearing courses, to their nearest four-year public college. After "placing" each non-attendee in a college, I then assign the non-attendees to courses: each of these students is assigned to a math class, an English class, and one other class in their respective college in the appropriate year. The courses are chosen randomly, with the probabilities of course placement equal to the observed fractions of students enrolled in that course. In other words, if 50% of freshmen enrolled in a math class at Washtenaw Community College in 2011 take college algebra, 25% take calculus, and 25% take statistics, a non-attendee assigned to Washtenaw Community College would have a 50% probability of being placed in college algebra, 25% probability of being placed in calculus, and a 25% probability of being placed in statistics.

I assign grades to these students in two ways. First, as a bounding exercise, I assign 0.0 GPAs to all students in the counterfactual sample. Second, to obtain a more realistic effect, I use out-of-sample prediction to give grades to these students. In the sample of students who have course grades available, I regress their course grades on the full set of observables, except for the value added of their high schools. I then use these estimates to impute the grades of the students who do not actually attend public colleges in Michigan. For instance, a student with a high 8^{th}-grade test score taking pre-algebra would likely receive a high grade, while a student with a low 8^{th}-grade score taking calculus would likely receive a low grade.

5.3 Robustness Results

Table 7 presents estimated coefficients on high-school value added, separated by subject, for each of the bias adjustments outlined in Section 5.2. While some of those adjustments shrink the point estimates, almost none of them alter the statistical significance of the results. The estimates are especially robust to the attenuation bias in the VAM and to across-district school choice mechanisms; neither the instrumented VAM nor the home-district sample changes the results meaningfully in any subject. The selection into college makes a slightly larger difference, but all of the corrections for this selection leave a positive and significant effect of at least 0.04 grade points in the all-subjects sample.

Even after all of these corrections, there remains residential sorting across districts. To place an upper bound on the effect of the sorting, I follow the methodology in Altonji and Mansfield (working). This involves estimating the contribution of school characteristics to outcomes, controlling for student-level and aggregated student characteristics. In place of Equation 13, I run a slightly-altered first-step regression, in which I replace the school fixed effects with school characteristics $\mathbf{W_{j,t-1}}$ that are not student aggregates: per-pupil expenditure (at the school and district level), pupil-teacher ratio, the fraction of students who transfer or drop out, and the average certification exam score of the school's teachers.

$$Y_{ijmnt} = \mu + \pi_1 Y_{im} + \mathbf{\Pi_2 X_i} + \mathbf{\Pi_3 \bar{X}_{j,t-1}} + \mathbf{\Pi_4 \bar{Z}_n} + \mathbf{\Pi_5 W_{j,t-1}} + \upsilon_t + \eta_{ijmnt} \tag{18}$$

I then find the variance of the contribution of those school observables $\mathbf{\Pi_5 W_j}$ and compare it to the variance of the VAM estimates from Section 4.2. The ratio of these variances $\frac{\text{var}(\mathbf{\Pi_5 W_j})}{\text{var}(ValueAdded_j t)}$ represents the relative contribution of school observables to school value added, which places a lower bound on the contribution of the school itself to value added and an upper bound on the contribution of students' sorting across schools and districts to value added. I multiply my regression coefficients by this ratio to determine the minimum effect size that can be attributed definitively to school quality and not to sorting.

These results are shown in Table 7 as well. I find that this lower bound estimate teeters along the edge of statistical significance, while its economic significance is somewhat limited, as a vast change in school effectiveness is met with a fairly modest change in college performance. This does cast some doubt upon my overall results if taken literally. However, it is an intentionally-conservative lower bound backed by a fairly prohibitive assumption: that all of schools' contribution to value added can be contained by observable variables such as per-pupil expenditure, teachers' certification scores, and student turnover. This leaves no room for schools to benefit students by teachers' creativity and hard work or by creating a positive learning environment, reducing the secret to student learning to a formula. In reality, this assumption is too punitive to be realistic; some of the effect can likely be attributed to students' sorting across districts, but there is room for schools to help students learn in ways that cannot be captured in a spreadsheet.

5.4 Results by Subgroup

One interpretation of the model in section 3 is that schools whose students have weaker academic backgrounds may be forced to focus on test preparation in order to meet high-stakes accountability standards, whereas schools with better-prepared students can focus on content. This would imply a stronger relationship between test-score value added and college performance in the latter schools, as they provide learning that students can build upon while the test preparation emphasized in the former schools has less application in other environments. I test this in Table 8, running the final specification from Table 5 separately by quartile of high schools' average 8^{th}-grade exam scores. The results do not support this hypothesis. The largest point estimates are found in the schools in the bottom quartile; students gain more from attending schools that raise scores from very low to somewhat low than from attending schools that raise scores from high to very high.

Brand and Xie (2010) find that students who come from groups less likely to complete college, such as black students and lower-income students, benefit most from attending college. This result could possibly extend to attending a good high school, but an opposing argument is also certainly plausible: that students from disadvantaged groups are underrepresented in higher-VAM schools (as shown in Table 1), and thus the schools are not designed to serve them, or they are socially excluded in these schools; either of these may lead to such students receiving fewer benefits from attending a "better" school.

I test this by estimating college grades results separately for black and white students, and separately for students who were economically disadvantaged and those who were not, presented in Table 9. Effect sizes are somewhat larger for black students than for white students; effects are also slightly larger for poor students than for non-poor students, but not significantly so. This result is reassuring, as it allays the fear that marginalized students are being left behind in high-VAM schools while white and wealthier students reap the benefits. High school VAM alone cannot close the state's gaps in college achievement, but it may be a part of the solution.

Finally, the ACT tests academic subjects and is used for admission at four-year colleges; schools that raise ACT scores may not be giving students the tools needed for success at community colleges as effectively, particularly in courses outside of the traditional academic fields. These schools may be focusing on their students who are on the four-year academic track at others' expense. I test this by running the college grades results separately at two-year and four-year institutions, as shown in Table 10. If anything, I find the opposite. The effect sizes are positive and significant in both two-year and four-year institutions; the returns to attending a high-quality high school are actually marginally bigger in two-year colleges.

