NBER WORKING PAPER SERIES

WHO BENEFITS FROM ATTENDING EFFECTIVE SCHOOLS? EXAMINING HETEROGENEITY IN HIGH SCHOOL IMPACTS

C. Kirabo Jackson
Shanette C. Porter
John Q. Easton
Sebastián Kiguel

Working Paper 28194
http://www.nber.org/papers/w28194

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2020

The authors thank the staff at Chicago Public Schools, particularly the Office of Social and Emotional Learning, and the University of Chicago Consortium on School Research for providing access to, and information about, the Chicago Public Schools data. This paper benefited from discussions with seminar participants at the UChicago Consortium and data management was facilitated by their archivist, Todd Rosenkranz. The authors acknowledge funding for this research from the Bill and Melinda Gates Foundation. The content is solely the responsibility of the authors. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Who Benefits From Attending Effective Schools? Examining Heterogeneity in High School Impacts
C. Kirabo Jackson, Shanette C. Porter, John Q. Easton, and Sebastián Kiguel
NBER Working Paper No. 28194
December 2020
JEL No. H0,I20,J0

ABSTRACT

We estimate the longer-run effects of attending an effective high school (one that improves a combination of test scores, survey measures of socio-emotional development and behaviours in 9th-grade) for students who are more versus less educationally advantaged (i.e., likely to attain more years of education based on 8th-grade characteristics). All students benefit from attending effective schools. However, the least advantaged students experience the largest improvements in high-school graduation, college-going, and school-based arrests. These patterns are driven by the least advantaged students benefiting the most from school impacts on the non-test-score dimensions of school quality. However, while there is considerable overlap in the effectiveness of schools attended by more and less advantaged students, it is the most advantaged students that are most likely to attend highly effective schools. These patterns underscore the importance of quality schools, and the non-test score components of quality schools, for improving the longer-run outcomes for less advantaged students.

C. Kirabo Jackson
Northwestern University
School of Education and Social Policy
Annenberg Hall, #204
2120 Campus Dr.
Evanston, IL 60208
and NBER
kirabo-jackson@northwestern.edu

Shanette C. Porter
Mindset Scholars Network
1201 Connecticut Ave. NW
Suite 300
Washington, DC 20036
shanette@gmail.com

John Q. Easton
UChicago Consortium on School Research
University of Chicago
1313 E. 60th St.
Chicago, IL 60637
jqeaston@uchicago.edu

Sebastián Kiguel
Northwestern University
School of Education and Social Policy
2120 Campus Drive
Evanston, IL 60208
skiguel@u.northwestern.edu

I Introduction

A growing body of research in the social sciences finds that schools have important causal effects on students' longer-term outcomes. For example, some charter schools increase college-going (Angrist et al. 2016; Sass et al. 2016), attending more selective schools improve educational attainment, wages, and health (Jackson 2010; Beuermann and Jackson 2020), and winning a school choice lottery may increase college-going for girls and reduce interaction with law enforcement among boys who are at highest risk of arrest (Deming et al. 2014; Deming 2011). In almost all these studies, the longer-run benefits are not fully explained by schools' impacts on test scores. Exploring mediators beyond test score impacts, Jackson et al. (2020) show that high schools' longer-run impacts reflect a combination of school impacts on test scores and socio-emotional development.[1] Despite all this evidence, questions remain about whether the benefit of attending a better school differs for better or worse prepared students, and whether schools that improve test scores benefit different students than those that improve socio-emotional development. We seek to understand two things: (a) If "effective schools" confer similar longer-run impacts on more and less *educationally advantaged* (i.e., likely to attain more years of education based on 8^{th}-grade characteristics) students, and (b) If schools that improve socio-emotional development versus test scores are similarly beneficial for more and less educationally advantaged students.

The motivations for this paper are twofold. First, even though studies are able to identify schools that improve student outcomes on average, there is relatively little evidence on the extent to which students with better or worse academic preparedness benefit equally. Because they may have more room for improvement, the least educationally advantaged students may benefit most from effective schools. On the other hand, if "*skills beget skills*" (Cunha et al., 2010), schools that are effective on average may have little impacts on the least advantaged. Despite these notions, existing empirical work on this topic is decidedly mixed.[2] Moreover, to deal with well-known selection problems these studies focus on either on a small group of oversubscribed charter schools (that use randomized admission lotteries) or a small set of elite schools (that use test score cut offs for admission). Because of the special nature of the schools examined, these studies may not generalize to a broad set of traditional schools. Also, because these studies rely on comparisons

[1] The notion that educational intervention's long-run impacts may reflect impacts on both hard skills and socio-emotional development was documented in Heckman et al. (2013) for Perry preschool, Chetty et al. (2011) for Kindergarten classrooms, Fredriksson et al. (2013) for class size, and Jackson (2018) for high-school teachers.

[2] Looking at charter schools, Angrist et al. (2012) and Walters (2018) finds that less advantaged Boston area charter applicants benefit more from attending oversubscribed charter schools, while Cohodes et al. (2020) find little evidence of this. Additionally, looking at Charter-like schools in India Kumar (2020) finds little difference between more and less advantaged students. Looking at elite schools, Oosterbeek et al. (2020) find negative effects of attending elite schools in Amsterdam for the lower-achieving students, Barrow et al. (2020) finds that selective enrollment high school in Chicago may have deleterious impacts on students from low- but not high-income homes, and Dustan et al. (2017) finds that less affluent students are more likely to dropout at elite schools than more affluent students. In contrast Shi (2020) finds larger elite school benefits for the least privileged students.

among *applicants* to these special schools (who may differ from the typical students in potential benefits) the patterns for students in these studies may be very different from those in the broader student population (Bruhn, 2020). In sum, while existing studies on this topic are *internally* valid, they may lack *external* validity. As such, the extent to which the causal impacts of attending a better school differ by academic advantage across a representative sample of schools or students in unknown. By exploring differences in the effect of attending more effective schools across all schools and all students in a large public school district, this study seeks to shed light on this issue.

The second motivation for this work is that both economists and social psychologists have found that differences in socio-emotional (or non-cognitive) development may explain attainment gaps by gender (Jacob, 2002) and socio-economic status (Liu 2020; Claro et al. 2016). Moreover, experimental studies in psychology find that (a) students from low-income families or who are academically lower-achieving might benefit from mindset interventions (Sisk et al., 2018), and (b) interventions that promote a sense of belonging are beneficial for the educational outcomes of minoritized (including Black and Latinx) youth (Gray et al. 2018; Walton and Cohen 2007; Walton and Cohen 2011; Murphy et al. 2020; Brady et al. 2020). As such, one might expect those schools that are effective at improving socio-emotional development to have particularly pronounced impacts for students from family disadvantage, lower-achieving students, males, and minoritized students. If so, test score measures of school quality may miss an important component of school quality for disadvantaged or minority populations. However, because few scholars have been able to identify schools that influence socio-emotional skills, whether this is true is an open question. To identify those schools that may be best able to improve the outcomes of the least educationally advantaged students is of considerable policy value. By exploring the differences across students in the impact of attending schools based on value-added to both cognitive dimensions and also socio-emotional dimensions and behaviours, we seek to shed light on this issue.

