ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 23-729

Understanding High Schools' Effects on Longer-Term Outcomes

Preeya P. Mbekeani
Brown University

John Papay
Brown University

Ann Mantil
Brown University

Richard J. Murnane
Harvard University

Improving education and labor market outcomes for low-income students is critical for advancing socioeconomic mobility in the United States. We explore how Massachusetts public high schools affect the longer-term outcomes of low-income students, using detailed longitudinal data. We estimate school value-added impacts on four-year college graduation and earnings. Similar students who attend schools at the 80th percentile of the distribution are 6 percentage points more likely to graduate from a four-year college and earn 13% (or $3,600) more annually at age 30 compared to peers who attend schools at the 20th percentile. We consider how school effectiveness across a range of short-term measures relates to longer-run impacts. Schools that improve students' test scores and college aspirations improve longer-run outcomes more.

VERSION: February 2023

Suggested citation: Mbekeani, Preeya P., John Papay, Ann Mantil, and Richard J. Murnane. (2023). Understanding High Schools' Effects on Longer-Term Outcomes. (EdWorkingPaper: 23-729). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/qwp6-hk05

Understanding High Schools' Effects on Longer-Term Outcomes

Preeya P. Mbekeani
John P. Papay
Ann Mantil
Brown University

Richard J. Murnane
Harvard University

February 10, 2023

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305H190035 to Brown University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. The authors thank staff at the Massachusetts Department of Elementary and Secondary Education for their ongoing partnership and support. We also thank Emma Zorfass and Aubrey McDonough for excellent research support.

Abstract

Improving education and labor market outcomes for low-income students is critical for advancing socioeconomic mobility in the United States. We explore how Massachusetts public high schools affect the longer-term outcomes of low-income students, using detailed longitudinal data. We estimate school value-added impacts on four-year college graduation and earnings. Similar students who attend schools at the 80th percentile of the distribution are 6 percentage points more likely to graduate from a four-year college and earn 13% (or $3,600) more annually at age 30 compared to peers who attend schools at the 20th percentile. We consider how school effectiveness across a range of short-term measures relates to longer-run impacts. Schools that improve students' test scores and college aspirations improve longer-run outcomes more.

Keywords: school effects, high schools, longer-run outcomes, value-added estimation, low-income students

Introduction

Parents and policymakers have long cared about differences among schools in how they promote student learning, development, and well-being. State testing under No Child Left Behind (NCLB) served to publicize the widely understood fact that schools differ substantially in their students' average test scores, and such scores were treated as proxies for school quality. Test score levels, however, reflect students' family backgrounds and prior schooling, not only the impact of schools. Moreover, test scores are proxies for what many parents, educators, and policymakers really want to know – namely, whether schools promote the development of skills, knowledge, capacities, and attitudes that enable students to "flourish" as adults (Brighouse, et al., 2016). We know little about how high schools affect students' longer-run educational and labor-market outcomes and whether the schools in which students show more growth in short-term outcomes are those that also promote students' longer-run success.

These questions are of particular interest for students living in poverty, who tend not to enjoy the myriad advantages and opportunities for enrichment experienced by their peers from wealthier families (Kaushal, Magnuson, & Waldfogel, 2011). In other words, school quality may matter more for students who grow up in poverty. However, such students often attend under-resourced schools with less experienced teachers and limited access to high quality instructional materials (Boyd et al. 2005; Clotfelter, Ladd, & Vigdor 2009; Kalogrides & Loeb, 2013; Lankford, Loeb, & Wyckoff, 2002). As a result, understanding the ways in which schools affect the longer-run outcomes and economic mobility of students living in poverty is central for efforts to promote equality of opportunity and social mobility (Duncan & Murnane, 2011; Chetty et al., 2020).

In this study, we use detailed longitudinal data from Massachusetts, including information on students' longer-run educational attainments and earnings, for first-time 9th grade public-school students in 2003 and 2004. We focus on students from families with low family income who attend high schools that serve large shares of students living in poverty. We estimate school value-added models, conditioning on lagged outcomes and a rich set of additional covariates not often found in administrative datasets, in order to account for the sorting of students into high schools and to identify schools' causal impacts on these longer-run outcomes. Past research suggests that such models yield results comparable to those from lottery-based studies (Deming, 2014; Angrist et al., 2017). We posit a multi-dimensional framework of school effectiveness and examine whether schools that improve particular short-term outcomes more than expected also have larger-than-expected effects on longer-run educational attainment and earnings.

We document substantial variation in high schools' effects on low-income students' longer-run outcomes. For the average low-income student in our sample, the difference in the probability of four-year college graduation associated with attending a school at the 20th percentile of the school effectiveness distribution as compared to one at the 80th percentile is 5.8 percentage points and the difference in annual earnings is 13% (approximately $3,600 for students with earnings near the sample mean). We show that our results are unlikely to be driven by differences in unobserved characteristics of the students enrolled in different high schools.

We also find that schools' impacts on short-term outcomes relate to their impacts on educational attainments and earnings. Schools with larger estimated effects on 10th grade test scores have larger estimated effects on both four-year college graduation and earnings. We find large and significant relationships between schools' impacts on students' college-going plans and their effects on students' longer-run outcomes. Relationships between other short-term outcomes and longer-run outcomes are also positive but smaller in magnitude. Taken together, our results indicate that what school a student living in poverty attends has substantial impacts on their longer-term educational attainments and earnings and that short-term proxies serve as important markers of ultimate impacts on these longer-term outcomes.

Background

Efforts to assess school "quality" in the United States have largely focused on student standardized test scores. This proxy measure has been used extensively, both by researchers and in state accountability systems, particularly as a result of NCLB (Linn, 2008; Loeb et al., 2018).

However, existing research has made clear that test score levels reflect not only the causal impacts of schools but also the impacts of students' family backgrounds, neighborhoods, and prior schooling experiences. In other words, student sorting, on both observed and unobserved characteristics, must be addressed in order to learn how schools differ in their effectiveness (Angrist, et al., 2021; Angrist, Hull, & Walters, 2022; Raudenbush & Willms, 1995).

These concerns have led to a second wave of analyses that use "value-added" or test score growth as measures of school quality. In the typical constant-effects "value-added" model preferred by researchers, the outcome is a student test score and the covariates include a test score from a prior grade (Angrist et al., 2022). Reardon (2018) found that, at the school level, the within-cohort growth in test scores is largely uncorrelated with the average test-score levels of entering students, and within-cohort growth is associated with longitudinal growth measures (Reardon et al., 2019). This research has led to substantial policy efforts to hold schools accountable for student academic growth occurring over the course of a single year. For example, many states use student growth percentiles, a distinct but related approach, that measures student performance in a given year relative to the set of students with the same test scores in a prior year (Betebenner, 2009).

While these measures that aim to assess students' academic growth or learning over the course of a year are substantial improvements over systems that do not take into account students' prior achievement, relying exclusively on standardized test scores reflects an overly narrow view of what schools can do to improve students' long-run opportunities (Loeb et al., 2018; Jackson et al., 2020; Jennings et al., 2015). Some standardized tests measure only basic skills; most do not assess the full range of academic knowledge and skills in core subjects. Schools also do many things beyond raising test scores that matter for students' longer-run success. Indeed, society's aims for schooling go beyond academic learning. Among other things, schools are expected to develop students' social and behavioral skills, including teamwork and personal responsibility. They should provide a general core of knowledge as well as opportunities for students to explore interests and develop their passions. Schools are also expected to prepare students to engage constructively in democratic citizenship (Brighouse et al., 2016; Gutmann, 1999). Research across a range of fields has demonstrated that a variety of skills, attitudes, and behaviors, such as self-discipline, motivation, work habits, and social skills, predict educational attainment and labor market outcomes (Allensworth & Clark, 2020; Heckman & Rubenstein, 2001; Duckworth & Seligman, 2005; Deming, 2017). A series of recent studies have used value-added approaches to explore how schools promote outcomes other than test scores, such as social-emotional learning and academic engagement (e.g., Jackson et al. 2020; Loeb et al. 2018).

Of course, while short-term measures of academic and social skill development are important, they do not address the ultimate impacts of schooling on students as adults. In recent years, advances in state longitudinal data systems have made it possible to examine a greater variety of longer-run student outcomes. Some of the strongest evidence of schools' effects on these outcomes comes from quasi-experimental studies exploiting school lotteries or other plausibly exogenous assignment mechanisms (e.g., Angrist et al. 2016; Angrist et al. 2017; Deming, 2011; Deming et al., 2014). For example, Deming (2011) found that "high-risk" male students who won a lottery to attend their first-choice high school in Charlotte-Mecklenburg Schools (CMS) were less likely to be arrested as young adults. Studies have also found positive impacts of winning school lotteries on four-year college enrollment and degree completion (Angrist et al., 2016; Deming et al., 2014).

These studies also make clear that school value-added models that incorporate rich sets of covariates, including prior test scores, generally recover the causal impacts of schools on student test scores. Using data from CMS, for example, Deming (2014) found that nonexperimental estimates of school value-added nearly matched estimates from the randomized lottery-based study. The results indicate that the covariates typically available in state administrative datasets are largely sufficient to account for the sorting of students across schools.

One critical question is whether high schools' effects on longer-run outcomes, such as educational attainments and earnings, are driven by impacts on student academic learning as measured on standardized tests. Few studies have examined this directly, and the existing evidence is mixed. Angrist and colleagues (2016) reported that charter school effects on four-year college enrollment were strongly correlated with effects on test scores. In contrast, Deming (2011) found no evidence of impacts on test scores for high-risk male lottery winners as compared to their peers who lost out in the same lotteries, despite finding impacts on later criminal activity. Similarly, Kemple and Willner (2008) found that attending a Career Academy high school, which combines academic and technical curricula, had no measurable effect on student test scores but improved longer-run earnings by 11%. More recently, Jackson and colleagues (2020) found that high schools' impacts on students' social-emotional development were more predictive of postsecondary outcomes than schools' test-score value-added.

A Multi-Dimensional Framework of School Effectiveness

We propose a multi-dimensional framework for how schools can affect students' longer-run outcomes, which we illustrate in Figure 1. For simplicity, we name four overlapping goals of schooling – supporting students' educational attainments, developing students' ability to succeed in the labor market, promoting civic engagement and participation, and supporting students' long-term health and well-being – although additional goals fit easily into this framework. In theory, schools can promote student success by influencing a range of capacities and dispositions that contribute to these outcomes (Brighouse et al., 2016). The most obvious of these is academic knowledge and skills. Schools differ in the quantity and quality of learning opportunities available to students. This includes the coherence, depth, and quality of instruction (Mehta & Fine, 2019; Rowan, 2011); access to advanced courses, such as Advanced Placement or dual enrollment in college courses; and how students are tracked into different course pathways, including the extent to which students receive advice and information about which courses to take.