6 Discussion and Conclusion

6.1 Policy Context

The students studied in this paper, if they complete their secondary education on time, are members of the high school classes of 2010, 2011, and 2012. During this period, Michigan public schools were subject to the accountability measures in the No Child Left Behind Act of 2001 ("NCLB"). High schools were judged on whether they met Adequate Yearly Progress in math and reading proficiency rates, as well as graduation rates, both for the student body as a whole and for subgroups of interest such as black students, Hispanic students, and students eligible for subsidized lunch. The necessary proficiency rates to achieve Adequate Yearly Progress grew more stringent year by year, putting schools under pressure to improve rapidly (Bielawski 2006). Additionally, the required 11^{th}-grade state standardized test incorporated the ACT college entrance exam during this time period. The ACT is specifically designed to be predictive of college performance, which implies a stronger relationship between state standardized test scores and college outcomes than one might expect from another standardized test. Additionally, the designers of the ACT state that a composite score of 21 is the benchmark for college readiness; this standard allows easier comparisons within the group of students who are expected to go to college. Finally, the Michigan Merit Curriculum required students during this sample period to take four years of math and English, three years of social studies, two years of a foreign language, and courses in biology, art, music, and either chemistry or physics (Jacob et al. 2017). This resulted in more challenging high-school coursework, and it prevented high schools from dedicating entire days to math and reading in the way that some elementary and middle schools did to prepare for standardized tests (McNeil and Valenzuela 2000).

From a policy perspective, or from the perspective of someone who cares about educational equality in the United States, the empirical results of this paper are encouraging. The model presents a bleak scenario in which schools with low-achieving incoming students cannot teach lasting lessons to students because they need to focus so completely on exam preparation. I do not find empirical evidence to support this claim; if anything, effect sizes are largest in these schools. Furthermore, effect sizes are notably large for black students and for economically disadvantaged students; exposing these students to better schools, either by improving integration of disadvantaged students into more-advantaged schools or by making investments in schools with more-disadvantaged populations, can make a meaningful difference in their long-run success.

6.2 Further Research

While this study makes a meaningful contribution to the literature about the student-level returns to high-school quality and the usefulness of value-added models, room for further investigation remains. For instance, other measures of high-school quality may predict future success and long-term learning more effectively than test-score value added does; replicating the exercises done here, replacing test-score value added with measures related to graduation rates or other quality metrics, and seeing which best predicts long-term learning, would be a useful extension. Additionally, the set of students who attend in-state public colleges immediately after graduation tend to be fairly stable and high-achieving students; looking at the effects of high-school quality on persistence, completion, and grades at for-profit colleges would also be valuable, although the selection concerns might run in the opposite direction from the ones in this paper.

Future researchers and policymakers alike may be interested in learning more about the characteristics of the schools that see high value added and large effects on college performance, particularly those that do so despite having low-performing students at entry. If there are patterns in the classroom practices most prevalent in these schools, or in how they spend money, assign teachers, and allocate other scarce resources, then other schools may mimic these patterns, hoping to achieve similar results. Conversely, there is also a chance that these schools have unique characteristics that other schools cannot replicate to the same effect.

6.3 Summary

I seek to measure the effect of high-school quality on first-year college course grades, in order to determine how much of high schools' value added truly comes in the form of persistent learning and skills. I match school-year level test-score value added measures onto college transcripts, looking separately at the effects on grades in all subjects, tested subjects, and non-tested subjects. Even after numerous adjustments for selection into high schools and colleges, a stable positive effect of high-school quality on college grades remains; students who attended a high school one standard deviation above the school-level average receive first-year grades between 0.04 and 0.1 grade points higher than their otherwise-identical counterparts who graduated from average schools. This implies that much of these high-quality schools' improvements in test scores are driven by durable student learning, a result that should ease some concerns of some skeptics of standardized testing.

References

[1] ACT Inc. (2010). *What Are ACT's College Readiness Benchmarks?*. http://files.eric.ed.gov/fulltext/ED510475.pdf.

[2] Ahn, T. (2016). A Theory of Dynamic Investment in Education in Response to Accountability Pressure. *Economics Letters*, 75-78.

[3] Allensworth, E.M., Moore, P.T., Sartain, L. & de la Torre, M. (2017). The Educational Benefits of Attending Higher Performing Schools: Evidence From Chicago High Schools. *Educational Evaluation and Policy Analysis* 1-23.

[4] Altonji, J.G., Elder, T.E., & Taber, C.R. *An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schools*. Working paper.

[5] Altonji, J.G. & Mansfield, R.K. *Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects*. Working paper.

[6] Angrist, J.D., Cohodes, S.R., Dynarski, S.M., Pathak, P.A., & Walters, C.R. (2016). Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice. *Journal of Labor Economics* 275-318.

[7] Angrist, J.D., Hull, P.D., Pathak, P.A., & Walters, C.R. *Leveraging Lotteries for School Value-Added: Testing and Estimation*. Working paper.

[8] Bastian, J. *The Rise of Working Mothers and the 1975 Earned Income Tax Credit*. Working paper.

[9] Bielawski, P. (2006). Michigan School Accountability: AYP and Education YES! *Michigan Department of Education, Office of Educational Assessment and Accountability.*

[10] Brand, J.E. & Xie, Y. (2010). Who Benefits Most from College? Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education. *American Sociological Review*, 273-302.

[11] Chetty, R., Friedman, J.N., & Rockoff, J.E. (2014). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, 2593-2632.

[12] Cohodes, S.R. (2016). Teaching to the Student: Charter School Effectiveness in Spite of Perverse Incentives. *Education Finance and Policy*, 1-42.

[13] Deming, D.J. (2014). Using School Choice Lotteries to Test Measures of School Effectiveness. *American Economic Review: Papers & Proceedings*, 406-411.

[14] Deming, D.J., Hastings, J.S., Kane, T.J., & Staiger, D.O. (2014). School Choice, School Quality, and Postsecondary Attainment. *American Economic Review*, 991-1013.

[15] Deming, D.J., Cohodes, S., Jennings, J., & Jencks, C. (2016). School Accountability, Postsecondary Attainment, and Earnings. *Review of Economics and Statistics*, 848-862.

[16] Dobbie, W. & Fryer Jr., R.G. (2015). The Medium-Term Impacts of High-Achieving Charter Schools. *Journal of Political Economy* 985-1037.