We leverage detailed data from Chicago Public Schools obtained from the UChicago Consortium on School Research. These data link K12 students to high schools and colleges along with test scores, administrative records, and self-reported survey measures of SED over time. Our project entails categorizing students as academically advantaged or not and then estimating the impacts of attending effective schools on these students. This involves three key steps: (1) First we categorize students. To this aim, we use student behaviours, survey measures, and test scores in 8^{th} grade to predict their educational outcomes years later (dropout, high school graduation, enroll in 2-year college, enroll in 4-year college) in an ordered probit model. We then use this model to create a latent educational advantaged index for each student. (2) Next we measure school effectiveness using value-added models. School value-added models seek to identify schools' causal impacts on student outcomes by comparing end-of-year outcomes across schools, while conditioning on lagged outcomes and other covariates. We estimate schools' impacts on test scores, behaviours, and socio-emotional measures in 9^{th} grade, and then combine effects across these outcomes to create an overall school effectiveness index. We validate these estimates as reflecting schools' causal impacts using within-sibling comparisons.[3] (3) Finally, we estimate the effect on educational attainment and school-based arrests of attending a more effective school for students with different levels of estimated educational advantage. We also disaggregate the effectiveness index and explore differences for schools that improve different dimensions (i.e., test scores, behaviours, survey measures of socio-emotional development).

The educational advantaged index differentiates between groups of students who are more or less likely to graduate high school, enroll in college, and attend a 4-year college. Students who are low in this index are more likely to have low 8^{th} grade test scores, low socio-emotional measures, and more absences and disciplinary incidents than those who are high on the index. Student low on this index are also more likely to come from low income homes, and be male and Black – precisely the student population that is thought to benefit the most from socio-emotional interventions. However, we find that *all* students benefit from attending more effective schools – rejecting a model in which only the most advantaged, or marginal, students benefit from better schools. Looking at short-run outcomes, one cannot reject that the marginal impact of attending a more effective school on test scores or socio-emotional development differs by educational advantage. However, attending a more effective school has much larger marginal effects on the behaviours (attendance and discipline) at the bottom of the educational advantage distribution. This may reflect (a) larger benefits for less-advantaged student and/or (b) the fact that these behaviours are relatively rare events for the most academically oriented.

Looking at the longer-run outcomes, those at bottom of the educational advantage distribution benefit the most from attending more effective schools (both in absolute and relative terms). Specifically, for those in the bottom decile of the distribution attending a school at the 85^{th} percentile of the effectiveness distribution versus one at the median is associated with a 3.4 percentage-point increase in high school graduation, a 2.2 percentage-point increase in college-going, and a 2.1 percentage-point reduction in being arrested – all statistically significant at the 1 percent level. The corresponding estimates for those in the top decile is a 0.6 percentage-point increase in high school graduation, a 1.9 percentage-point increase in college-going, and a 0.1 percentage-point reduction in being arrested– many statistically significant only at the 10 percent level.

Next we examine mechanisms. Looking at college type, all students are more likely to attend *some* college. However, attending a more effective school leads to increased 2-year and 4-year college going for those at the bottom of the distribution, but shifts students away from 2-year colleges toward 4-year colleges in the middle and top of the distribution. To show that this is not

[3] These models have been used extensively, and generally yield results similar to causal estimates from randomized lotteries (Deming et al. 2014; Angrist et al. 2017).

driven by race or gender differences, we document that these pattern holds both across demographic groups and also within gender and race groups. To help explain these differential impacts by educational advantage, we also look at the different components of school effectiveness. While test score impacts matter for all students, across both the educational and arrest outcomes, students at the bottom of the education advantage distribution reap particularly sizable benefits from attending schools that improve soft skills (as measured by impacts on surveys and behaviours).

Because the results suggest that the least advantaged students may benefit the most from access to effective schools, we also examine the distribution of school effectiveness by educational advantages. While we find considerable overlap in the distribution of school effectiveness for more and less advantaged students, we do find that the most advantaged students are most likely to attend highly effective schools. If the least educationally advantaged students (bottom decile) attended the same schools as the most advantaged (top decile), our estimates indicate that they would be 1.3 percentage points more likely to graduate high school, 1 percentage point more likely to attend college, and about 0.9 percentage points less likely to have a school-based arrest. While differences in school effectiveness *do not* account for most of the differences in outcomes across students with differing levels of educational advantage, the potential gains to a more equitable distribution of students across schools are economically meaningful.

By examining impacts for all schools in a district (as opposed to elite schools or charter schools) we contribute to the broader school quality literature. We demonstrate that across all public schools in a large district, all students benefit from attending more effective schools. Importantly, we show sizable increases in college going even among groups with very low college going rates – reinforcing the policy importance of access to effective schools for disadvantaged students. We also contribute to this literature by moving beyond a test score measure of effectiveness, and showing how students with varying levels of educational advantage benefit from schools that raise cognitive skills versus socio-emotional skills and behaviours. Importantly, we show how test-score measures of school quality may understate the benefits of effective schools– particularly for disadvantaged student populations.

The remainder of the paper proceeds as follows: Section II described the data used, Section III details the methods we use to categorize students and to measure school effectiveness. Section IV validates our methodology as representing causal impacts. The results are presented in Section V, and Section VI concludes.

II Data

We use administrative data from Chicago Public Schools (CPS) obtained from the UChicago Consortium on School Research. CPS is a large urban school district with 133 public (neighborhood /charter/ vocational/ magnet) high schools. CPS students in our data are 42% Black and 44%

Latinx, and 86% are from families with disadvantaged economic backgrounds. The full data-set includes cohorts of 9th-grade students who attended one of these schools between 2011 and 2017 (*n*=157,027). For high school graduation and school-based arrests we focus on the cohorts of 9^{th} graders between 2011 and 2015 (*n*=81,929), and for college outcomes, we focus on the cohorts of 9^{th} graders between 2011 and 2014 (*n*=55,347) because these students are old enough to have attended college. We only include first time 9^{th} graders to remove any sample selection biases due to grade repetition. The data are summarized in Table 1 and are discussed below.