Importantly, we can divide academic skills into two categories: those measured by state tests and those not. For example, state standardized tests wholly exclude some content areas (e.g., history/social science in Massachusetts) and do not assess some critical skills in tested subjects (e.g., oral communication skills as part of English language arts). Unmeasured academic knowledge and skills are important for students' longer-run outcomes.

Beyond academics, schools influence a range of skills that pay off in the long run. For example, schools can develop students' employment skills; this may occur through hands-on learning experiences, career and technical education (CTE) courses, and opportunities that build capacities for specific careers. School culture, including behavioral norms and disciplinary practices, shapes students' social skill development. Schools may also have explicit curricula focused on building students' social-emotional skills. Schools affect students' attitudes and dispositions about their futures. For example, schools influence students' knowledge of and ideas about postsecondary education and career options through counselors, teachers, peers, and other networks (Hoxby & Avery, 2013; Mantil, 2022; Mulhern, 2020). As such, they serve an important role in shaping students' aspirations for future education and careers.

The effectiveness of schools in supporting the development of these skills, capacities, and dispositions depends on at least three factors: their priorities, their resources, and their students. Schools may choose to prioritize some dimensions of learning or skill development more than others. Priorities may be influenced by a range of factors, including the goals of the local community, the priorities of the adults in the school, the school's position in a state's accountability system, and broader social forces. For example, a school facing sanctions may focus more on raising high-stakes test scores than on developing students' socioemotional skills. By contrast, in the wake of the pandemic some schools have re-invested in efforts to support student well-being and social-emotional development.

School resources, broadly defined, are key determinants of how well schools can achieve their priorities. For example, between two schools equally interested in promoting academic skills development, one may do better because of greater financial resources, stronger leadership, more effective teachers, or a more robust curriculum. Most obviously, financial resources play a key role in shaping the quality of the educator workforce, physical facilities, and learning materials

(Jackson, Johnson & Persico, 2016; LaFortune, Rothstein, & Schanzenbach, 2018). Schools differ substantially in their human resources as well, with critical effects on student learning and development. Beyond the resources themselves, how efficiently they are deployed and the extent to which this deployment is aligned with school priorities shape how well schools can achieve their aims.

Finally, a key component of high schools is the student body. Peers matter, and student body composition has important implications not only for classroom learning but also for the social and relational capital students can access within and beyond school (Harris, 2010; Jencks & Mayer, 1990; Xu et al., 2022). Possible mechanisms through which peers may positively impact a student’s longer-run outcomes include promoting an engaged school culture, contributing to a college-going ethos, and providing connections to employment networks.

Research Questions

We focus on high schools’ effects on two longer-term outcomes, four-year college graduation and adult labor market earnings. We then explore whether schools’ impacts on these longer-term outcomes correlate with their impact on more proximal measures aligned to our framework. While we do not have direct measures of all of the dimensions described above (e.g., employment skills), our data enable us to construct reasonable proxy measures for several shorter-term outcomes. Throughout the paper, we focus attention on students with low family income who attend high schools that serve large shares of students living in poverty. We address two questions: First, what is the variation in high schools’ effects on four-year college graduation and adult earnings? And second, are schools with larger than expected effects on these longer-run outcomes those that also improve test scores, improve students’ attendance, influence their academic trajectories, and/or promote college going?

Research Design

Data & Sample

We use student-level data from the Massachusetts Department of Elementary and

Secondary Education (DESE), including student demographic and enrollment information, standardized test scores, and rich data from student questionnaires that students complete when they take state tests in 8th and 10th grade. DESE collects information on college outcomes from the National Student Clearinghouse (NSC). NSC data include nearly all colleges and universities (public, private, 2-year, and 4-year) in the United States. The coverage rate in Massachusetts approached 95% in 2011 and has improved since then (Dynarski, Hemelt, & Hyman, 2015; National Student Clearinghouse, 2022). De-identified data on labor market earnings come from the Massachusetts Unemployment Insurance (UI) system, which matches records to directory information for former public-school students.

Our analytic sample consists of students who entered 9th grade for the first time during the 2002-03 or 2003-04 school years. We focus on these years because they provide a sufficient time horizon to examine our longer-term outcomes of interest, particularly earnings. Labor market earnings vary substantially throughout early adulthood, as youth are in and out of college and the workforce and may not have consistent employment. Earnings are quite variable in the decade after students graduate from high school, but individuals’ placement in the earnings distribution tends to stabilize by age 30 (Chetty et al., 2014); therefore, we examine earnings when the average student in our sample is 30 or 31.

We focus on public high schools in Massachusetts that serve large numbers of students from low-income families. We use eligibility for free- or reduced-price lunch (FRPL) in 8th grade as our indicator of low family income and refer to students who qualify for FRPL as “low-income” and students who do not as “higher-income.” While some higher-income students in these schools may be relatively better off than students identified as low-income (as opposed to others whose families’ earnings fall just above the cutoff), these students are unlikely to be from families with truly “high” incomes. We identified 106 high schools in which there were at least 20 low-income students per cohort and at least 25% of all students were low-income.[1] These schools served 76% of low-income 9th grade public school students in the state in 2002-03 and 2003-04. The share of low-income students in these schools ranges from 25% to 87%, with an average of 52%. Sixty-six of the schools are located in urban areas, 9 are charter schools, and 27 are CTE schools, such as vocational and technical high schools. Given that more than a quarter of our sample consists of CTE schools, we present findings specific to CTE and non-CTE schools after describing the main results.

We limit our analytic sample to students who were in a Massachusetts public school in 8th grade and have non-missing data on 8th grade test scores, FRPL eligibility, special education status, English learner status, and attendance. Our analytic sample excludes approximately 15% of first-time 9th graders in the state due to missing 8th grade test scores; nearly all of these are students who first entered Massachusetts public schools in 9th grade. In Table 1, we present descriptive characteristics of our analytic sample. Approximately half are low-income. On average, students in our sample have substantially lower 8th grade test scores (-0.5 standard deviations) and worse short-term and longer-term outcomes than the average Massachusetts student; these patterns are even more pronounced for the low-income students in the sample, on whom we focus.

Measures

Our first longer-run outcome is whether a student graduated from a four-year college within ten years of entering ninth grade (by about age 24, for those who progress on time). While more than one-third of students statewide had earned a four-year college degree within this time frame, only 18% of students in our analytic sample did so, and among low-income students in the sample, the share was 11%. We also examine whether a student earned any degree, i.e. two- or four-year, within the same time period. Our results are quite similar across these definitions. We focus on four-year college graduation because of its stronger relationship with other longer-run outcomes,

[1] We excluded from our analytic sample the three exam high schools in Massachusetts, all located in Boston, that use selective admissions procedures. We show in supplementary analyses, available upon request, that our results are robust to their inclusion.

including increased civic participation, such as volunteering and voting, and positive health behaviors (Ma, Pender, & Welch, 2019).

Our second longer-run outcome is adult labor market earnings. We express in 2022 dollars individuals' logged average quarterly earnings from the three years when students are approximately age 30 (14 to16 years after 9th grade).[2] We omit quarters with no reported earnings and scale the average quarterly value by 4 to reflect annual earnings. For succinctness, we refer to this outcome interchangeably as adult earnings or earnings at age 30. The median earnings of the students in our sample was approximately $36,000 and $32,500 for low-income students.

We lack earnings data for approximately 30% of the analytic sample. The state UI data do not include earnings from self-employment, employment in the federal government or military, under-the-table pay, or earnings from employment in another state. As such, we cannot detect whether individuals with no reported earnings are unemployed, out of the labor force, or have non-reported earnings. As expected, students who leave Massachusetts for college and students who do not earn a high school degree are more likely to be missing earnings data, but probably for different reasons. We are reassured that our estimates reflect impacts on true earnings for several reasons. First, Opportunity Insights' Opportunity Atlas reports earnings at approximately the same age for students who grew up in Massachusetts (Chetty et al., 2018). Nearly 70% of students remain in the same commuting zone, echoing the share of students in our sample for whom we have reported earnings. Furthermore, students who leave the commuting zone have reasonably similar earnings (about 5% higher) than those who stay in the zone. Finally, estimates from the American Community Survey (ACS) suggest that of 29–31-year-olds born in Massachusetts who still live in state and are employed, 94% are employed in state while 6% are employed out of state.[3]

[2] This is our preferred earnings outcome because it avoids potentially underestimating earnings from students being in and out of the workforce over the course of a year. We create an alternative earnings outcome measure—the three-year average of students' total annual earnings, expressed in 2022 dollars and logged. We present these supplementary results after our main results in Footnote 8.
[3] Authors' calculations using ACS 2018 1-Year Public Use Microdata Sample (PUMS) available from the U.S. Census Bureau.

Our central question predictor is the first public high school a student attended. Approximately 25% of low-income students transfer between schools during their high school careers. Given that transfer is endogenous and may reflect the policies and practices of the first high school, we focus on first high school in an intent-to-treat-style analysis. In other words, our estimates likely understate true variation in schools' effects; we view them as conservative.

We measure all covariates in 8th grade, before students enroll in high school. As stated above, we use eligibility for FRPL in 8th grade as our indicator for family income. In 2004, the maximum annual income for reduced-price lunch eligibility for a family of four was $34,873 (equivalent to approximately $54,000 in 2022 dollars). We use information on gender, race/ethnicity, attendance, and whether a student received special education services or was classified as an English learner. As shown in Table 1, 6% of students identify as Asian, 17% as African-American or Black, 23% as Hispanic, less than 1% as Native American, and 54% as White. Nineteen percent of students receive special education services, and about 7% are English learners. We operationalize attendance as a student's attendance rate, which is calculated by dividing days in attendance by days in membership in 8th grade. The average attendance rate among low-income students in our sample is 92%.

A critical covariate is each student's score on the Massachusetts Comprehensive Assessment System (MCAS) grade-8 mathematics examination, the only MCAS test required for 8th graders in these cohorts. We standardize students' grade-8 raw scores across the population of test takers in the state for each cohort, such that the average score in each statewide cohort is zero, with a standard deviation of one. We refer to all of the previously described 8th grade measures (test scores, race/ethnicity, gender, FRPL, SPED, EL, and attendance rate) as "Standard" controls, as they are typical measures in statewide longitudinal data systems.