[17] Dynarski, S., Hyman, J., & Schanzenbach, D.W. (2013). Experimental Evidence on the Effect of Childhood Investments on Postsecondary Attainment and Degree Completion. *Journal of Policy Analysis and Management*, 692-717.

[18] Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M.J. (2014). The Sensitivity of Value-Added Estimates to Specification Adjustments: Evidence From School- and Teacher-Level Models in Missouri. *Statistics and Public Policy*, 19-27.

[19] Hanushek, E.A., & Rivkin, S.G. (2010). Generalizations about Using Value-Added Measures of Teacher Quality. *American Economic Review: Papers & Proceedings*, 267-271.

[20] Herrmann, M., Walsh, E., & Isenberg, D. (2016). Shrinkage of Value-Added Estimates and Characteristics of Students with Hard-to-Predict Achievement Levels. *Statistics and Public Policy*, 1-10.

[21] Jackson, C.K. *Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9^{th} Grade Teachers in North Carolina*. Working paper.

[22] Jacob, B.A. (2005). Accountability, Incentives and Behavior: the Impact of High-Stakes Testing in the Chicago Public Schools. *Journal of Public Economics*, 761-796.

[23] Jacob, B., Dynarski, S., Frank, K., & Schneider, B. (2017). Are Expectations Alone Enough? Estimating the Effect of a Mandatory College-Prep Curriculum in Michigan. *Educational Evaluation and Policy Analysis*, 333-360.

[24] Jennings, J.L. & Bearak, J.M. (2014). "Teaching to the Test" in the NCLB Era: How Test Predictability Affects Our Understanding of Student Performance. *Educational Researcher*, 381-389.

[25] Jennings, J.L., Deming, D., Jencks, C., Lopuch, M., & Schueler, B.E. (2015). Do Differences in School Quality Matter More Than We Thought? New Evidence on Educational Opportunity in the Twenty-first Century. *Sociology of Education*, 56-82.

[26] Master, B., Loeb, S., & Wyckoff, J. (2017). More Than Content: The Persistent Cross-Subject Effects of English Language Arts Teachers Instruction. *Educational Evaluation and Policy Analysis*, 1-19.

[27] McNeil, L. & Valenzuela, A. (2000). The Harmful Impact of the TAAS System of Testing in Texas: Beneath the Accountability Rhetoric. *Educational Resources Information Center.*

[28] Merseth, K.K. (2010). High-Performing Charter Schools: Serving Two Masters? In R.J. Lake, *A Balanced Look at American Charter Schools in 2009* (pp. 27-38). Bothell, WA: National Charter School Research Project.

[29] Muralidharan, K. & Sundararaman, V. (2011). Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy*, 39-77.

[30] Neal, D. & Schanzenbach, D.W. (2010). Left Behind by Design: Proficiency Counts and Test-Based Accountability. *Review of Economics and Statistics*, 263-283.

[31] Oster, E. (2016). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business and Economic Statistics*, 1-18.