Survey Measures: Some of our key variables are survey measures of social-emotional development (SED). The SED constructs captured by these surveys are hypothesized to be particularly important for the success of disadvantaged youth. Responses are collected by CPS on a survey administered to students in 2008-09, and then every year from 2010-11 onward. These survey items are not part of Chicago's accountability system and response rates were high (78%). However, nonresponse was higher for low-achievers (Appendix Table A1). Note that our analysis of impacts on longer-run outcomes is based on all students irrespective of survey completion. Each survey measure was comprised of several items and students responded to each item using point scales to indicate agreement (e.g., 1=Strongly disagree, to 4=Strongly agree). Rasch analysis was used to model responses and calculate a score for each student on each construct (for measure properties see Appendix Table A2). Two of the SED survey measures relate to one's relationship with others in the school. The first is Interpersonal Skills, and the second is a measure of Belonging.[4] The other three survey measures capture students' orientation toward hard work. These are Academic Effort, the perseverance facet of Grit, and Academic Engagement.[5] Following Jackson et al. (2020), we combine the interpersonal-related questions into a **Social Index** and the work-related questions into a **Work Hard Index**. To create each index we standardize each construct, compute the average of the included measures, and then standardize the index to be mean zero and unit variance.

Behavior Measures: Motivated by work showing that impacts on behaviours measure skills not well captured by test score impacts (e.g., Jackson (2018); Liu and Loeb (2019), Heckman et al. (2013), Petek and Pope (2020)), the second set of non-test score measures we use are student behaviors from CPS administrative data. These include the number of excused and unexcused absences, the number of severe disciplinary incidents (eligible for suspension), and the number

[4] **Interpersonal Skills** includes: I can always find a way to help people end arguments. I listen carefully to what other people say to me. I'm good at working with other students. I'm good at helping other people. **Belonging** includes: I feel like a real part of my school. People here notice when I'm good at something. Other students in my school take my opinions seriously. People at this school are friendly to me. I'm included in lots of activities at school.

[5] **Academic Effort** includes: I always study for tests. I set aside time to do my homework and study. I try to do well on my schoolwork even when it isn't interesting to me. If I need to study, I don't go out with my friends. **Grit** includes: I finish whatever I begin. I am a hard worker. I continue steadily towards my goals. I don't give up easily. **Academic Engagement** includes: The topics we are studying are interesting and challenging. I usually look forward to this class. I work hard to do my best in this class. Sometimes I get so interested in my work I don't want to stop.

of days a student is suspended, in each grade. In the analytic sample, the average 9^{th} grader is absent 15.12 days and suspended 0.82 days. Approximately 7.8% of these are involved in a severe disciplinary incident. We summarize these three measures in 9^{th} grade using a **Behaviours Index**. This index is the average of standardized days absent, days suspended, and severe disciplinary incidents in 9^{th} grade. We standardize the summary measure to be mean zero and unit variance.

Test Score Measures: The "hard" skills measure in our data is standardized test scores. To allow for comparability across grades, test scores were standardized to be mean zero unit variance within grade and year among all CPS test takers. For each student we average the standardized math and English scores, and then standardize (i.e., make it mean zero with unit variance) this average to create a **Test Score Index**.

Long-Run Outcomes: A key longer-run outcome is having a school-related arrest (among those old enough to have graduated high school). These are arrests for activities conducted on school grounds, during off-campus school activities, or due to a referral by a school official. During our sample period, 3.8 percent of first time 9^{th} graders had a school-based arrest, 5.3 percent of males, and 7.9 percent of Black males. Roughly 20 percent of juvenile arrests in 2010 were school-based arrests (Kaba and Edwards, 2012), so that these have important long-term implications. Our other longer-term outcomes include high school graduation and enrollment and persistence in college. High school completion is obtained from school leaving files from the years 2010 through 2018. We define a student as having graduated high school if they are marked as leaving high school because they graduated. The high school completion rate in our data is 0.74, indicating that about 74 percent of first time 9^{th} graders in CPS graduate high school. Our second key long run outcome is enrollment in college. Our college data come from the National Student Clearinghouse (NSC) and are merged with all CPS graduates. We code a student as enrolling in college if they are observed in the NSC data within two years of expected high school graduation (2011 through 2014 cohorts only). Using this measure, 53 percent of first-time 9^{th} graders enrolled in college. We further divide college enrollment into 2-year and 4-year college. In our sample, 34 and 27 percent of students enroll in a 4-year or 2-year college within 2 years of expected graduation, respectively.

III Methods

Our analysis involves three main steps: (1) First, we calculate an educational advantage score for each student by estimating their predicted educational attainment based on a rich set of covariates using an ordered probit. We place students into deciles from least to most likely to attain more years of education. (2) Second, following Jackson et al. (2020), we identify schools that improve students' SED and test scores in 9^{th} grade. In addition, we estimate school value-added on student behaviors using the same method. We combine school effects on the different 9^{th}-grade measures - which are predictive of students' long-term outcomes - into an index of school effectiveness. (3)

Finally, we estimate the effect of attending a more effective school among students of differing educational advantage to assess who benefits from attending better schools. We also explore effects on each individual value-added dimension to shed light on whether schools that are better in some dimensions (cognitive, socio-emotional, or behaviours) are better for some students than for others.

III.1 Classifying Students

To classify students along a single dimension, we rank students by their likelihood to attain more years of education. We refer to students who are more likely to attain more years of education (based on observed characteristics *before* entering high school) as more educationally advantaged. To classify students, we exploit the fact that we have a rich set of observable characteristics that may predict educational attainment and also multiple measures of educational attainment. In principle, with a single measure of educational attainment (say high school graduation) one could predict high-school completion based on observed covariates in 8^{th} grade. However, because some characteristics may matter more for higher levels of education (such as 4-year college attendance) it is helpful to model the relationship between these covariates and 4-year college going also. If the underlying education advantage predicts both high-school completion and college-going (or any other educational attainment level), one can model a student's underlying educational advantage (in a way that will predict multiple educational attainment margins) using a rank-ordered probit.

The basic idea is that some underlying educational advantage, y^*, is a linear function of observable characteristics X so that $y^* = X\pi + \varepsilon$. Individuals with higher levels of educational advantage attain higher levels of education, where there are some unobserved thresholds between education levels. That is, for all individuals i

$$y_i = \begin{cases} \text{No High School} & y_i^* \leq \mu_1 \\ \text{Graduate high School} & \mu_1 > y_i^* \leq \mu_2 \\ \text{Attend a 2-Year College} & \mu_2 > y_i^* \leq \mu_3 \\ \text{Attend a 4-Year College} & y_i^* > \mu_3 \end{cases}$$

The probability of observing outcome $y_i = k$ is then $Pr(y_i = k) = Pr(\mu_{k-1} < X\pi \leq \mu_k)$. The probability of observing the data is the product of these probabilities across all individuals i. Assuming a normally distributed error term, we solve for the set of estimates $(\hat{\pi}, \hat{\mu}_{k-1}, \hat{\mu}_{k-1}, \hat{\mu}_{k-1})$ that are most consistent with the observed data by estimating an ordered probit model by maximum likelihood.