As part of the MCAS administration in these years, students responded to questionnaires that asked them about, among other things, their course-taking, their perception of school safety, their plans for their education after high school, and their parents' highest level of education. We refer to these 8th grade survey data as "Expanded" controls, as they are not typically available in statewide data systems. We convert the categorical responses to indicator variables. Of particular interest is information on parents' highest level of education and students' plans after high school.

One limitation of these survey data is substantial missingness due to both item and survey non-response. For example, the questionnaires for 30% of students in our analytic sample lack responses about their mother's or father's highest level of education. To retain these students in our sample, we include an indicator for whether a student's response is missing. Our results are quite similar if we restrict our sample to students who have non-missing information on parental education from the 8th grade survey.

We construct four variables that serve as short-term high-school outcomes aligned to our framework of school effectiveness: (1) a test score index, (2) an index indicating whether the student is academically on-track, (3) an attendance index, and (4) an indicator of whether a student plans to attend a four-year college after high school. While we focus our results on low-income students, we standardize these four measures, such that the mean is 0 and the standard deviation is 1, within our analytic sample that includes both low- and higher-income students. We do this for ease of interpretation and to facilitate comparisons across these outcomes. We construct these measures using the administrative data and 10th grade student questionnaire responses, as follows.

- Test-score index: We form a standardized composite of students' 10th grade mathematics and ELA MCAS scores by averaging each student's standardized raw scores.
- Academic on-track index: We form the academic index from student reports of their 9th and 10th grade math courses and an indicator calculated from the administrative data of whether they took their 10th grade MCAS tests on time based on when they entered 9th grade. We create a composite of these correlated items by using weights derived from the first component of a principal components analysis (PCA).
- Attendance index: Following recent literature (Jackson, 2018; Kraft, 2019), we create an index that includes a student's attendance (logit function) and two indicators of whether they were

suspended in-school and out-of-school in their test year. We use the PCA procedure described above to form the index.

- College plans: We create a binary measure using information from the questionnaire of whether a student planned to attend a four-year college after high school. We refer interchangeably to this measure as college plans or college aspirations.

Finally, we construct a composite measure of peer effects, using three measures: peer achievement, peer college aspirations, and peer social capital. We proxy for peer achievement with the leave-out average (i.e., omitting a student's own score from average) 8th grade mathematics test scores of students in the same cohort and high school.[4] To create the measures of peers' educational aspirations and social capital, we calculate the share of students in the same cohort and high school who indicated in 8th grade that they plan to attend a four-year college and the share that indicated that one or both of their parents had completed a four-year college degree. We construct the composite of these three measures using PCA as described above.

Analytic Strategy

We estimate high school effects on our longer-run outcomes using school value-added models (VAMs) that estimate a school's effect on a given student outcome conditional on a student's prior achievement and demographic characteristics (see e.g., Jackson et al., 2020a; Jackson et al., 2020b; Jennings et al., 2015; Loeb et al., 2018; Lloyd & Schachner, 2021). Conceptually, the effect or value-added of a given school *j* represents its contribution to student *i*'s outcome relative to the student's expected outcome averaged across all other high schools in the sample. These measures are relative – they do not measure the impact of a student attending a given high school compared to not attending school, but compared to attending the average school in the sample. Below, we provide evidence supporting the validity of these models to uncover the

[4] These peer effect proxies necessarily include only peers who had non-missing 8th grade data. The share of students with missing data ranges from 0% to 39% across schools, suggesting that these proxies may not fully reflect a given student's actual peers. We find that the share of missingness is largely uncorrelated with school impacts on our longer-run outcomes.

causal effects of schools on students short- and longer-term outcomes.

We estimate two-level mixed models with random effects for high schools. These models take the general form:

$$Y_{ijt} = \alpha + \mathbf{B}\mathbf{X}_i + \mathbf{\Gamma}\mathbf{E}_i + \delta_t + (u_j + \epsilon_{ijt}) \qquad \text{Eq. (1)}$$

where Y_{ijt} represents one of our longer-term outcomes for student (i) in school (j) and cohort (t). School-specific random effects are captured by u_j. The standard control variables are represented by the vector $\mathbf{X}_i$ and the additional expanded controls from the 8th grade questionnaire are represented by the vector $\mathbf{E}_i$. We pool data across the two cohorts and include a fixed effect for cohort, δ_t. ϵ_{ijt} is the random individual-level error term.

We obtain model-based estimates of the standard deviation of high school effects ($\hat{\sigma}_{u_j}$) for a given student outcome. We also extract school-specific random effects ($\hat{u}_j$), which represent empirical Bayes (EB) estimates of each school's impact on the outcome of interest, meaning they have been shrunken based on their estimated reliability. We fit complementary models in which we replace the school random effects with school fixed effects. Variance estimates based on these fixed-effect estimates (FE) are biased upward, while variance estimates based on EB random effect estimates are biased downward, with the true variance, $\sigma^2_{u_j}$, bounded between these two approaches (Raudenbush & Bryk, 2002). We focus our presentation on the random effects model-based estimates of the standard deviation, which we see as conservative. In some cases, for reference, we also provide results from the estimated fixed effects, which, as expected, are larger.

Because of the expanded data available to us, we are able to control for a richer set of lagged covariates than is typical in the value-added literature. Deming (2014) finds that nonexperimental estimates of school value-added that control for students' prior test scores and demographic characteristics closely approximate estimates of school effects from experimental designs. Our models control for similar characteristics (race/ethnicity, gender, FRPL status, English learner status, special education status, a cubic polynomial of math test scores, and a cubic polynomial of attendance rate). However, we also include our set of expanded controls from student questionnaire items, including their parents' level of education, a variable not often found in administrative data. Ideally, we would condition on two years of prior test scores; however, we do not have 7th grade test score data for the earlier cohort in our sample. We assess the sensitivity of our results by estimating models on the 2004 cohort only, including 7th grade English Language Arts (ELA) scores with their higher-order terms in addition to the other lagged covariates. We estimate very high correlations between these value-added estimates and estimates that include only one prior year of test scores (r=0.97 for four-year college graduation and r=0.96 for earnings).

Because we are particularly interested in schools' impacts on students from low-income families, we estimate school effects specifically for them in two ways. First, we restrict our sample to only low-income students and estimate models on this sample. From this, we are able to obtain model-based estimates of the standard deviation of school effects for low-income students. All of our cross-school comparisons occur within this sample of low-income students. Second, to compare effects for low- and higher-income students in the same schools, we fit complementary models using the whole analytic sample, fully interacting all terms with our indicator of low-income status. From these models, we obtain school-by-income random effects, and we explore how these effects are correlated within school.

Our second research question concerns whether a school's impacts on short-term measures explain the variation in a school's effects on longer-term outcomes. We explore this in two ways. First, we fit value-added models as described previously, but with each of the four near-term measures as the outcome.[5] We estimate correlations between schools' effects on these measures with their effects on the two longer-run outcome measures. These correlations provide descriptive evidence of whether schools with larger than expected effects on longer-run outcomes are also those that improve test scores, improve students' attendance, promote college plans, or support

[5] Our value-added models control for lagged attendance but do not control for lagged suspensions because that information is not available in the state data.

students to stay academically on-track. Given the imperfect reliability of these measures, our estimates of these correlations are likely biased downwards. Following Kraft (2019), we disattenuate our estimated correlations for measurement error using an approach analogous to Spearman's (1904) correction. We first estimate the reliability of each of the school effects for each of the six outcomes using the following formula:

$$r_{u_j u_j} = \frac{\sigma_{u_j}^2}{\sigma_{u_j}^2 + \sigma_{\epsilon_j}^2} \qquad \text{Eq. (2)}$$

We use the model-based estimates of the variance ($\hat{\sigma}_{u_j}^2$) for each outcome, and we approximate $\sigma_{\epsilon_j}^2$ as the average of the squared standard errors of the predicted EB school effects. We then calculate adjusted correlations using the Spearman correction, multiplying the estimated correlation by the inverse of the square root of the product of the reliability of each of the outcomes.[6]

We also address our second question more directly with a model-based approach, including these short-term measures as covariates in our value-added models. Doing so allows us to examine whether, after conditioning on students' prior test scores and demographic and academic characteristics, these measures predict our longer-run outcomes of interest. We further examine how much of the variance in schools' effects these short-term measures explain.

Results

High Schools' Effects on Longer-Run Outcomes for Low-Income Students

High schools differ substantially in their estimated effects on four-year college graduation and earnings. For a low-income ninth-grade student in 2003 or 2004 with sample average characteristics, the probability of four-year college graduation ranges from 4% to 26% based solely

[6] The formula is as follows: $r_{xy}^* = \frac{\hat{r}_{xy}}{\sqrt{r_{xx} r_{yy}}}$. We calculate the adjusted correlations using the reliabilities calculated from Eq. 2. There are two caveats to note regarding this adjustment. First, the estimated reliability likely does not account for all the sources of error in these measures; consequently, even the disattenuated correlations may be biased downward (Boyd et al., 2013). Second, the adjustment assumes that the errors in each measure are uncorrelated, an assumption likely violated given that these effects are estimated using the same students across outcomes, resulting in upwardly biased estimates of the disattenuated correlations (Kraft, 2019).

on the high school they attended. The difference in the probability of four-year college graduation associated with attending a school at the 20th percentile compared to a school at the 80th percentile of the distribution is 5.8 percentage points (pp). In Table 2, we present estimates of this school-level variation from our models. For low-income students, the model-based estimate of the standard deviation of the school random effects is 4.6 pp (5.9 pp for school fixed effects).

We also find differences among high schools in their estimated effects on age 30 earnings. The standard deviation of the estimated school random effects is 0.10 (0.20 for fixed effects). In other words, one standard deviation in school quality relates to a 10% difference in later earnings for low-income students. The difference in annual earnings associated with attending a school at the 20th versus 80th percentile rank of the quality distribution is 13%, or $3,600 for the average low-income student in our sample.

To facilitate comparison across the two longer-run outcome measures, we scale the estimated effects on four-year college graduation and earnings outcomes by their sample standard deviations. A one standard deviation improvement in school value-added raises low-income students' probability of four-year college graduation by 0.15 standard deviations (Table 2) and earnings by 0.11 standard deviations. Given all of the factors that influence students' later-life outcomes, these estimated school effects are quite large.