Figure 1: Distribution of Value Added, Student Level

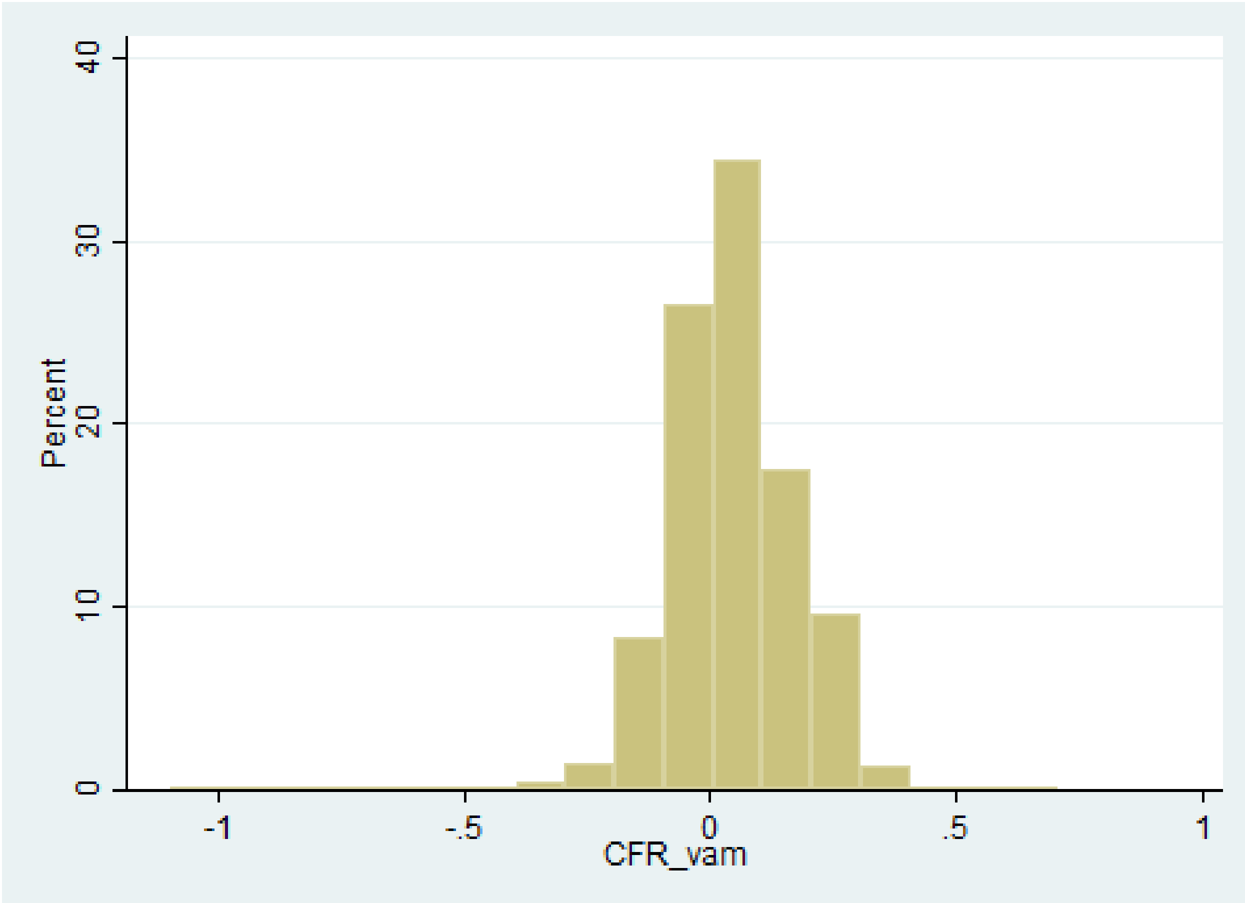

Figure 2: Distribution of Value Added, School Level

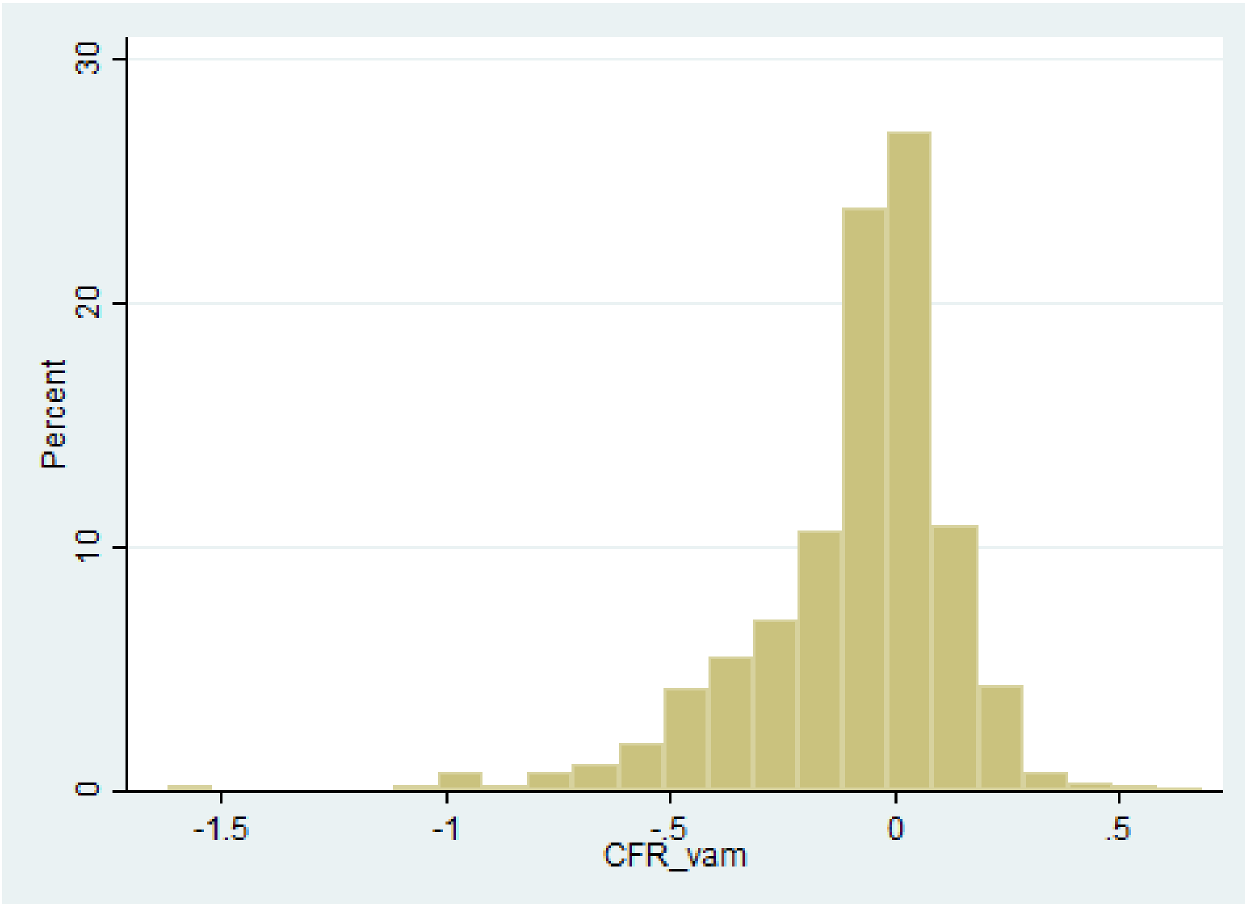

Table 1: Student Characteristics by VAM Quartile

	Lowest	2^{nd}	3^{rd}	Highest
Fraction Black	0.351	0.151	0.1	0.134
Fraction Hispanic	0.054	0.043	0.041	0.029
Fraction Asian	0.011	0.013	0.02	0.046
Fraction in Special Education	0.144	0.129	0.112	0.101
Fraction Limited English Proficiency	0.044	0.025	0.021	0.021
Fraction Economically Disadvantaged	0.578	0.375	0.277	0.194
Fraction in Charter Schools	0.052	0.023	0.022	0.03
Fraction in Magnet Schools	0.111	0.16	0.097	0.094
Average 8th-Grade Standardized Test Score	-0.341	-0.035	0.164	0.384
Average 11th-Grade Standardized Test Score	-0.369	0.028	0.164	0.38
Fraction Graduating High School	0.686	0.805	0.85	0.873
Fraction Entering College	0.598	0.703	0.765	0.838
Number of Observations	113,666	92,481	84,728	75,791

One observation per student. VAM is estimated following the method in Chetty, Friedman, and Rockoff (2014).

Table 2: College Placement by VAM Quartile

	Lowest	2nd	3rd	Highest
No College	0.402	0.297	0.235	0.162
In-State Community College	0.346	0.375	0.351	0.301
In-State Public Four-Year	0.117	0.187	0.259	0.354
In-State Private	0.067	0.073	0.076	0.065
Out of State	0.068	0.069	0.08	0.118
Any College	0.598	0.703	0.765	0.838
In Grade Sample	0.392	0.513	0.563	0.636
Takes Math, If Ever in Michigan Public College	0.401	0.45	0.472	0.484
Takes English, If Ever in Michigan Public College	0.493	0.517	0.499	0.476
Number of Observations	113,666	92,481	84,728	75,791

One observation per student. VAM is estimated following the method in Chetty, Friedman, and Rockoff (2014).