Our predictors of the education outcomes include measures of lagged test scores (quadratics of 8^{th} grade math and ELA), 8^{th} grade survey measures, and lagged behaviors[6]. We also include demographics (lunch status, race, gender, and interactions between race and gender). Once the

[6] Because these variables do not have a lot of variation in early grades, we include an indicator for being in the top quartile of absences in 8th grade and an indicator for having any severe disciplinary incidents in 7th or 8th grade

parameter estimates have been estimated, we take the fitted values of latent variable, $X\hat{\pi}$, as our estimated latent educational advantage. *Note that, we use leave-year-out models to avoid mechanical correlation between our predicted and actual education levels for each student i.* As such, each student's predicted educational advantage index is based on the relationship between covariates and educational attainment in *other* cohorts. However, to show the relationship between the advantage index and the observable covariates we present the coefficient estimates from the ordered probit model for the full sample in Appendix Table A3.

Differences in Incoming Attributes by Educational Advantage

To shed light on how the attributes of students with high and low educational advantage differ, we present summary statistics for the top and bottom deciles of the education advantage distribution in the middle and right panels of Table 1. This categorization captures important differences between students, both in terms of demographics and achievement. For example, the top decile contains almost three times more females than the bottom decile (69.8% versus 24.3%), about 8 times fewer students in special education (5.6% versus 45.2%), and less than half the share of students who qualify for free lunch (43.3% versus 95.3%). The top decile has more white students than the bottom decile (23.1% versus 4.1%), more Asian students (18.6% versus 0.13%), but with lower shares of Latinx students (33% versus 43.1%) and Black students (24.3% versus 52.3%). Regarding academic achievement, students in the lowest decile have 8^{th} and 9^{th}-grade test scores more than two standard deviations below those in the top decile. Students in the top decile also have fewer absences (5.5 days compared to 34.4 days) and days suspended (.06 days vs. 2.95 days), and are involved in fewer severe incidents (.007 vs .29), relative to the lowest decile in 9^{th}-grade.

Differences in Outcomes by Educational Advantage

To illustrate the differences in our main longer-run outcomes by the latent educational advantage index, we compute the average of our key outcomes for by each percentile of the index. This is presented graphically in Figure 1. This figure highlights a few important facts. First, at the bottom of the index (the bottom 20 percent), even though about 40 percent of students graduate from high school, few (about 17 percent) go to any college, and even fewer (8 percent) attend a 4-year college. Indeed, at the very bottom decile, under 5 percent attend a 4-year college. In the middle of the distribution (between the 40^{th} and 60^{th} percentiles), the high school graduation rate is about 75%, the college-going rate is about 50% and both the 4-year and 2-year college-going rates are around 25%. As one looks to the top of the distribution (the top 20%), the high school graduation rate is above 90%. Interestingly, the 4-year college going rate increases to about 70%, while the 2-year college rate remains at 25%. That is, as one goes up the educational advantage distribution, 4-year college going increases but 2-year college going does not. Indeed at the very top of the educational advantage distribution, the 2-year college rate declines with educational advantage. Even though the index is predicted based on educational attainment, we also report the school-based arrest rate by education advantage. School-based arrests are largely concentrated among students with very low educational advantage. For the bottom 20% the arrest rate is roughly 8 percent, while for those above the median it is almost zero (0.02%). Indeed, in the very bottom decile, the arrest rate is a sizable 12.6 percent (see Table 1).

III.2 Classifying Schools

To isolate schools' causal impacts, we use value-added models to estimate schools' impacts on 9^{th}-grade SED, behaviours, and test scores. We then combine these value-added estimates to form an overall school effectiveness index.

Identifying School Impacts on SED, Behaviours, and Test Scores

We seek to isolate the causal effects of individual schools on student measure $q \in Q = \{$test scores, work hard, social, behaviours$\}$ by comparing measure q at the end of 9^{th} grade to those of similar students (with the same survey measures, course grades, incoming test scores, discipline, attendance, and demographics, all at the end of 8^{th} grade) at other schools. School j's value-added on measure q reflects how much school j increases measure q between 8^{th} and 9^{th} grade relative to the changes observed for similar students (based on all the attributes above) who attended different schools. We model the 9^{th} grade measure q of student i who attends school j with characteristics Z_{ijt} in year t as below. Z_{ijt} includes lagged measures (i.e., 8^{th} grade test scores, surveys, behaviours), gender, ethnicity, free-lunch status, and the socio-economic status of the student's census block.[7] We include school-level averages of all individual lagged outcomes. For each measure q, to obtain estimates of the impacts of attending school j in year t relative to the average school (i.e., $\theta^{VA}_{jt,q}$), we estimate (1) below, where $\upsilon_{ijt,q} = \theta^{VA}_{j,q} + \varepsilon_{ijt,q}$.

$$q_{ijt} = \beta_q Z_{ijt} + \upsilon_{ijt,q} \tag{1}$$

Where $\upsilon_{ijt,q}$ is the true student-level error from (1), $u_{ijt,q}$ is the empirical student-level residual obtained after estimation of (1). The average school-year level residuals from this regression is our estimated impact on measure q of attending a school in a given year. Where N_{jt} is the number of students attending school j in year t, this is

$$\hat{\theta}^{VA}_{jt,q} = \sum_{i \in jt}^{jt} (u_{ijt,q})/N_{jt} \tag{2}$$

[7]The census block SES measure is the average occupation status and education levels in the block.

If unobserved determinants of student outcomes are unrelated to our value-added estimates, $\hat{\theta}^{VA}_{jt,q}$ will be an unbiased estimate of the value-added of school j in year t for measure q.

When using value-added to *predict* outcomes for a particular cohort, we exclude data for that cohort when estimating value-added to avoid mechanical correlation. As in Jackson et al. (2020), these leave-year-out (or out-of-sample) predictions of school effectiveness are based on the value-added for the same school in *other years*. If the value-added in year $t+1$ were equally predictive of outcomes in year t as those in $t+4$ or any other year, then the best leave-year-out predictor for a school would be the average value-added for that school *in all other years*. However, adjacent years tend to be more highly correlated with one another than less temporally proximate years (see the top panel of Table 2). Accordingly, Following Chetty et al. (2014) we use value-added with drift which places more weight on value-added for years that are more highly correlated with the prediction year. Our leave-year-out predictor for measure q in year t is

$$\hat{\mu}_{jt,q} = \sum_{m=t-l}^{t-1} \hat{\psi}_{m,q}[\hat{\theta}^{VA}_{jm,q}] \tag{3}$$

The vector of weights $\hat{\psi}_q = (\hat{\psi}_{t-l,q}, ..., \hat{\psi}_{t-1,q}, \hat{\psi}_{t+1,q}, .., \hat{\psi}_{t+l,q})'$ are selected to minimize mean squared forecast errors (Chetty et al., 2014). A school's predicted value-added on measure q is our best prediction *based on other years* of how much that school will increase measure q between 8^{th} and 9^{th} grade relative to the improvements of similar students at other schools. We use leave-year-out predictions for all analyses, but for brevity, refer to them simply as value-added.