We illustrate the magnitude of these school effects in Figure 2. The first bar in each series represents the standard deviation in each outcome from an unconditional model with only a fixed effect for cohort. This suggests substantial variation across schools in students' ultimate outcomes ($\hat{\sigma}_{u_j}$=0.087 for college graduation and $\hat{\sigma}_{u_j}$=0.21 for earnings). These differences represent both selection of students to schools and schools' effects on outcomes. The remaining bars represent estimates of the standard deviations from different models, sequentially adding (a) 8th grade mathematics test scores (cubic function), (b) standard controls, (c) expanded controls excluding parental education, and (d) parental education. Our estimated school effects from the final model represents about half of the difference in outcomes across schools. Confirming the results discussed below and supporting the validity of our value-added models, estimates from models with standard controls are quite similar to those with expanded controls. In particular, once we control for lagged measures of course-taking, college plans, and other resources from the 8th grade student questionnaire, adding measures of parental education provides essentially no additional information. Given the critical role that parental education plays in students' educational attainments and life outcomes, the fact that including it in our models has minimal effect suggests that we have sufficiently accounted for sorting of students to schools.

Relationships Between Schools' Effects on Short-Run and Longer-Run Outcomes

We next explore whether schools' impacts on short-term measures predict their impacts on longer-term outcomes. In particular, we look at schools' success in improving students' 10th grade test scores, promoting academic progress, improving student attendance, and promoting college aspirations. In Table 3, we present both raw and disattenuated correlations between schools' effects on short-term and longer-term measures. Disattenuated correlations are presented above the diagonal and raw correlations below, with the estimated reliabilities on the diagonal; we report disattenuated correlations in the text. The first two rows illustrate the disattenuated relationship between each of these measures and our two longer-term outcomes, while the last three show the relationships among schools' effects on the short-term measures. In nearly all cases, correlations are positive, substantively meaningful, and statistically significant.

Looking first at the interrelationships among the short-term measures, we find positive correlations among schools' effects on all measures. However, many of the correlations are relatively modest, suggesting that the short-term measures are capturing somewhat different dimensions of schooling. In particular, schools that are most effective in raising 10th grade test scores are not necessarily the schools that are most effective in promoting other short-term outcomes.

Looking at the longer-run outcomes, we find that schools that promote four-year college graduation rates more tend to be those that improve students' test scores and college aspirations.

In particular, schools that increase college aspirations more (net of students' 8th grade plans) appear to be setting students up to graduate from a four-year college (*r*=0.65), the strongest relationship that we observe. Schools' effects on 10th grade test scores are also positively correlated with four-year college completion outcomes (*r*=0.45). In Panel A of Figure 3, we show the relationships between schools' effects on four-year college graduation and these two short-run outcomes, with college-going aspirations on the left and test scores on the right. Estimated correlations between schools' effects on four-year college graduation and both the academic on-track index (*r*=0.23) and the attendance index (*r*=0.30) are positive and significant, but smaller in magnitude.

We find more modest, yet notable, correlations between schools' effects on earnings and their effects on three of the short-term measures. Schools that improve 10th grade test scores more than expected raise earnings more (*r*=0.39). We find a similar correlation between schools' effects on college plans and students' later earnings (*r*=0.34). These relationships are presented in Panel B of Figure 3. Interestingly, the relationship between schools' estimated impacts on the attendance index and earnings (*r*=0.37) is about the same magnitude as for test scores and earnings. Notably, this relationship is also somewhat stronger than the relationship between attendance and four-year college graduation (*r*=0.30), suggesting that this measure might capture skills that pay off in the labor market in ways that do not operate through educational attainment. The correlation between schools' effects on earnings and effects on the academic index is positive but not significant.

Correlations provide insights into the relationships between these measures but not into their relative magnitudes. In Tables 4 (for four-year college graduation) and 5 (for earnings) we also present estimates of how each of our short-term measures individually predicts outcomes conditional on our standard and expanded controls (Columns 2-5). Echoing the relatively modest correlations among the short-term measures, each of these measures predicts longer-run outcomes when they are entered simultaneously in the regression model (Column 6). A one-standard-deviation difference in the test score composite is associated with an 11-percentage point difference in the probability of graduating from a four-year college (Column 2), and this magnitude is reduced only by two percentage points when controlling for the other short-term measures (Column 6). The related coefficients on the other near-term measures range from just over one percentage point for the academic index to four percentage points for the indicator of students’ college plans, conditional on the other measures.

We find strong evidence that schools that enhance both social and academic skills have bigger impacts on four-year college graduation than the sum of their individual effects would indicate. As seen in Column 7, the interactions of test scores and each short-term measure are positive, with coefficients ranging from 1.5 to 3.7 percentage points, and statistically significant. This suggests that improving test scores matters more in schools that also promote social skills, and vice versa.

Finally, the bottom rows of Table 4 show that these near-term measures explain a large share of the differences among high schools in their impacts on four-year college graduation. Our baseline value-added model with the rich set of student covariates and lagged outcomes (Column 1) accounts for just over two-thirds of the between-school variation in four-year college graduation. The remaining one-third of between-school variation includes the actual school effects and residual variation. The estimated variance of school-level effects is 0.0025. Schools’ impacts on test scores explains approximately one-third of the school effect (Column 2). Accounting for all of our shorter-term measures and interaction terms explains nearly two-thirds (Column 7).

As seen in Table 5, the pattern of results is quite similar for earnings, with one important exception. Like college graduation, the intermediate measures predict earnings alone, conditional on lagged covariates (Table 5, Columns 2-5). A one standard deviation difference in schools’ effects on 10^{th} grade test-scores is associated with a 0.24 difference in log earnings (in other words, a 25% difference in adult earnings), a sizable effect. We also see meaningful and statistically significant impacts on earnings of schools’ effects on college-going plans, academic course-taking, and attendance. The patterns across measures echo the results from the correlational analysis described above. Schools’ effects on short-term measures remain significant and are reasonably strong, independent predictors when used jointly to predict earnings at age 30. Schools' effects on these short-term measures explain about one-third of the variation in schools' effects on longer-term outcomes. However, unlike with college graduation, we see no evidence of interaction effects (Column 8). The effects of test scores on earnings do not appear to depend on schools' effects on non-test score outcomes.

Relationship Between Effects on Attainments and Earnings

We know that a four-year college degree is a pathway to higher earnings (Autor, 2014). But, do schools that improve college-going also boost students' earnings? We find that they do. In Table 3, the correlation between schools' estimated effects on earnings and four-year college graduation for low-income students is positive and relatively strong (r=0.55).

We explore more directly the extent to which schools' effects on earnings operate through improving postsecondary attainment by adding measures of two- and four-year degree completion to the regression model. In Table 5, we show that a four-year degree is associated with a 0.38 increase in log earnings (46% difference in earnings), and a two-year degree is associated with a 0.18 increase in log earnings (20% difference in earnings). But, as suggested above, the short-term measures remain significant predictors of earnings, indicating that high schools influence earnings above and beyond impacts on postsecondary educational attainments.[7]

The Role of Peer Effects

As suggested in our framework (Figure 1), one way that attending school may affect student outcomes is through the effects of peers. As seen in Table 6, our peer composite measure

[7] We also examined schools' effects on two alternative versions of our main outcomes: any college degree and average annual earnings. In each case, estimated effects on these alternative outcomes are highly correlated with effects on our preferred outcomes (r=0.93 for the college outcomes, and r=0.99 for earnings). Our results are very similar using these alternative measures. The standard deviation of schools' effects on any college degree is 0.042 (0.046 for four-year college degree), and the correlation between schools' effects on any degree and earnings is 0.47 (compared to main result of 0.55). The standard deviation of schools' effects on the alternative earnings measure is 0.11 (compared to 0.10) and the magnitude of the correlation with effects on four-year college is 0.49 (compared to main result of 0.55).

predicts longer-run outcomes, conditional on our rich set of lagged covariates. A one-standard-deviation difference in the peer composite is associated with a 3.6-percentage point difference in the probability of graduating from a four-year college (Column 1), and a one-standard-deviation difference in the peer composite is associated with a 3% difference in later earnings (Column 3).

We find that peer effects largely operate through our short-term measures. In Columns 2 and 4 of Table 6, we add our short-run measures to the models. In the case of earnings, nearly all of the relationship between the peer effect composite and earnings is explained by the short-run measures. In the case of four-year college graduation, the coefficient on the peer composite is reduced by about one-third when the short-run measures are included. This suggests that for four-year college graduation, while some of the effects of peers is explained by the short-run measures, peers matter above and beyond these near-term outcomes. However, the estimated coefficients on the short-run measures are largely unchanged when we account for peer effects (for comparison see Column 6 of Tables 4 and 5).

Effects By Family Income

We have focused so far on how schools that serve large shares of low-income students affect outcomes for these low-income students. However, knowing how they affect the relatively more affluent students in the same school is important – are these schools just more effective overall, or do they disproportionately improve outcomes for low-income students? Again, given that we focus on schools that serve large shares of low-income students, the "higher-income" students in our sample are not among the most advantaged in the state.

In general, we find quite similar patterns for low- and higher-income students. Schools' estimated effects on both longer-run and shorter-run outcomes (Table 2) are quite similar, particularly in relation to the sample standard deviations for each group. Furthermore, in Table 7 we present estimated correlations in schools' effects between these groups. Effects on four-year college graduation are reasonably consistent for the two groups (r=0.74), while the correlation between effects on earning is somewhat lower (r=0.59). The adjusted correlations on the shorter-term measures are much higher, ranging from 0.90 on plans to attend four-year college and the academic index to 0.95 on the attendance index.

Effects Among CTE Schools

As noted in our framework, school priorities and resources affect outcomes. Massachusetts has a robust set of Career and Technical Education (CTE) schools that serve relatively large shares of students from low-income families. These schools explicitly have different priorities than comprehensive high schools. The 27 CTE schools in our sample serve approximately 25% of the low-income students.[8] The four-year college graduation rate among these students is 5%, substantially lower than among low-income students not in CTE schools (12.5%), while the share who complete two-year degrees is nearly identical (5.6%). Median earnings are approximately $1,000 less for low-income students in CTE schools than for those not in CTE schools.

In contrast to non-CTE schools, the impact of CTE schools on later earnings does not run through educational attainments. In Figure 4, we show the relationship between high schools' estimated effects on our two longer-run outcomes, with non-CTE high schools denoted in black and CTE schools in gray. All the CTE schools in the sample have below average effects on four-year college graduation (i.e., left of the vertical dashed line). Their effects on earnings, however, range across the distribution. In contrast, there are very few non-CTE high schools that have above average impacts on earnings (above the dotted line) without also having above average impacts on four-year college graduation.