Table 3: College Attendance and Graduation - Probits

	Takes ACT	Graduates HS	Any College	In Grade Sample
Scaled High School VAM	0.016***	0.027***	0.044***	0.028***
	(0.005)	(0.005)	(0.004)	(0.006)
8th-Grade Test Score	0.078***	0.1***	0.162***	0.149***
	(0.002)	(0.002)	(0.001)	(0.003)
Female	0.021***	0.046***	0.083***	0.065***
	(0.001)	(0.002)	(0.002)	(0.002)
Black	-0.002	0.009**	0.092***	0.05***
	(0.004)	(0.004)	(0.005)	(0.006)
Hispanic	-0.038***	-0.037***	-0.031***	-0.041***
	(0.005)	(0.004)	(0.005)	(0.007)
Asian	-0.018**	-0.013	0.033***	0.05***
	(0.008)	(0.01)	(0.012)	(0.011)
Special Education	-0.02***	-0.017***	-0.067***	-0.077***
	(0.003)	(0.002)	(0.002)	(0.003)
Limited English Proficiency	0.008	0.025***	0.064***	0.043***
	(0.007)	(0.007)	(0.011)	(0.015)
Economically Disadvantaged	-0.076***	-0.099***	-0.093***	-0.109***
	(0.002)	(0.002)	(0.002)	(0.002)
Block-Level and School-Level Variables?	Yes	Yes	Yes	Yes
Number of Observations	366,663	366,663	366,665	366,663

One observation per student. Average marginal effects shown. Value added is estimated following the method in Chetty, Friedman, and Rockoff (2014), normalized in terms of its school-level standard deviation. Missing values of covariates are recoded to 0; missing indicators are included but not shown. Bootstrapped standard errors in parentheses, clustered by high school.

Table 4: Falsification Tests

	7^{th}-Grade Score	7^{th}-Grade Attendance	Block-Group Pct. In Poverty	Block-Group Pct. With BA
Scaled High School VAM	0.044***	0.004***	-0.065	1.84***
	(0.006)	(0.001)	(0.318)	(0.439)
Student-Level Variables? Yes	Yes	Yes	Yes	
Block-Level Variables?	Yes	Yes	No	No
School-Level Variables?	Yes	Yes	Yes	Yes
Number of Observations	352,629	350,399	366,666	366,666

One observation per student. Value added is estimated following the method in Chetty, Friedman, and Rockoff (2014), normalized in terms of its school-level standard deviation. Missing values of covariates are recoded to 0; missing indicators are included but not shown. Standard errors in parentheses, clustered by high school.

Table 5: All Grades, Full Sample

	(1)	(2)	(3)	(4)	(5)	(6)
Scaled High School VAM	0.234***	0.185**	0.153***	0.105***	0.098***	0.089***
	(0.024)	(0.019)	(0.017)	(0.011)	(0.014)	(0.014)
8th-Grade Test Score			0.328***	0.274***	0.332***	0.316***
			(0.007)	(0.006)	(0.007)	(0.007)
Female				0.257***	0.258***	0.244***
				(0.006)	(0.006)	(0.005)
Black				-0.426***	-0.309***	-0.295***
				(0.013)	(0.013)	(0.013)
Hispanic				-0.146***	-0.115***	-0.115***
				(0.018)	(0.017)	(0.018)
Asian				0.026	0.04**	0.027
				(0.017)	(0.016)	(0.017)
Special Education				-0.113***	-0.093***	-0.093***
				(0.013)	(0.013)	(0.011)
Limited English Proficiency				0.181***	0.203***	0.184***
				(0.02)	(0.019)	(0.019)
Economically Disadvantaged				-0.169***	-0.152***	-0.155***
				(0.009)	(0.009)	(0.007)
Course ID Fixed Effects?	No	Yes	Yes	Yes	Yes	Yes
Block-Level and School-Level Variables?	No	No	No	No	Yes	Yes
Bootstrapped or Weighted?	Weighted	Weighted	Weighted	Weighted	Weighted	Bootstrapped
Number of Observations	594,239	594,239	594,239	594,239	594,239	594,239

One observation per student-course combination. Value added is estimated following the method in Chetty, Friedman, and Rockoff (2014), normalized in terms of its school-level standard deviation. Outcome variable is the grade in the student's first-year courses in college, on a 4.0 scale. Math and English courses must also be the lowest-numbered such course in the first semester in which a math or English class is taken. Missing values of covariates are recoded to 0; missing indicators are included but not shown. Standard errors in parentheses, clustered by high school.

Table 6: Grades by Subject, Full Sample

	Math	English	Other	Other
Scaled High School VAM	0.122***	0.065***	0.099***	0.087***
	(0.015)	(0.016)	(0.014)	(0.015)
8th-Grade Test Score	0.288***	0.215***	0.367***	0.343***
	(0.013)	(0.01)	(0.008)	(0.007)
Female	0.287***	0.346***	0.223***	0.21***
	(0.009)	(0.009)	(0.007)	(0.006)
Black	-0.312***	-0.299***	-0.309***	-0.289***
	(0.023)	(0.02)	(0.015)	(0.014)
Hispanic	-0.132***	-0.148***	-0.109***	-0.103***
	(0.03)	(0.027)	(0.02)	(0.021)
Asian	0.033	0.042	0.039**	0.024
	(0.032)	(0.027)	(0.015)	(0.015)
Special Education	-0.101***	-0.102***	-0.082***	-0.086***
	(0.021)	(0.02)	(0.013)	(0.012)
Limited English Proficiency	0.245***	0.163***	0.19***	0.176***
	(0.037)	(0.029)	(0.021)	(0.025)
Economically Disadvantaged	-0.094***	-0.16***	-0.16***	-0.166***
	(0.014)	(0.011)	(0.01)	(0.009)
Course ID Fixed Effects?	Yes	Yes	Yes	Yes
Block-Level and School-Level Variables?	Yes	Yes	Yes	Yes
Bootstrapped or Weighted?	Bootstrapped	Bootstrapped	Weighted	Bootstrapped
Number of Observations	79,728	86,566	427,945	427,945

One observation per student-course combination. Value added is estimated following the method in Chetty, Friedman, and Rockoff (2014), normalized in terms of its school-level standard deviation. Outcome variable is the grade in the student's first-year courses in college, on a 4.0 scale. Math and English courses must also be the lowest-numbered such course in the first semester in which a math or English class is taken. Missing values of covariates are recoded to 0; missing indicators are included but not shown. Standard errors in parentheses, clustered by high school.