Creating an Overall School Effectiveness Index

In principle, each value-added measure represents school impacts on a different dimension. However, each may be sensitive to some deeper underlying school quality. To shed light on this, we correlate the school impacts across these four measures (see Table 2). The correlations between the test score and SED value-added are quite high (all above 0.4) suggesting that many schools tend to be either good in all dimensions or poor in all dimensions. Given that each of these value-addeds is measured with error, the true correlations are likely higher than this. It is notable, however, that the behaviours value-added are only weakly correlated with the others - this may reflect greater measurement errors for behaviour value-added or indicate that there is important independent variation in schools' value-added on behaviours (we will show evidence of the later). To better understand the raw correlations, we conduct factor analysis of the school effects (Table 3). The model finds that a single underlying factor explains almost all the common variation in these value-addeds.[8] This single factor is positively related to all the value-addeds indicating that

[8]The proportion explained by this factor is greater than one because the model also includes factors with negative eigenvalues.

it is related to the schools' quality across all dimensions. As such, we combine our value-addeds (work hard, social, test-scores, and behaviours) into a single index of school effectiveness. Our overall index is the predicted first principal factor of these four variables. The overall index, $\hat{\omega}_{jt}$, is a weighted average of the different value-added estimates given by (4) and represents a measure of school impacts on 9^{th} grade measures that is shared across the SED (work hard and social), test score, and behavior dimensions.

$$\hat{\omega}_{jt} = (0.09)\hat{\mu}_{jt,testscores} + (0.43)\hat{\mu}_{jt,workhard} + (0.44)\hat{\mu}_{jt,social} + (0.05)\hat{\mu}_{jt,behaviors} \tag{4}$$

We standardize the overall school quality index to be mean zero, unit variance. As we show in Section IV, the index is generally a better predictor of school impacts on longer-run outcomes than the value-added on the individual measures. However, both test score and behaviours value-added have a high level of uniqueness. This suggests that either these measures have a lot of error or that there remains independent variation in school quality captured by one or both of these dimensions. While the overall index is a good summary measure of quality, to assess the possibility that independent variation in each dimension may predict longer-run outcomes, we explore the impacts of school value-added on the individual measures in Section V.

III.3 Estimating School Effectiveness Impacts by Educational Advantage

To quantify the effect of attending a school with one standard deviation higher predicted overall effectiveness, we regress each outcome on the standardized school effectiveness index (plus controls). Specifically, where Y_{ijt} is an outcome, and $\hat{\omega}_{jt}$ is the standardized out-of-sample predicted effectiveness, we estimate the following model by OLS.

$$Y_{ijt} = \delta \hat{\omega}_{jt} + \beta_1 Z_{ijt} + \tau_t + \varepsilon_{ijt} \tag{5}$$

All variables are as defined above and τ_t is a year fixed-effect. Standard errors are adjusted for clustering at the school level.[9] To estimate differences in the marginal impacts by student type, we estimate Equation (5) separately for each decile of the estimated education advantage index.

To take the estimated impacts of effectiveness as reflecting schools' causal impacts requires that, on average, there are no unobserved differences in the determinants of outcomes between students that attend high- and low-effectiveness schools. We assess this in Section IV, where we show that value-added is unrelated to observable determinants of student outcomes, validate our estimates using quasi-random variation based on school attendance zones, and show that our estimates are

[9] Individuals with missing 8^{th} grade surveys or test scores are given imputed values. We regress each survey measure or test score on all observed pre-8^{th} grade covariates. We then obtain predicted 8^{th} grade values based on these regressions. and replace missing values with the predictions. Results are similar with and without imputation.

similar using within-family variation. These tests support a causal interpretation of our results.

IV Validating the Method

Validating the School Effectiveness Index

Before exploring differences in school impacts by educational advantage, we establish average impacts. Table 4 reports the coefficient on the educational index in a regression of various outcomes on the index and controls for the full sample. The point estimate is the difference in outcomes associated with attending a school with 1σ higher estimated effectiveness (i.e., going from a school at the median to one at the 85^{th} percentile of the effectiveness distribution). As basis for comparison, we also report the estimated effect of the value-added on the individual dimensions also. We refer to schools with a higher estimated overall school effectiveness index as more effective schools.

The top row shows that more effective schools improve 9^{th}-grade test scores, socio-emotional development in 9^{th} grade (as measured by surveys), and behaviours in 9^{th} grade. Specifically, *on average*, a 1σ increase in effectiveness increases test scores by 6.68 percent of a standard deviation, socio-emotional development by 7.9 percent of a standard deviation, and behaviours by 4.28 percent of a standard deviation. Note that social and work hard are very highly correlated (0.9) so that we combine these two SED measures into a single survey measure.[10] Not surprisingly, more effective schools also improve longer-run outcomes on average. A 1σ increase in effectiveness increases high school graduation by 1.89 percentage points, college going (within 2 years of high school completion) by 2.17 percentage points, and the likelihood of have a school-based arrest by 0.786 percentage points. All of these estimate are significant at the 1 percent level.

Our use of the index (as opposed to simply using test score impacts) is motivated by Jackson et al. (2020) showing that a combination of school impacts on test scores and surveys better predict both short and long-run outcomes than test scores alone, and Jackson (2018) showing that a combination of teacher impacts on test scores and behaviours better predict long-run outcomes than test scores alone. We show this to be the case here also. In the third row, we show the estimated impact of a one standard deviation increase in test-score value-added on these same outcomes. One can see that test score value-added *does* predict impacts on both short- and long-run outcomes, but that these impacts are smaller than those based on the effectiveness index. For each of the six outcomes, the improvement in outcomes associated with a 1σ increase in effectiveness is greater than that of a 1σ increase in test score value-added. For the longer-run outcomes, the marginal impacts of the effectiveness index are between 50 and 100 percent larger than that for test scores alone.

To shed light on the extent to which the effectiveness index outperforms all of the individual value-added, we also present the estimated impacts of value-added on the surveys, and behaviours.

[10] We provide analogous entries of Table 4 in Appendix Table A4 where the two survey measures are separated.

Note that the school impacts on social and work hard are highly correlated and they predict longer-run outcomes similarly. As such, to be concise, we combine these two SED measures into a single surveys value-added. Remarkably, for surveys, test scores, high-school graduation, and college enrollment, the school effectiveness index is more predictive of impacts than any of the individual measures. This indicates that the effectiveness index is a good summary measure of school ”effectiveness” for these outcomes. However, the behaviours value-added does appear to have more predictive power for behaviours and school-based arrests than the overall effectiveness index. This indicates that school impacts on behaviors capture some meaningful dimension of school quality that is not captured by the index which is predictive of behaviour and school-based arrests. We shed further light on this in section V.2 after presenting patterns for the overall effectiveness index.

IV.1 Testing For Selection

Because students are not randomly assigned to schools, one may worry that our effectiveness index is related to unobserved predictors of outcomes so that our estimates are biased. While there is no way to prove that the effectiveness estimates are unrelated to unobserved determinants of outcomes, we present several tests to show that this is likely satisfied in our setting.