Even with these differences, the associations between the shorter- and longer-term outcomes are directionally consistent across these two samples of schools. In Table 8, we present regression results separately for CTE and non-CTE schools from our main model that includes all short-run outcomes. For four-year college graduation (Columns 1 and 2), the magnitudes (except the academic on-track index) are meaningfully smaller; this is unsurprising given the much lower

[8] Students in non-CTE schools may also participate in CTE programming; however, they do so in the context of a comprehensive high school rather than a CTE high school.

college graduation rates for students from CTE schools. Nonetheless, all the measures remain positive and statistically significant for CTE schools. For earnings, the coefficients on the short-term measures are quite similar in magnitude. The one exception is a larger relationship between attendance and earnings in the CTE school sample (0.11 vs. 0.07).

Threats to Validity

Student Sorting

The central threat to validity is that, even after controlling for observed student characteristics, we have not sufficiently accounted for the sorting of students to schools in ways that are related to outcomes. We leverage our rich set of controls to provide evidence on this point.

First, we compare estimated school effects from models with and without the expanded controls from the student questionnaire data. The expanded controls are important predictors of the outcomes we examine, explaining about 10% of the between-school variation in outcomes. Nonetheless, we find very high correlations between estimated school effects on earnings (r=0.993) and four-year college graduation (r=0.989) with and without these controls. In other words, adding these expanded controls, including student-reported parent education, does not affect estimates substantially. While we cannot completely rule out potential bias from omitted variables, missing predictors would have to be much stronger than the important ones that we include to generate substantial bias (Altonji, Elder, & Taber, 2005; Frank et al., 2013; Oster, 2019).

A second piece of validity evidence comes from the reasonably high intertemporal stability of these effect estimates. These correlations may reflect, in part, some common unmeasured experiences of students in a given cohort in a given school. We re-estimate our main models in a sample of students in the ninth-grade cohorts of 2005 and 2006.[9] The estimated school effects are generally quite stable across cohorts, as shown in Table 9. The disattenuated correlations range

[9] Data limitations in the later years mean that the student-survey controls are not identical with those in our main analyses (e.g., we do not have information on parents' educational attainments in these cohorts), so we use a more limited but common set of survey items in these models.

from 0.72 for the academic index to 0.96 on four-year college graduation.[10] Given that schools may improve their effectiveness at different rates, and even at different rates across each of these measures, we would not expect perfect correlations in school effects across years. Nonetheless, we interpret the general stability of school effects across cohorts as evidence supporting their validity.

Student Mobility and Attrition

In our intent-to-treat-style analysis, student mobility across schools in the state is not a concern. And, mobility out of the state's public-school system does not matter for four-year college graduation, as we observe outcomes for all students who attend a Massachusetts public high school in 9th grade. However, our analyses looking at short-term outcomes and potentially earnings are more sensitive. Students who leave the state may be less likely to return to work in Massachusetts. And, we only observe 10th grade measures for students who continued in the state public school system. Of particular concern for our analyses are students who did not take the 10th grade MCAS examinations and did not complete the associated 10th grade questionnaire, as data from these sources form several of our key near-term measures.

The central concern here is any endogenous attrition – in other words, if schools affect which students drop out or move out of the state system, our estimates may be biased. Approximately 23% of low-income students in our sample do not have MCAS scores, primarily because they dropped out of school or transferred out of the state public school system before the end of 10th grade. On average, such students have substantially lower 8th grade math scores than students who persisted (-1.04 as compared to -0.63). If schools are differentially effective in having students persist to 10th grade, our estimates may be biased.

To examine the extent to which this latter concern is a threat to the validity of our findings, we correlate the share of low-income students in a high school who are missing MCAS scores and the share of low-income students in a high school who dropped out with the schools' estimated

[10] We do not examine the correlation across cohorts in earnings because we lack sufficient years of earnings data for the later cohorts.

effect for low-income students. We find negative correlations (Table 10), indicating that high schools that consider have the most attrition do least well in promoting the shorter- and longer-term outcomes we. As such, if anything, student mobility appears to attenuate the differences in estimated effects across schools.

Discussion

We advance the literature on high school quality by examining schools' effects on students' longer-run educational attainments and earnings. Using longitudinal data on 9th grade students in public high schools in Massachusetts serving large shares of low-income students in 2003 and 2004, we show that high school quality matters a great deal for students' probability of four-year college graduation and labor market earnings at age 30. We also find that schools' impacts on all four of our short-term measures predict longer-run outcomes independently. In particular, schools that have larger than expected effects on students' 10th grade test scores and college-going plans also have larger than expected effects on educational attainments and earnings.

These findings have several implications for education research, policy, and practice. Perhaps most importantly, some high schools are much more successful at improving longer-term outcomes for students living in poverty than others. One standard deviation in school quality relates to a 4.6 percentage point difference in four-year college graduation and a 10% difference in later earnings. This is an optimistic story, as it suggests that high schools can make substantial differences in the life outcomes of low-income students. Which high school a low-income student attends has real implications for socioeconomic mobility. Furthermore, the schools that improve outcomes the most are not simply those that serve relatively few economically disadvantaged students. In Figure 5, we plot each school's effect on the longer-run outcomes against the share of students from low-income families in the school. Across the distribution, there are schools that improve outcomes more and those that do not serve students well. In other words, some schools that serve high proportions of economically disadvantaged students have substantial positive impacts on their later life outcomes.

We also find strong evidence supporting our multi-dimensional model of school effectiveness. We highlight three key results. First, schools influence students' longer-run outcomes through pathways beyond tested academic knowledge and skills. We know that a wide range of academic and social skills are important for future educational attainments and pay off in the labor market. We document that schools that improve proxy measures of these skills more tend to have larger impacts on longer-run outcomes. In particular, schools that improve students' academic skills and knowledge in tested areas and improve their college-going plans (what some scholars have called "college press") support students' educational attainments and labor market earnings. But, high schools' effects on all of the academic and social skill measures we examine predict schools' impacts on longer-run outcomes. This has important implications for school improvement and accountability, which we discuss below.

Second, our results suggest that improvement efforts aimed at a particular outcome could in fact enhance multiple outcomes. In other words, schools that improve students' social skills more tend to see added gains from test-score improvements (and vice versa). This suggests that schools should not treat skills in isolation and that a holistic approach to students' academic and social-emotional development appears to be more effective. This is particularly relevant for school and district leaders focused on school improvement efforts.

Third, schools can take different paths to improving students' life chances. For example, some CTE schools in our sample were effective at raising students' longer-run labor market earnings without larger than expected effects on college completion, unlike the non-CTE schools in our sample. This finding aligns with past research, using experimental and quasi-experimental designs, demonstrating that CTE schools increase students' labor market earnings but have no impacts on college attainment (Brunner, Dougherty, & Ross, 2021; Kemple & Willner, 2008; Page, 2012). It also suggests that school priorities are essential inputs into any system designed to measure school effectiveness. Examining school effectiveness in relation to school priorities is an important avenue for future research.

For policymakers and practitioners, these findings suggest possible next steps in designing measures of school effectiveness for accountability and improvement. From a policy perspective, it is helpful to know that schools' impacts on short-run measures predict their effects on longer-run outcomes. Ultimately, education stakeholders care about students' longer-run success, but assessing schools on these longer-run outcomes is largely impractical. We provide some evidence of the validity of shorter-term outcomes as markers of school effectiveness.

Related, while test scores have formed the cornerstone of accountability policies for schools and teachers, our results suggest that examining school effects across a broader range of outcomes beyond test scores may better account for the multiple ways in which schools influence students' longer-run life outcomes. Our results are consistent with ongoing policy efforts at the federal level and in many states to move beyond accountability systems based only on test scores. Further work should be done to develop and refine measures of students' experiences in schools, including, for example, student behavior, engagement, and college plans. Identifying measures that are difficult to manipulate is an important avenue for future research. In this regard, markers of college access, like enrollment, might prove promising; we find that schools' impacts on four-year college enrollment are strongly correlated with their impacts on four-year college completion (r=0.89) and with earnings (r=0.59).

But, accountability is not the only way in which a broader understanding of schools' impacts can inform practice. Instead, many of the measures we describe, or improved versions of them supported by state and district data efforts, can provide diagnostic information to school and district leaders to foster ongoing improvement. One clear example here is schools' impacts on students' college plans, which seems to be an important pathway through which schools improve longer-run outcomes. While survey measures of college plans may be too easy to manipulate to be valid indicators in an accountability system, they can be valuable metrics to help educators gauge their success in supporting students.

In short, developing systems of school support and accountability that account for the various ways schools may influence students' longer-run outcomes is important. One key feature of any such system – whether of formative indicators to aid school improvement or more formal accountability structures – is that the end goal should be to support not only changes in the measures themselves but the development of students' underlying skills, abilities, and capacities. Supporting schools to improve the broader social and academic skills of low-income students has real potential to pay off in longer-term outcomes, reduce persistent opportunity gaps, and advance socioeconomic mobility.

References

Allensworth, E. M., & Clark, K. (2020). High school GPAs and ACT scores as predictors of college completion: Examining assumptions about consistency across high schools. *Educational Researcher*, *49*(3), 198-211. https://doi.org/10.3102/0013189x20902110

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed an unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, *113*(1), 151-184.

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*, *34*(2), 275-318.

Angrist, J. D., Hull, P. D., Pathak, P. A., & Walters, C. R. (2017). Leveraging lotteries for school value-added: Testing and estimation. *The Quarterly Journal of Economics*, *132*(2), 871-919. https://doi.org/10.1093/qje/qjx001

Angrist, J. D., Hull, P. D., Pathak, P. A., & Walters, C. R. (2021). Race and the mismeasure of school quality (Working Paper 29608). National Bureau of Economic Research. http://www.nber.org/papers/w29608

Angrist, J. D., Hull, P. D., & Walters, C. R. (2022). Methods for measuring school effectiveness (Working Paper 30803). National Bureau of Economic Research. https://www.nber.org/papers/w30803

Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science*, *344*(6186), 843-851.