Table 7: Estimates Accounting for Bias

	All	Math	English	Other
Instrumented VAM	0.068***	0.09***	0.066***	0.064***
	(0.011)	(0.016)	(0.015)	(0.012)
	[594,239]	[79,728]	[86,566]	[427,945]
Home-District Sample	0.103***	0.094***	0.132***	0.091***
	(0.016)	(0.017)	(0.02)	(0.017)
	[482,889]	[64,953]	[69,758]	[348,178]
College-Ready Sample	0.041***	0.076***	0.039**	0.035***
	(0.014)	(0.023)	(0.018)	(0.012)
	[277,361]	[36,024]	[32,812]	[208,515]
Oster (2016) Bias-Adjusted Treatment Effect	0.058***	0.088***	0.02	0.058***
	(0.014)	(0.015)	(0.016)	(0.015)
	[594,239]	[79,728]	[86,566]	[427,945]
Counterfactual Sample, 0.0 GPAs	0.098***	0.066***	0.082***	0.127***
	(0.014)	(0.014)	(0.021)	(0.017)
	[1,136,000]	[243,420]	[254,377]	[638,203]
Counterfactual Sample, Imputed Grades	0.051***	0.029***	0.035***	0.071***
	(0.008)	(0.005)	(0.006)	(0.013)
	[1,136,000]	[243,420]	[254,377]	[638,203]
Altonji and Mansfield (working) Lower Bound	0.023*	0.032**	0.017	0.023
	(0.014)	(0.015)	(0.016)	(0.015)
	[594,239]	[79,728]	[86,566]	[427,945]
Course ID Fixed Effects?	Yes	Yes	Yes	Yes
Block-Level and School-Level Variables?	Yes	Yes	Yes	Yes

One observation per student-course combination. Estimates shown are the coefficients on the high-school value added term in the respective regression specification. Value added is estimated following the method in Chetty, Friedman, and Rockoff (2014), normalized in terms of its school-level standard deviation. Outcome variable is the grade in the student's first-year courses in college, on a 4.0 scale. Math and English courses must also be the lowest-numbered such course in the first semester in which a math or English class is taken. Missing values of covariates are recoded to 0; missing indicators are included but not shown. Bootstrapped standard errors in parentheses, clustered by high school. Sample sizes in brackets.

Table 8: Estimates by School 8^{th}-Grade Exam Quartile

	All	Math	English	Other
Lowest Quartile	0.131***	0.159***	0.139***	0.122***
	(0.02)	(0.029)	(0.03)	(0.022)
	[138,781]	[18,913]	[22,927]	[96,941]
2^{nd} Quartile	0.042*	0.078***	0.006	0.043*
	(0.022)	(0.03)	(0.036)	(0.025)
	[148,373]	[19,994]	[22,862]	[105,517]
3^{rd} Quartile	0.069***	0.117***	0.014	0.071***
	(0.021)	(0.036)	(0.027)	(0.027)
	[153,237]	[20,632]	[21,788]	[110,817]
Highest Quartile	0.056**	0.085***	0.015	0.057**
	(0.025)	(0.034)	(0.034)	(0.026)
	[153,848]	[20,189]	[18,989]	[114,670]
Course ID Fixed Effects?	Yes	Yes	Yes	Yes
Block-Level and School-Level Variables?	Yes	Yes	Yes	Yes

One observation per student-course combination. Estimates shown are the coefficients on the high-school value added term in the respective regression specification. Value added is estimated following the method in Chetty, Friedman, and Rockoff (2014), normalized in terms of its school-level standard deviation. Outcome variable is the grade in the student's first-year courses in college, on a 4.0 scale. Math and English courses must also be the lowest-numbered such course in the first semester in which a math or English class is taken. Missing values of covariates are recoded to 0; missing indicators are included but not shown. Bootstrapped standard errors in parentheses, clustered by high school. Sample sizes in brackets.

Table 9: Estimates by Race and Income

	All	Math	English	Other
Black	0.13***	0.134***	0.125***	0.13***
	(0.021)	(0.037)	(0.021)	(0.024)
	[80,442]	[11,176]	[13,072]	[56,194]
White	0.061***	0.105***	0.029*	0.059***
	(0.014)	(0.02)	(0.016)	(0.012)
	[479,533]	[63,867]	[68,852]	[346,814]
Economically Disadvantaged	0.114***	0.117***	0.124***	0.11***
	(0.015)	(0.027)	(0.024)	(0.019)
	[136,079]	[18,753]	[22,269]	[95,057]
Not Economically Disadvantaged	0.075***	0.117***	0.038***	0.074***
	(0.012)	(0.023)	(0.014)	(0.013)
	[458,160]	[60,975]	[64,297]	[332,888]
Course ID Fixed Effects?	Yes	Yes	Yes	Yes
Block-Level and School-Level Variables?	Yes	Yes	Yes	Yes

One observation per student-course combination. Estimates shown are the coefficients on the high-school value added term in the respective regression specification. Value added is estimated following the method in Chetty, Friedman, and Rockoff (2014), normalized in terms of its school-level standard deviation. Outcome variable is the grade in the student's first-year courses in college, on a 4.0 scale. Math and English courses must also be the lowest-numbered such course in the first semester in which a math or English class is taken. Missing values of covariates are recoded to 0; missing indicators are included but not shown. Bootstrapped standard errors in parentheses, clustered by high school. Sample sizes in brackets.