No Selection on Observables

First, to show that our effectiveness index is likely unbiased, we show that it is unrelated to observed determinants of students' long-run outcomes. That is, we estimate the relationship between "predicted" outcomes (based on all of the observed covariates) and our effectiveness estimates. To form predicted outcomes, we regress each outcome (graduate high school, enroll in college, etc.) on *all* of the observed covariates and use the fitted values as our predicted values. To avoid mechanical correlation, we form this prediction based on the regression from other years. We then regress these predicted outcomes for each student on the estimated school effectiveness of the school they attended. More formally, where $(\hat{Y}_{ijt}|Z)$ is the predicted outcome given all the observed covariates, we estimate the following model by Ordinary Least Squares (OLS).

$$(\hat{Y}_{ijt}|Z) = \delta_p \hat{\omega}_{jt} + \tau_t + \upsilon_{ijt} \tag{6}$$

Figure 2 shows a binned scatterplot of the predicted outcome against the actual outcomes used in this paper. The predicted outcomes track actual outcomes well.[11] The parameter estimates of δ_p provides a test of whether the predicted outcome is correlated with estimated school effectiveness. If strong observable predictors of the outcomes are unrelated to our effectiveness estimates, then it

[11] The R-squared is above 0.2 for surveys, behaviours, and test scores. they are also above 0.2 for graduation, any college enrollment, 4-year college enrollment. The R-squared are somewhat lower for 2-year college going and arrests are 0.13 and 0.086 respectively. For all outcome one rejects that the leave-year-out predicted outcome is unrelated to the actual outcome at the 1 percent significance level.

is plausible that *unobservable* predictors are also – so that our estimates are unbiased. Column 4 of Table 5 reports the coefficient of effectiveness on predicted outcomes. In all models, effectiveness is not significantly related to predicted outcomes and the point estimate is very small. While this evidence supports a causal interpretation of our estimates, we also present tests of selection in unobserved dimensions below.

Attendance Boundary Instruments

Even though we show no evidence of selection on observables, one may worry about selection on unobservables. To address this, we construct instruments that remove the sorting bias that may exists when individuals chose to attend a school outside their zoned area. We propose an instrumental variables approach that instruments for the effectiveness of the school attended with the effectiveness of the residentially assigned school. This approach eliminates all selection to non-zoned schools that could have led to bias. The first stage regression is strong – yielding first stage F-statistics above 500. The two-stage-least-squares (2SLS) regressions are reported in the second column of Table 5. The OLS estimates are reported as a basis for comparison in columns 1 and 5. For all long-run outcomes the point estimates are positive and significant at the one percent level. While the 2SLS estimates are somewhat larger than the OLS, they are on the same order of magnitude. In sum, our effectiveness measure does not appear to be biased by selection on unobservables. These 2SLS estimates will only be biased if those families that attend the zoned schools tend to self-select into neighborhoods along unobserved dimensions that are correlated with school effectiveness. To rule out this possibility, we address this below.

Sibling Comparisons

To account for the possibility that families may select into neighborhoods in ways that would lead to bias in our 2SLS approach, we also estimate models that rely on within-family comparisons. For a small subset of the data we can identify siblings. That is, we can identify siblings in the data after 2015. As such, for families that have more than one sibling who were in CPS after 2015 we can make within-family comparisons. We were able to identify 13,150 families in which more than one sibling is observed in 9^{th} grade. Of these that have multiple children old enough to have graduated from high school, we have 3822 such families. For those old enough to have enrolled in college, this number falls to 1581 families.[12] We can remove any correlation with potentially confounding family characteristics by comparing students from the same family who attended different schools. This is achieved by adding a family fixed effect to our main model in equation (5). The within-family estimates are presented in columns 3 and 7 of Table 5. While the standard errors are much

[12] Because we cannot identify *all* siblings prior to 2015, these data are imperfect and incomplete. However, if we are able to find similar effects in this small sub-sample as in the broader sample, it would be compelling evidence that our estimates are not biased by family selection to neighborhoods.

larger in the sibling models, the point estimates are very similar to the OLS estimates. For college enrollment, the point estimate is no longer significant, but that for high school graduation (for which there is much more variation) remains significant at the 1 percent level, and that for in-school arrests is significant at the 10 percent level. This indicates that selection of families does not drive the estimates.

Considering all the Selection Tests Together

Taken together, we show that (a) our effectiveness estimates are unrelated to observed covariates, (b) our estimates are not driven by selection to schools outside one's attended zone, and (c) our estimates are not biased by certain kinds of families sending their children to different schools. If our results were driven by selection to schools across families, it would bias our IV results but not our sibling results. If our results were driven by selection to schools within families, it would bias our sibling results but not our 2SLS results. If there were selection (either within or across families) one would expect that strong predictors of outcomes would be related to our estimated value-added– but this is not the case. While none of these tests is dispositive in isolation, together they are compelling evidence that our estimated school impacts, and the main results, reflect true causal impacts and are not driven by any selection bias.

V Results

V.1 Heterogeneous Impacts of School Effectiveness on 9th Grade Measures

We now consider how effects vary for students with different levels of *ex-ante* educational advantage. We estimate these same regression for each measure and for each decile. To summarize these thirty regressions, we plot the point estimate and 95% confidence intervals for each estimate in Figure 3. The 9th grade measures are in the top panel. In principle, one could explore heterogeneous impacts using test score measures alone. However, if the effect heterogeneity is in dimensions other than those measured by standardized tests, one would not observe it. As such, as a starting point, we explore heterogeneous test score impacts first, and the look to other short run outcomes.

The middle top panel of Figure 3 shows the effect of attending a school one standard deviation higher in school effectiveness on students' 9th grade test scores. All students benefit from attending a more effective school. Attending a school one standard deviation higher in school effectiveness increases the 9th grade test scores of students in the lowest decile of predicted educational attainment by 0.067 standard deviations. The effect is slightly higher (but statistically indistinguishable) for students in the 5th decile, at 0.083 standard deviations. The effect is somewhat smaller (0.041 standard deviations) for students in the top decile, but the impacts of attending a more effective school on 9th grade test scores are similar throughout the educational advantage distribution.

We now turn to the survey measures. The left panel shows the effect of attending a school one standard deviation higher on the school effectiveness index on students' socio-emotional measures in 9^{th} grade by decile of predicted educational attainment. As with test scores, all students benefit from attending a more effective school. For students in the lowest and top deciles of predicted educational attainment (those most and least likely to drop out of high school), attending a school 1 standard deviation higher on school effectiveness leads to a 0.080 and 0.98 standard deviation increase in 9^{th} grade socio-emotional development, respectively. This suggests slightly larger effects for those at the top of the distribution, but these effect are statistically indistinguishable from the average effect reported in Table 4. Overall, the impacts on socio-emotional development (as measures by the surveys) are largely the same throughout the educational advantage distribution.