Betebenner, D. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, *28*(4), 42-51. https://doi.org/10.1111/j.1745-3992.2009.00161.x

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2005). Explaining the short careers of high-achieving teachers in schools with low-performing students. *American Economic Review*, *95*(2), 166-171. https://doi.org/10.1257/000282805774669628

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2013). Measuring test measurement error: A general approach. *Journal of Educational and Behavioral Statistics*, *38*(6), 629-663. https://doi.org/10.3102/1076998613508584

Brighouse, H., Ladd, H. F., Loeb, S., & Swift, A. (2016). Educational goods and values: A framework for decision makers. *Theory and Research in Education*, *14*(1), 3-25. https://doi.org/10.1177/1477878515620887

Brunner, E. J., Dougherty, S. M., & Ross, S. L. (2021). The effects of career and technical education: Evidence from the Connecticut technical high school system. *The Review of Economics and Statistics*, 1-46. https://doi.org/10.1162/rest_a_01098

Chetty, R., Friedman, J. N., Hendren, N., Jones, M. R., & Porter, S. R. (2018). The opportunity atlas: Mapping the childhood roots of social mobility (Working Paper 25147). National Bureau of Economic Research. https://www.nber.org/papers/w25147

Chetty, R., Hendren, N., Kline, P., Saez, E., & Turner, N. (2014). Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility. *American Economic Review*, *104*(5), 141-147. https://doi.org/10.1257/aer.104.5.141

Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2020). Race and economic opportunity in the United States: An intergenerational perspective. *The Quarterly Journal of Economics*, *135*(2), 711-783. https://doi.org/10.1093/qje/qjz042

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2009). Are teacher absences worth worrying about in the United States? *Education Finance and Policy*, *4*(2), 115-149. https://doi.org/10.1162/edfp.2009.4.2.115

Deming, D. J. (2011). Better schools, less crime? *The Quarterly Journal of Economics*, *126*(4), 2063-2115. https://doi.org/10.1093/qje/qjr036

Deming, D. J. (2014). Using school choice lotteries to test measures of school effectiveness. *American Economic Review*, *104*(5), 406-411. https://doi.org/10.1257/aer.104.5.406

Deming, D. J., Hastings, J. S., Kane, T. J., & Staiger, D. O. (2014). School choice, school quality and postsecondary attainment. *American Economic Review*, *104*(3), 991-1013. https://doi.org/10.1257/aer.104.3.991

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, *16*(12), 939-944.

Duncan, G. J., & Murnane, R. J. (Eds.). (2011). *Whither opportunity? Rising inequality, schools, and children's life chances*. Russell Sage Foundation.

Dynarski, S. M., Hemelt, S. W., & Hyman, J. M. (2015). The missing manual: Using National Student Clearinghouse data to track postsecondary outcomes. *Educational Evaluation and Policy Analysis*, *37*(1S), 53S-79S. https://doi.org/10.3102/0162373715576078

Frank, K. A., Maroulis, S. J., Duong, M. Q., & Kelcey, B. M. (2013). What would it take to change an inference? Using Rubin's causal model to interpret the robustness of causal inferences. *Educational Evaluation and Policy Analysis*, *35*(4), 437-460. https://doi.org/10.3102/0162373713493129

Gutmann, A. (1999). *Democratic education*. Princeton University Press.

Harris, D. N. (2010). How do school peers influence student educational outcomes? Theory and evidence from economics and other social sciences. *Teachers College Record*, *112*(4), 1163-1197.

Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, *91*(2), 145-149.

Hoxby, C. M., & Avery, C. (2013). The missing "one-offs": The hidden supply of high-achieving, low-income students. *Brookings Papers on Economic Activity*, *2013*(Spring), 1-65. https://doi.org/10.1353/eca.2013.0000

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, *126*(5), 2072-2106.

Jackson, C. K., Porter, S. C., Easton, J. Q., Blanchard, A., & Kiguel, S. (2020). School effects on socioemotional development, school-based arrests, and educational attainment. *American Economic Review: Insights*, *2*(4), 491-508. https://doi.org/10.1257/aeri.20200029

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*, *88*(1), 56-82. https://doi.org/10.1177/0038040714562006

Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher*, *42*(6), 304-316. https://doi.org/10.3102/0013189x13495087

Kaushal, N. M., Magnuson, K. A & Waldfogel, J. (2011). How is family income related to investments in children's learning? In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 187-206). Russell Sage Foundation.

Kemple, J. J., & Willmer, C. (2008). *Career Academies: Long-term impacts on labor market outcomes, educational attainment, and transitions to adulthood.* MDRC.

https://www.mdrc.org/publication/career-academies-long-term-impacts-work-education-and-transitions-adulthood

Kraft, M. A. (2019). Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources*, *54*(1), 1-36.

Lafortune, J., Rothstein, J., & Schanzenbach, D. W. (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics*, *10*(2), 1-26. https://doi.org/10.1257/app.20160567

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*(1), 37-62. https://doi.org/10.3102/01623737024001037

Linn, R. L. (2008). Methodological issues in achieving school accountability. *Journal of Curriculum Studies*, *40*(6), 699-711. https://doi.org/10.1080/00220270802105729

Lloyd, T., & Schachner, J. N. (2020). School effects revisited: The size, stability, and persistence of middle schools' effects on academic outcomes. *American Educational Research Journal*, *58*(4), 748-784. https://doi.org/10.3102/0002831220948460

Loeb, S., Christian, M. S., Hough, H. J., Meyer, R. H., Rice, A. B., & West, M. R. (2018). School effects on social-emotional learning: Findings from the first large-scale panel survey of students (Working Paper). Policy Analysis for California Education. https://files.eric.ed.gov/fulltext/ED591089.pdf.

Loeb, S., Soland, J., & Fox, L. (2014). Is a good teacher a good teacher for all? Comparing value-added of teachers with their English learners and non-English learners. *Educational Evaluation and Policy Analysis*, *36*(4), 457-475. https://doi.org/10.3102/0162373714527788

Ma, J., Pender, M., & Welch, M. (2020). *Education pays 2019: The benefits of higher education for individuals and society* (Trends in Higher Education Series). College Board. https://research.collegeboard.org/media/pdf/education-pays-2019-full-report.pdf

Mantil, A. (2021). Crossing district lines: The impact of urban–suburban desegregation programs on educational attainments. *Educational Evaluation and Policy Analysis*, *44*(1), 127-148. https://doi.org/10.3102/01623737211030504

Mehta, J., & Fine, S. M. (2019). *In search of deeper learning: The quest to remake the American high school*. Harvard University Press.

Mulhern, C. (2020). Beyond teachers: Estimating individual guidance counselors' effects on educational attainment. (Working Paper). Retrieved from: http://papers.cmulhern.com/Counselors_Mulhern.pdf

National Student Clearinghouse (2022, March). "Enrollment Coverage Workbook." *Working With Our Data*. Retrieved December 12, 2022 from https://nscresearchcenter.org/workingwithourdata/

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187-204. https://doi.org/10.1080/07350015.2016.1227711

Page, L. C. (2012). Understanding the impact of Career Academy attendance: An application of the principal stratification framework for causal effects accounting for partial compliance. *Evaluation Review*, *36*(2), 99-132. https://doi.org/10.1177/0193841X12447248

Raudenbush, S., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, *59*(1), 1-17.

Raudenbush, S., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, *20*(4), 307-335.

Reardon, S. F. (2018). Educational opportunity in early and middle childhood: Variation by place and age. (Working Paper 17-12). Stanford Center for Education Policy Analysis. https://edopportunity.org/papers/wp17-12-v201803.pdf

Reardon, S. F., Papay, J. P., Kilbride, T., Strunk, K. O., Cowen, J., An, L., & Donohue, K. (2019). Can repeated aggregate cross-sectional data be used to measure average student learning rates? A validation study of learning rate measures in the Stanford Education Data Archive. (Working Paper 19-08). Stanford Center for Education Policy Analysis. https://edopportunity.org/papers/wp19-08-v201911.pdf

Rowan, B. (2011). Intervening to improve the educational outcomes of students in poverty: Lessons from recent work in high-poverty schools. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances.* (pp. 523-537). Russell Sage Foundation.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72-101.

Xu, D., Zhang, Q., & Zhou, X. (2022). The impact of low-ability peers on cognitive and noncognitive outcomes. *Journal of Human Resources*, *57*(2), 555-596. https://doi.org/10.3368/jhr.57.2.0718-9637R2

Tables

Table 1

Student Characteristics 2003-2004 9th Grade Cohorts

	MA 9th Graders		Analytic Sample					
	All Students		All Students		Low-Income Students		Higher-Income Students	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Demographics and Prior Achievement								
Asian	0.044	0.205	0.060	0.237	0.081	0.272	0.037	0.188
Black	0.081	0.272	0.166	0.372	0.225	0.418	0.101	0.302
Hispanic	0.104	0.305	0.231	0.421	0.357	0.479	0.092	0.289
Native American/American Indian	0.003	0.053	0.003	0.056	0.003	0.055	0.003	0.056
White	0.768	0.422	0.541	0.498	0.335	0.472	0.767	0.423
Free/Reduced-price lunch	0.267	0.442	0.524	0.499	1.000	.	0.000	.
Special education	0.157	0.364	0.193	0.394	0.223	0.416	0.160	0.366
English learner	0.031	0.173	0.067	0.250	0.111	0.314	0.019	0.136
Parent completed four-year college	0.502	0.500	0.286	0.452	0.194	0.396	0.375	0.484
8th grade attendance rate	0.944	0.065	0.932	0.072	0.924	0.078	0.941	0.063
8th grade math standardized score	0.012	0.984	-0.495	0.871	-0.719	0.787	-0.250	0.893
Short-Term Outcomes								
10th grade math standardized score	0.031	0.984	-0.444	0.959	-0.695	0.906	-0.203	0.947
10th grade ELA standardized score	0.040	0.953	-0.387	1.003	-0.692	1.022	-0.094	0.892
Plan to attend four-year college	0.695	0.460	0.581	0.493	0.561	0.496	0.598	0.490
High school graduation	0.770	0.421	0.679	0.467	0.591	0.492	0.775	0.417
Two-year college enrollment	0.215	0.411	0.263	0.440	0.179	0.384	0.355	0.479
Four-year college enrollment	0.456	0.498	0.245	0.430	0.234	0.423	0.257	0.437
Longer-Run Outcomes								
Four-year degree	0.355	0.478	0.180	0.384	0.107	0.309	0.260	0.439
Two-year degree	0.061	0.240	0.064	0.246	0.056	0.231	0.073	0.261
Median earnings at age 30 (2022$)	43,781	43,298	36,089	35,464	32,477	31,033	41,296	39,493
N (Students)	151,855		50,430		26,406		24,024	

Note. The analytic sample excludes students attending exam schools and students missing 8th grade covariates. High school graduation is measured on-time, and college enrollment is measured within one year of high school graduation. The inter-quartile range (not standard deviation) is presented for median earnings. ELA = English language arts; MA = Massachusetts; SD = standard deviation.