Table 10: Estimates by College Type

	All	Math	English	Other
Four-Year Colleges	0.076***	0.095***	0.049**	0.076***
	(0.015)	(0.022)	(0.023)	(0.015)
	[332,516]	[41,019]	[36,923]	[254,574]
Two-Year Colleges	0.098***	0.137***	0.075***	0.095***
	(0.016)	(0.023)	(0.021)	(0.018)
	[254,192]	[37,613]	[48,495]	[168,084]
Course ID Fixed Effects?	Yes	Yes	Yes	Yes
Block-Level and School-Level Variables?	Yes	Yes	Yes	Yes

One observation per student-course combination. Estimates shown are the coefficients on the high-school value added term in the respective regression specification. Value added is estimated following the method in Chetty, Friedman, and Rockoff (2014), normalized in terms of its school-level standard deviation. Outcome variable is the grade in the student's first-year courses in college, on a 4.0 scale. Math and English courses must also be the lowest-numbered such course in the first semester in which a math or English class is taken. Missing values of covariates are recoded to 0; missing indicators are included but not shown. Bootstrapped standard errors in parentheses, clustered by high school. Sample sizes in brackets.

A Mathematical Appendix

As stated above, when the high-stakes accountability constraint binds, the result must satisfy the accountability constraint $KS^{\gamma}L^{\delta} = \theta$ and the budget constraint $\Omega = c_S S + c_L L$. Solve the accountability constraint for S:

(19) $$S^{\gamma} = \frac{\theta}{AL^{\delta}}; S = \frac{\theta^{\frac{1}{\gamma}}}{K^{\frac{1}{\gamma}}L^{\frac{\delta}{\gamma}}}$$

Substitute this into the budget constraint:

(20) $$\Omega(1 + A^{**}) = c_S \frac{\theta^{\frac{1}{\gamma}}}{K^{\frac{1}{\gamma}}L^{\frac{\delta}{\gamma}}} + c_L L$$

This cannot be solved analytically unless $\delta = \gamma$. However, if $\frac{\partial T}{\partial S} > \frac{\partial T}{\partial L}$ at the unconstrained optimum under high-stakes accountability, then the school will reallocate toward short-term resources (i.e. focus more on test preparation) in order to move toward the threshold. Since $T = KS^{\gamma}L^{\delta}$:

(21) $$\frac{\partial T}{\partial S} = \gamma K S^{\gamma-1}L^{\delta}; \frac{\partial T}{\partial L} = \delta K S^{\gamma}L^{\delta-1}$$

Evaluate the derivatives at the unconstrained maximizing values of S and L:

(22) $$\frac{\partial T}{\partial S} = \gamma K \left(\frac{\Omega(1 + A^{**})}{c_S}\left(\frac{\alpha\gamma}{\alpha\gamma + \alpha\delta + \beta\zeta}\right)\right)^{\gamma-1}\left(\frac{\Omega(1 + A^{**})}{c_L}\left(\frac{\alpha\delta + \beta\zeta}{\alpha\gamma + \alpha\delta + \beta\zeta}\right)\right)^{\delta}$$

(23) $$\frac{\partial T}{\partial L} = \delta K \left(\frac{\Omega(1 + A^{**})}{c_S}\left(\frac{\alpha\gamma}{\alpha\gamma + \alpha\delta + \beta\zeta}\right)\right)^{\gamma}\left(\frac{\Omega(1 + A^{**})}{c_L}\left(\frac{\alpha\delta + \beta\zeta}{\alpha\gamma + \alpha\delta + \beta\zeta}\right)\right)^{\delta-1}$$

Start by dividing both equations by $\left(\frac{\Omega(1+A^{**})}{c_S}\left(\frac{\alpha\gamma}{\alpha\gamma+\alpha\delta+\beta\zeta}\right)\right)^{\gamma-1}\left(\frac{\Omega(1+A^{**})}{c_L}\left(\frac{\alpha\delta+\beta\zeta}{\alpha\gamma+\alpha\delta+\beta\zeta}\right)\right)^{\delta-1}$; then $\frac{\partial T}{\partial S} > \frac{\partial T}{\partial L}$ if:

(24) $$\gamma\left(\frac{\Omega(1 + A^{**})}{c_L}\left(\frac{\alpha\delta + \beta\zeta}{\alpha\gamma + \alpha\delta + \beta\zeta}\right)\right) > \delta\left(\frac{\Omega(1 + A^{**})}{c_S}\left(\frac{\alpha\gamma}{\alpha\gamma + \alpha\delta + \beta\zeta}\right)\right)$$

Then divide through by $\frac{\gamma\Omega(1+A^{**})}{\alpha\gamma+\alpha\delta+\beta\zeta}$, and obtain that $\frac{\partial T}{\partial S} > \frac{\partial T}{\partial L}$ if:

$$\frac{\alpha\delta + \beta\zeta}{c_L} > \frac{\alpha\delta}{c_S} \tag{25}$$

If the costs are equal, the school will always move toward short-term resources if constrained, because $\beta\zeta > 0$.

B Data Appendix

B.1 Covariates In Fully-Specified Regressions

The first stage of the value-added model contains the following covariates.

- Individual test scores: standardized 8^{th}-grade math score, standardized 8^{th}-grade reading score, interaction of standardized 8^{th}-grade math and reading scores.
- Individual demographics: black, Hispanic, Asian, female, economically disadvantaged, limited English proficiency, special education.
- Census block group averages: household income, has a high school diploma, has a bachelor's degree, white, black, Asian, Hispanic, married, employed, home owner, below poverty line.
- School averages: black, economically disadvantaged, 8^{th}-grade test score.
- School-level variables: enrollment.
- Missing indicators for all of the above.
- Cohort fixed effects.
- School fixed effects (variable of interest).

The second stage of the value-added model contains the following covariates.

- Average residualized scores: one year forward, two years forward, one year backward, two years backward.
- Missing indicators for all of the above.

The Altonji and Mansfield (working) decomposition regression contains the following covariates.

- Individual test scores: standardized 8^{th}-grade math score, standardized 8^{th}-grade reading score, interaction of standardized 8^{th}-grade math and reading scores.
- Individual demographics: black, Hispanic, Asian, female, economically disadvantaged, limited English proficiency, special education.
- Census block group averages: household income, has a high school diploma, has a bachelor's degree, white, black, Asian, Hispanic, married, employed, home owner, below poverty line.
- School averages: black, economically disadvantaged, 8^{th}-grade test score.
- School-level variables: enrollment, per-pupil instructional expenditure, district per-pupil instructional expenditure, pupil/teacher ratio, fraction of students leaving the school without graduating, teacher certification exam score.
- Missing indicators for all of the above.
- Cohort fixed effects.