We report the results for 9^{th} grade behaviours in the right top panel of Figure 3. Unlike the socio-emotional and test score measures, the effect on behavior is not similar across the educational advantage distribution. Effective schools have the strongest effect on behavior for students in the lower end of the distribution. For a student in the lowest (first) decile, attending a school 1 standard deviation higher in school effectiveness improves the behavior index by 0.13 standard deviations. Meanwhile, for students in the top (tenth) decile, the behavior index only improves by 0.012 standard deviations. While each effect is statistically significantly different from zero, the impacts at the top and the bottom of the distribution are statistically significantly different from each other. One interpretation of this patters in that schools have heterogeneous effects on students across the distribution. However, it is also likely that the small impacts for students at the top of the distribution are driven by a lack of variation among these students. Specifically, students in the top decile are very unlikely to be involved in a disciplinary incident (0.007) and have a low absence rate (5.5 days compared to 34 days in the bottom decile), so that there is relatively little room for improvement. Given that the other two measures (where the variation is similar for all students) show limited evidence of differential school effectiveness impacts by educational proclivity, we take the differences for behaviors (where the is a possible truncation problem) as merely suggestive.

In sum, the short run outcomes indicate that all students benefit from attending more effective schools in all dimensions. However, there is clear evidence that students at the lower end of the educational advantage distribution experience improved behaviours (compared to those at the top). Because students at the bottom have more room for improvement, we take this as only suggestive of larger relative impacts for those at the bottom of the educational advantage distribution. To shed further light on this, we not turn to impact on longer-run outcomes.

V.2 Heterogeneous Impacts on Longer-Run Outcomes

Having shown the effect on short-run measures in 9^{th} grade, we now examine similar figures for the longer-run outcomes (the middle and lower panels of Figure 3). Looking at high school graduation, one can see that the marginal impacts of school effectiveness are much larger for students at the bottom of the educational advantage index than those at the top. Indeed, for those in the bottom decile, a 1σ increase in effectiveness increases high school completion by 3.4 percentage points (p-value<0.01) compared to only 0.6 percentage points (p-value>0.10) in the top decile. Relative to each groups' baseline level, this is about a 10 percent increase for those at the bottom of the distribution compared to a 1 percent increase for the top. One may wonder if this pattern is due to more students at the bottom being on the margin of high school graduation. To assess the second possibility, one can compare the graduation rates among the bottom 30 percent. Among the bottom 30 percent, the marginal effect on graduation rates is almost identical even though the average graduation rates go from 33 percent for the bottom decile to almost 60 percent at the 30th percentile. This suggests that the large high school graduation rate improvements for the bottom 30 percent of students is not *only* due to more of these students being marginal for high school completion. Another potential explanation is that school effectiveness leads to larger skill improvements for student lower in the educational advantage distribution. While the test score impacts and the survey impacts are very similar for all students, larger improvement in cognitive or SED are unlikely explanations, However, there are larger improvement in behaviours at the bottom of the distribution. Because this could be driven by nature of the behaviours studied, we take this as merely suggestive.

The next outcome we examine is enrolling in any college (2-year or 4-year) within two years of expected high-school completion. The point estimates generally suggest larger increases at the bottom of the distribution, but these differences are not statistically significant. That is for the bottom third, a 1σ increase in effectiveness increases college-going by about 3 percentage points (p-value<0.01) compared to about 1.9 percentage points (p-value<0.05) in the top third. Given the large differences in base rates, the differences in relative marginal impacts are sizable. The estimates indicate that for the bottom third, a 1σ increase in effectiveness increases college-going by 15 percent compared to 2.5 percent in the top third. In our setting, students in the middle third have college going rates around 50 percent, so that *if all of the differences are due to differences in the proportion of marginal students*, one might expect the largest college-going impacts for this group. The results are inconsistent with this idea; the largest increases are among the bottom third (about 3 percentage points), and the impacts for the middle and bottom thirds are largely the same (about 1.8 percentage points). Instead, the patterns are more consistent with larger skill or behavior benefits for the bottom of the distribution.

Looking at college type reveals some interesting patterns. In previous work Jackson et al. (2020) found that test score value added and surveys value added had little impacts on 2-year college going. Looking at the heterogeneous impacts provides an explanation for that null result on average. Among students who are least likely to attend any college, attending a more effective school increases 2-year college going, but among those who are more likely to attend college, attending a more effective school reduces 2-year college going. Given the increase in college-going overall, this suggests that more effective schools have particularly pronounced increases on 4-year college going among those in the top of the education advantage distribution. Indeed, the lower panel show this to be the case. For the bottom third, a 1σ increase in effectiveness increases 4-year college-going by about 2.5 percentage points compared to over 4 percentage points for the middle third and about 3 percentage points for the top third. Taken together, the results reveal an overall increase in college going (both at 2-year and 4-year colleges) among those who are least likely to attend college, and an increase in 4-year college going among those who are more likely to attend college driven both by increased college attendance and also switching from 2-year to 4-year institutions. One noteworthy result is that the increase in 4-year college going is similar for those in the top third and bottom third even though the base rates are very different (15 versus 66 percent). This indicates that the increases in college going, and those for 4-year institutions are not limited only to populations with students on the margin. These results show that attending an effective high school can lead to sizable increases in college going even among student populations for which that may seem unlikely.

To explore whether these increases in college going result in students persisting in college, we examine impacts on college persistence beyond freshman year. As one can see, there are positive impacts throughout the educational advantage distribution. Much like the impacts on 4-year college going, one cannot likely reject equality of impacts through the distribution.

Finally, we examine whether a student had ever had a school-based arrest. Because this is a relatively rare outcome among student at the top of the education advantage distribution, one would not expect much effect at the top of the distribution. Indeed, this is precisely what one observes. Among students in the bottom decile, a 1σ increase in effectiveness decreases in-school arrests by 2.1 percentage points (p-value$<$0.01) compared to only 0.1 percentage point in the top decile (p-value$<$0.1). Even though there are significant effects even among those at the top of the educational advantage distribution, the marginal effects are much more pronounced for those at the bottom. Given the long terms implication of these school-based arrests, this implies sizable long term benefits to attending effective schools particularly for those who are least likely to complete high school. It is also worth noting that this likely represents a *lower* bound on the effect of arrests because students who may have dropped out of school will not receive a school-based arrest.

The Impact of School Effectiveness on Long-run Outcomes: By Dimension

Our measure of school effectiveness reflects a combination of school impacts on test scores, surveys and behaviours. One may wonder if the heterogeneous impacts we document are driven by one particular dimension. To shed light on this, we present the marginal impacts of attending a school on the disaggregated components of the effectiveness index (test scores, surveys, and behaviours). Similar to Figure 3, we plot the estimated marginal impacts of standardized value-added on each dimension for each decile of the educational advantage index (Figure 4) . As a point of reference, we also plot the impacts of standardized overall school effectiveness.