Table 2

Random and Fixed Effect Estimates of School Effects, Raw and Scaled

	Four-Year Degree (Percentage points)		Earnings (Log 2022 dollars)	
	Low-Income Students	Higher-Income Students	Low-Income Students	Higher-Income Students
	Random Effect Estimates			
Standard Deviation	0.046	0.070	0.100	0.085
95% CI	[0.038, 0.055]	[0.059, 0.082]	[0.077, 0.130]	[0.066, 0.110]
Scaled (SD Units)	0.148	0.160	0.107	0.095
	Fixed Effect Estimates			
Standard Deviation	0.059	0.078	0.203	0.178
Scaled (SD Units)	0.191	0.177	0.217	0.199

Notes. Random effect estimates are the model-based standard deviation of the school effects from our preferred value-added model with our standard and expanded controls and cohort fixed effects. We scale the random effect results by the sample standard deviations of the outcomes for low- and higher-income students. Fixed effect estimates are from a regression model with school fixed effects, standard and expanded controls, and fixed effects. Estimates for low- and higher-income students come from separate models. SD = standard deviation.

Table 3

Disattenuated and Raw Correlations of Schools Effects on Longer-Run and Short-Term Outcomes for Low-Income Students with Estimated Reliabilities on Diagonal

	Four-year degree	Earnings	Test-score index	College plans	Academic on-track index	Attendance index
Four-year degree	0.814	0.548***	0.452***	0.647***	0.230**	0.297***
Earnings	0.414***	0.703	0.390***	0.340***	0.110	0.365***
Test-score index	0.384***	0.308***	0.887	0.292***	0.142	0.350***
College plans	0.525***	0.257***	0.247**	0.811	0.419***	0.266***
Academic on-track index	0.196**	0.088	0.127	0.358***	0.900	0.137
Attendance index	0.252***	0.288***	0.310***	0.225**	0.122	0.886

Note. N = 106. Disattenuated correlations are presented above the diagonal and raw correlations below. Disattenuated correlations are adjusted for attenuation bias due to measurement error using post-hoc predicted BLUP school random effect estimates derived from a model with only low-income students and conditioning on the standard and expanded controls and cohort fixed effect. Estimates are pooled across cohorts. Estimated reliabilities are on the diagonal and are calculated as described in the text. BLUP = Best Linear Unbiased Prediction.
*p<0.1; **p<0.05; ***p<0.01

Table 4

Regression of Four-Year College Graduation on Lagged Covariates and Short-Term Measures, Low-Income Students 2003-2004 Cohorts

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Test-score index		0.106***				0.085***	0.097***
		(0.005)				(0.005)	(0.005)
College plans			0.050***			0.038***	0.049***
			(0.003)			(0.003)	(0.003)
Academic on-track index				0.025***		0.014***	0.020***
				(0.003)		(0.003)	(0.003)
Attendance index					0.046***	0.031***	0.049***
					(0.003)	(0.003)	(0.003)
Test Score X College Plans							0.035***
							(0.003)
Test Score X Academic On-Track							0.015***
							(0.003)
Test Score X Attendance							0.037***
							(0.003)
School-Level Variation (Var(*u_j*))	0.0025	0.0018	0.0019	0.0023	0.0024	0.0014	0.0008
Proportion Explained	0.685	0.772	0.767	0.706	0.702	0.829	0.896

Note. The sample is 15,299 students in 106 high schools. All models include standard and expanded controls as described in the text and cohort fixed effect. Standard errors in parentheses.
*p<0.1; **p<0.05; ***p<0.01

Table 5

Regression of Earnings on Lagged Pre-Treatment Covariates, Short-Term Measures, and College Graduation, Low-Income Students 2003-2004 Cohorts.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Test-score index		0.236***				0.197***	0.198***	0.158***
		(0.015)				(0.015)	(0.015)	(0.015)
College plans			0.065***			0.040***	0.044***	0.023***
			(0.008)			(0.008)	(0.009)	(0.008)
Academic on-track index				0.062***		0.041***	0.041***	0.032***
				(0.010)		(0.010)	(0.011)	(0.010)
Attendance index					0.106***	0.073***	0.076***	0.060***
					(0.010)	(0.010)	(0.012)	(0.010)
Test Score X College plans							0.016	
							(0.010)	
Test Score X Academic on-track							0.001	
							(0.010)	
Test Score X Attendance							0.005	
							(0.010)	
Four-year degree								0.376***
								(0.024)
Two-year degree								0.178***
								(0.028)
School-Level Variation (Var(*u_j*))	0.0061	0.0040	0.0052	0.0061	0.0064	0.0041	0.0041	0.0038
Proportion Explained	0.735	0.830	0.774	0.739	0.722	0.823	0.822	0.835

Note. The sample includes 11,360 students in 105 schools. All models include standard and expanded controls as described in the text and cohort fixed effects. We lack information on suspensions in the earnings data, so we include only standardized test-year attendance. Standard errors in parentheses.
*p<0.1; **p<0.05; ***p<0.01

Table 6

Regression of Long-Run Outcomes on Lagged Covariates, Short-Term Measures, and Peer Effect Measures, Low-Income Students 2003-2004 Cohorts

	Four-Year Degree (n= 15,299)		Earnings (n = 11,360)	
	(1)	(2)	(3)	(4)
Peer composite	0.036***	0.022***	0.033**	0.009
	(0.006)	(0.005)	(0.014)	(0.014)
Test-score index		0.083***		0.196***
		(0.005)		(0.015)
College plans		0.038***		0.039***
		(0.003)		(0.008)
Academic on-track index		0.014***		0.041***
		(0.003)		(0.010)
Attendance index		0.030***		0.073***
		(0.003)		(0.010)

Note. All models include standard and expanded controls as described in the text and cohort fixed effects. Standard errors in parentheses.
*p<0.1; **p<0.05; ***p<0.01

Table 7

Relationship Between Estimated Schools Effects for Low-Income and Higher-Income Students in the Same High School

	Disattenuated Correlations	Raw Correlations
Four-Year Degree	0.736	0.609
Earnings	0.593	0.408
Test-score index	0.915	0.801
College plans	0.898	0.766
Academic on-track index	0.900	0.810
Attendance index	0.945	0.824

Note. N = 106. School effects estimated for low- and high-income students in the same school using fully interacted mixed effects value-added models with standard and expanded controls and cohort fixed effect. Disattenuated correlations are adjusted for measurement error as described in the text.

Table 8

Robustness of Regression of Four-year College Graduation and Earnings to Alternate Samples of Schools

	Four-Year College		Earnings	
	Non-CTE Schools	CTE Schools	Non-CTE Schools	CTE Schools
Test-score index	0.093***	0.050***	0.201***	0.175***
	(0.006)	(0.007)	(0.018)	(0.031)
College plans	0.043***	0.022***	0.038***	0.045***
	(0.004)	(0.004)	(0.010)	(0.016)
Academic on-track index	0.012***	0.012***	0.040***	0.034**
	(0.004)	(0.004)	(0.012)	(0.017)
Attendance index	0.037***	0.019***	0.066***	0.114***
	(0.004)	(0.004)	(0.012)	(0.022)
N (Students)	11,033	4,266	8,229	3,138
N (Schools)	79	27	78	27

Note. All models include standard and expanded controls and cohort fixed effects. Standard errors in parentheses. CTE = Career and Technical Education
*p<0.1; **p<0.05; ***p<0.01

Table 9

Stability of School Value-Added Estimates: Disattenuated and Raw Correlations, 2003-04 Cohorts and 2005-06 Cohorts

	Disattenuated	Raw
Four-year degree	0.965	0.807
Test-score index	0.808	0.726
College plans	0.839	0.674
Academic on-track index	0.755	0.683
Attendance index	0.634	0.564

Note. These value-added models are estimated using standard controls and a common set of survey items, which do not include parent education. Disattenuated correlations are adjusted for measurement error as described in the text.

Table 10

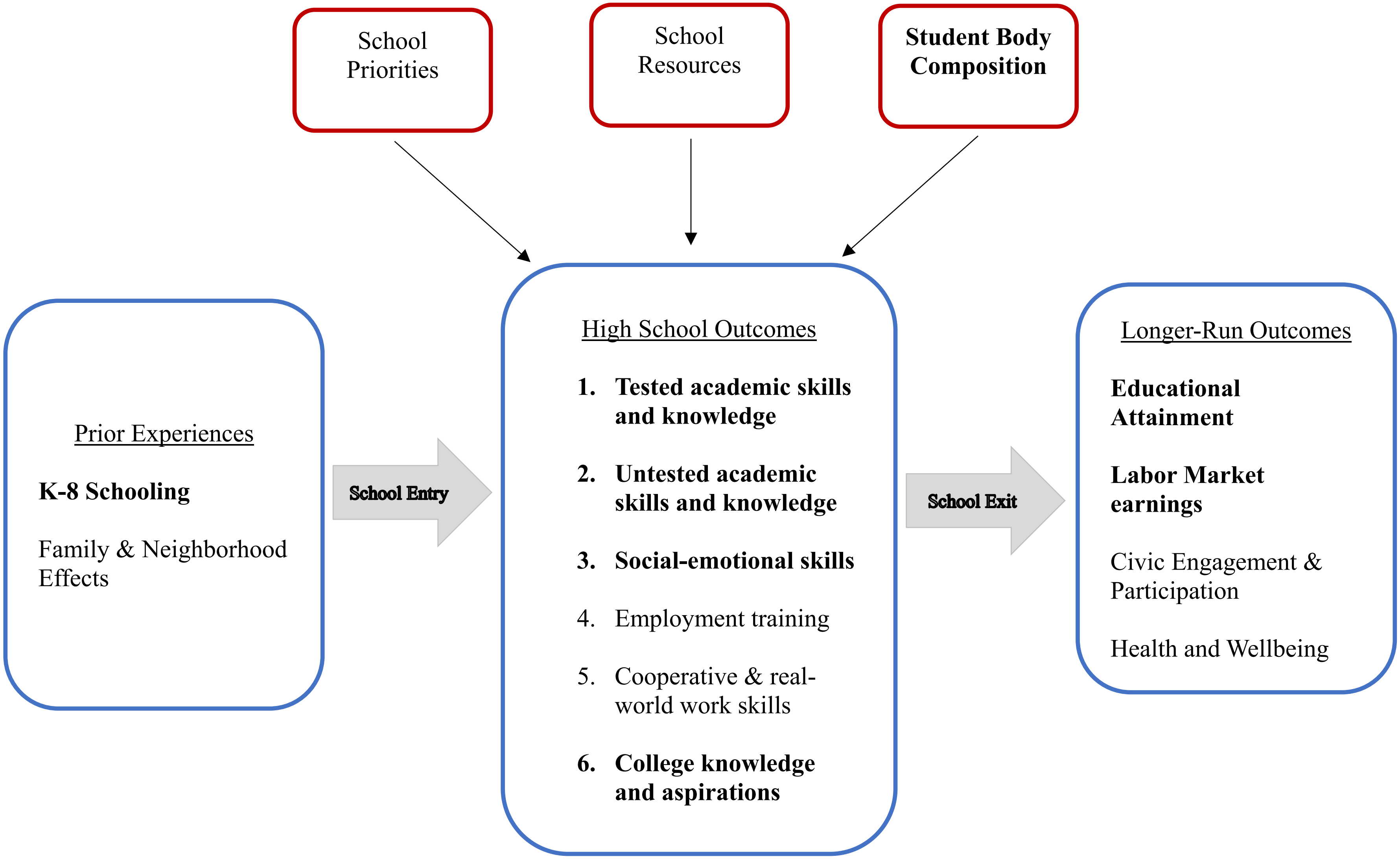
Correlations of School-Level Mobility and Attrition Measures with Estimated School Effects