The outcome regressions contain the following covariates.

- Individual test scores: standardized 8^{th}-grade scores in a fourth-order polynomial.
- Individual demographics: black, Hispanic, Asian, female, economically disadvantaged, limited English proficiency, special education.
- Census block group averages: household income, has a high school diploma, has a bachelor's degree, white, black, Asian, Hispanic, married, employed, home owner, below poverty line.
- School averages: black, economically disadvantaged, 8^{th}-grade test score, attendance.
- School-level variables: enrollment, magnet, charter.
- Missing indicators for all of the above.
- High school value added, expressed in terms of its school-level standard deviation (variable of interest).

B.2 Subject Abbreviations

The following abbreviations correspond to English courses at the listed institutions:

- "ENG": Alpena Community College, Central Michigan University, Delta College, Gogebic Community College, Grand Valley State University, Henry Ford Community College, Jackson Community College, Kalamazoo Valley Community College, Kirtland Community College, Marygrove College, Mid Michigan Community College, Muskegon Community College, North Central Michigan College, Northwestern Michigan College, Oakland Community College, Schoolcraft College, St. Clair County Community College, University of Michigan-Flint, Washtenaw Community College, Wayne County Community College, Wayne County Community College District, Wayne State University, West Shore Community College
- "ENGL": Bay de Noc Community College, Eastern Michigan University, Ferris State University, Kellogg Community College, Lake Michigan College, Lake Superior State University, Macomb Community College, Monroe County Community College, Montcalm Community College, Mott Community College, Saginaw Valley State University, Southwestern Michigan College, The Robert B. Miller College, Western Michigan University
- "DEN": College for Creative Studies
- "COM": Glen Oaks Community College
- "EN": Grand Rapids Community College, Northern Michigan University
- "WRIT": Lansing Community College
- "WRA": Michigan State University
- "UN": Michigan Technological University
- "WRT": Oakland University
- "ENGLISH": University of Michigan-Ann Arbor
- "COMP": University of Michigan-Dearborn

The following abbreviations correspond to math courses at the listed institutions:

- “MTH”: Alpena Community College, Central Michigan University, Delta College, Gogebic Community College, Grand Valley State University, Jackson Community College, Kirtland Community College, Marygrove College, Michigan State University, North Central Michigan College, Northwestern Michigan College, Oakland University, St. Clair County Community College, University of Michigan-Flint, Washtenaw Community College, West Shore Community College

- “MATH”: Bay de Noc Community College, Eastern Michigan University, Ferris State University, Henry Ford Community College, Kalamazoo Valley Community College, Kellogg Community College, Lake Michigan College, Lake Superior State University, Lansing Community College, Macomb Community College, Monroe County Community College, Montcalm Community College, Mott Community College, Muskegon Community College, North Central Michigan College, Saginaw Valley State University, Schoolcraft College, Southwestern Michigan College, The Robert B. Miller College, University of Michigan-Ann Arbor, University of Michigan-Dearborn, Western Michigan University

- “NSM”: Glen Oaks Community College

- “MA”: Grand Rapids Community College, Michigan Technological University, Northern Michigan University

- “MAT”: Mid Michigan Community College, Oakland Community College, Wayne County Community College, Wayne County Community College District, Wayne State University

- “MMTH”: West Shore Community College

B.3 Appendix Tables

Table 11: College Attendance and Graduation - Probits, College-Ready Sample

	Takes ACT	Graduates HS	Any College	In Grade Sample
Scaled High School VAM	0.015***	0.024***	0.028***	0.005
	(0.004)	(0.004)	(0.004)	(0.009)
8th-Grade Test Score	0.052**	0.056***	0.151***	0.258***
	(0.021)	(0.022)	(0.041)	(0.058)
Female	0.006***	0.015***	0.045***	0.032***
	(0.001)	(0.002)	(0.002)	(0.003)
Black	-0.016***	-0.012**	0.022***	-0.03***
	(0.003)	(0.006)	(0.007)	(0.011)
Hispanic	-0.021***	-0.023***	-0.012	-0.025***
	(0.004)	(0.005)	(0.008)	(0.009)
Asian	-0.021***	-0.022***	-0.002	0.036***
	(0.003)	(0.004)	(0.009)	(0.011)
Special Education	-0.032***	-0.044***	-0.052***	-0.073***
	(0.004)	(0.005)	(0.006)	(0.009)
Limited English Proficiency	-0.037***	-0.037***	-0.037***	-0.06***
	(0.006)	(0.01)	(0.009)	(0.015)
Economically Disadvantaged	-0.037***	-0.052***	-0.064***	-0.091***
	(0.002)	(0.003)	(0.003)	(0.004)
Block-Level and School-Level Variables?	Yes	Yes	Yes	Yes
Number of Observations	115,116	115,116	115,116	115,116

One observation per student. Marginal effects at means shown. There are as many "college-ready" students as have a maximum ACT composite of 21 or higher; this same number of students is then chosen based on their 8th-grade test score. Value added is estimated following the method in Chetty, Friedman, and Rockoff (2014), normalized in terms of its school-level standard deviation. Missing values of covariates are recoded to 0; missing indicators are included but not shown. Bootstrapped standard errors in parentheses, clustered by high school.

Table 12: Falsification Tests, Home-District Sample

	7^{th}-Grade Score	7^{th}-Grade Attendance	Block-Group Pct. In Poverty	Block-Group Pct. With BA
Scaled High School VAM	0.054***	0.003**	-0.145	1.92***
	(0.007)	(0.001)	(0.365)	(0.583)
Student-Level Variables? Yes	Yes	Yes	Yes	
Block-Level Variables?	Yes	Yes	No	No
School-Level Variables?	Yes	Yes	Yes	Yes
Number of Observations	275,601	276,897	287,348	287,348

One observation per student. Only students who attend non-charter, non-magnet high schools in the district in which they reside are included. Value added is estimated following the method in Chetty, Friedman, and Rockoff (2014), normalized in terms of its school-level standard deviation. Missing values of covariates are recoded to 0; missing indicators are included but not shown. Standard errors in parentheses, clustered by high school.