	Mobility Measure	
School VA Measure	Share who did not take MCAS	Share who dropped out
Four-year degree	-0.129	-0.251
Earnings	-0.382	-0.510
Test-score index	-0.206	-0.330
College plans	0.002	-0.172
Academic on-track index	0.026	-0.108
Attendance index	-0.331	-0.314

Figures

Figure 1

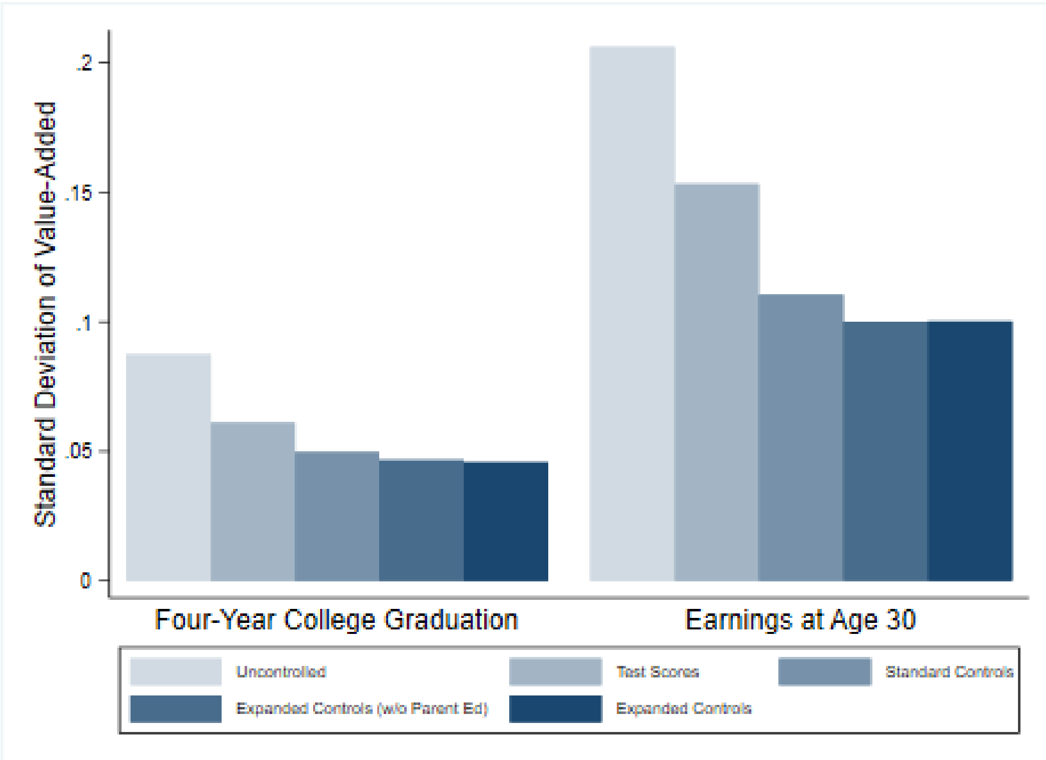
Multi-Dimensional Framework of School Effectiveness and Longer-Run Impacts



Note. Bolded items are included in our analyses.

Figure 2

Estimated School Effects from Alternative Value-Added Models

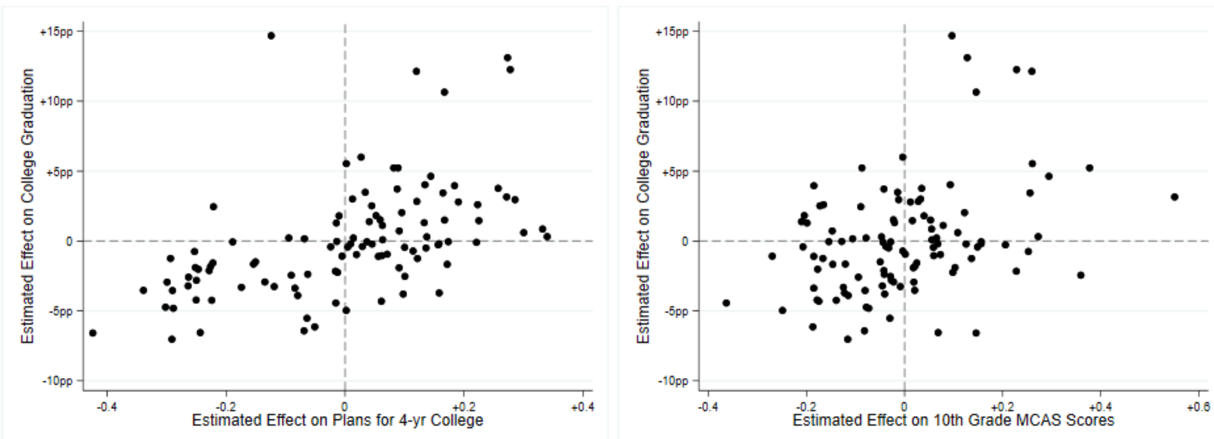


Note. The unconditional model includes only a cohort fixed effect. Controls for other models are described in the text.

Figure 3

Relationship between Schools' Estimated Effects on Longer-Run Outcomes, Four-Year College Graduation (Panel A) and Earnings (Panel B), and Short-Run Outcomes, College Plans (Left) and Test Score Index (Right)

Panel A: Four-Year College Graduation



Panel B: Earnings

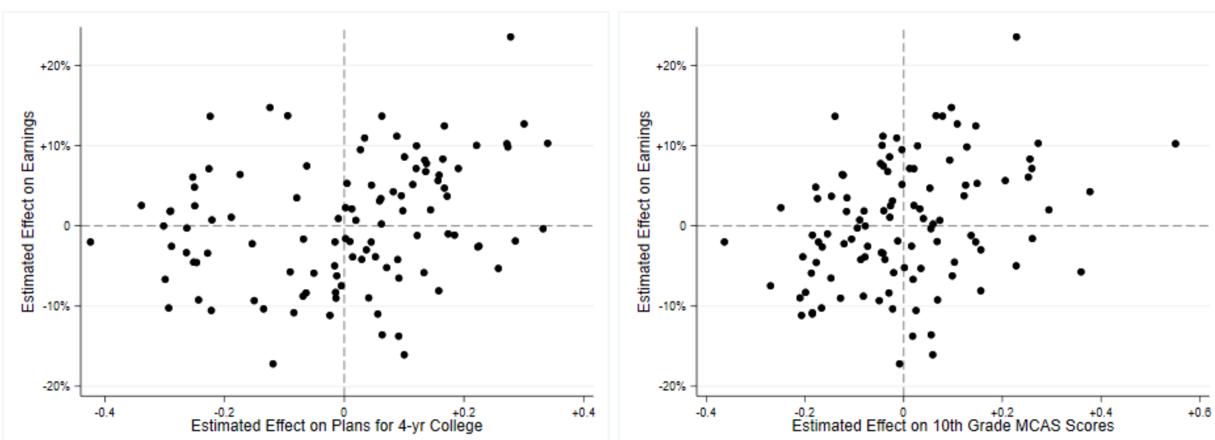


Figure 4

Relationship Between School Effects on Earnings and Four-Year College Graduation

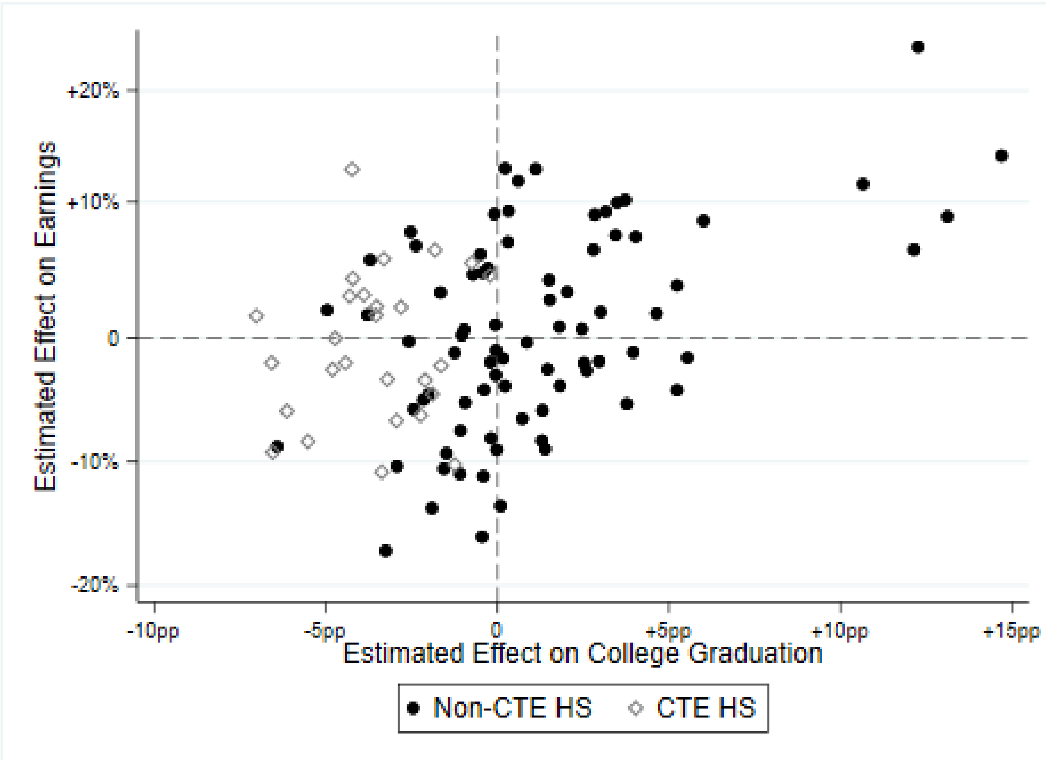


Figure 5

Relationship Between School Effects on Four-Year College Graduation (Left) and Earnings (Right) by the School Share of Low-Income Students