The results for high school graduation are in the top left panel. In general, students at the lower end of the educational advantage distribution benefit more from schools that improve 9^{th} grade skill measures than those at the top. However, this difference is particularly pronounced for the school impacts on socio-emotional development (SED), measured by surveys. In particular, raising test score value-added and SED value-added by 1 standard deviation increases high school graduation for the top decile by only about 0.2 and 0.6 percentage points, respectively. However, the effects are quite different for the bottom decile; raising test score value-added and SED value-added by 1 standard deviation among those in the bottom decile increases high school graduation by 1.8 and 3.1 percentage-points respectively. That is, while the effects of test score and SED value-added on high school graduation are similar for student at the top decile, the marginal effects are much larger for a 1 standard deviation increase in SED value-added than test score value-added among the lowest decile. Unlike the other dimensions (which show larger benefits for the less advantaged), the impact of school behaviours value-added on high school graduation is similar throughout the educational advantage distribution. The similarity in the pattern of effects between the overall index and SED value-added suggests that the *primary* reason for the larger high-school graduation impacts for students at the lower end of the educational advantage distribution is due to these students being particularly sensitive to improvements in SED value-added. These patterns are consistent with work is psychology suggesting that less advantaged students may enjoy particularly large benefits from interventions that promote socio-emotional development (Sisk et al. 2018,Gray et al. 2018; Walton and Cohen 2007; Walton and Cohen 2011).

Looking to college-going (top middle panel of Figure 4) the pattern of larger impacts at the lower end of the distribution than the top is echoed for both test score and SED value-added. Even though the estimates are somewhat imprecise, the college-going effects are most pronounced for SED value-added for student in the lowest decile of educational advantage – further evidence that less-advantaged students *may* enjoy particularly large benefits from interventions that promote socio-emotional development. The marginal impact of a 1 standard deviation increase in SED value-added on college-going for the top and bottom decides are 1.6 and 2.1 percentage points, respectively (a 31 percent difference). In contrast, this difference or test score value-added is under 10 percent. Much like high school graduation, the marginal effect of behaviours value-added on college-going is relatively similar throughout the educational advantage distribution (about 2 percentage points) though there is suggestive evidence of larger effects at the very top of the distribution and smaller impacts at the very bottom. Looking at 4-year college going (lower left panel), a large part of the benefit of effective schools on 4-year college going is due to effects on SED.

Indeed, through most of the educational advantage distribution, the impact on increasing SED value-added is appreciably larger than that of test-score value-added or behaviours value-added. For the lowest decile, the marginal impact of a standard deviation increase in value-added is between 50 and 100 percent larger for SED value-added than behaviour or test score value-added. For the second decile (which has larger college going impacts in general) this gap is even larger; the marginal impact of a standard deviation increase in value-added is between 125 and 260 percent larger for SED value-added than behaviour or test score value-added. While this gap varies in size throughout the educational advantage distribution, it exists for all but the top decile. This indicates that school impacts on socio-emotional development appear to capture a set of skills and dispositions that are particularly important at promoting 4-year college going for most students.

Looking at school-based arrests (lower right panel) it is clear that, *on average*, behaviours value-added provides much more predictive power than test score or SED value-added. Indeed, as shown in Table 4, on average, a 1 standard deviation increase in behaviours value-added reduces the likelihood of a school-based arrest by 1.28 percentage points, compared to only 0.6 percentage points for SED value-added, 0.35 percentage points for test score value-added. Consistent with this, the behaviours value-added predicts larger reductions in arrests than the other value-added (including the overall index) for most students. However, for the bottom decile of educational advantage, a standard deviation increase in SED value-added predicts larger reduction in arrests (about 2.1 percentage points) than behaviours value-added (about 1.4 percentage points). For all the other deciles, the predictive power is larger for behaviours value-added. Taken together the results indicate that (a) behaviours value-added captures important dimensions of school quality that best predict non-academic outcomes such as arrests, but that (b) among the very least advantaged populations SED value-added may be particularly important (even more so that behaviours value-added). Another notable pattern is that test score value-added predicts small reductions in arrests, while behaviours value-added and SED value added predict large benefits, particularly among those at the bottom of the educational advantage distribution.

In sum, we document that for most outcomes, the benefit to attending a more effective school are larger for the least academically advantaged students. Looking at particular dimensions of school quality, the patterns indicate that this is mainly due to less-advantaged students benefiting the most from schools that improve socio-emotional development or promoting positive behaviours. This supports the notion that cognitive skills only capture a fraction of the skills needed to be successful academically (and in general), and that soft skills play an important role (Farrington et al. 2012; Duckworth et al. 2007; Dweck 2006; Lindqvist and Vestman 2011; Heckman and Rubinstein 2001; Borghans et al. 2008; Waddell 2006 Kautz et al. 2014). Another important implication of the pattern of results is that test-score based measures of school effectiveness may drastically understate the benefits to attending "better" schools – particularly for the least educationally advantaged students.

Differences by Race and Gender and School Type

The summary statistics in Table 1 show that students in the bottom and top of the educational advantage distribution differ along both sex and ethnicity dimensions. As such, one may wonder if these patterns reflect gender or race differences, or if these are broad patterns that exist within demographic groups. To assess this, we implement analogous analyses using students from a particular group (males, females, black, Latinx). By and large, the patters of results that we document across all groups exists within groups (See appendix Figures A1 and A2). As such, our results are not an artifact of making comparisons across sex or ethnic groups. There are, however, some differences that we discuss below.

Looking at males and females separately, the educational effects are similar for the two groups; the average effects on high school graduation are slightly larger for males and the effects on college going are slightly larger for females. In contrast, the arrests rates of males are clearly more responsive to school quality than those of females. The average effect of a 1SD increase in school effectiveness for males is about 1.1 percentage points while that for females is about 0.5 percentage points. However, for both males and females, students at the very bottom of the educational advantage distribution experience larger reductions in arrests from attending a more effective school.

Next we examine effects for Black and Latinx students separately (other ethnic groups are too small to examine heterogeneous impacts). The arrest outcomes are much more sensitive to school effectiveness for Black students than Latinx students, while the educational attainment effects are particularly pronounced for Latinx students. In particular, among Black students in the bottom decile of the educational advantage distribution, a standard deviation increase in school effectiveness reduces the likelihood of a school-based arrest by over 3 percentage points (p-value$<$0.01), while that for Latinx students is less than one percentage point. Looking at educational outcomes, for Latinx students in the bottom of the educational advantage distribution, a standard deviation increase in school effectiveness increases the likelihood of high-school graduation by over 5 percentage points (p-value$<$0.01), and in the middle of the distribution it increases the four-year college going rate by around 7 percentage-points.[13] The analogous numbers for Black students are 2.1, and 1.8 percentage points for high school graduation and college going, respectively.

Given that much of the evidence of differential school effectiveness is based on small samples of oversubscribed charter schools, one may wonder if our results persist if one were to focus only on traditional public schools. To assess this, we implement the entire analysis looking only at traditional public schools (See appendix Figure A3). The patterns we document are very similar when restricted only to traditional public schools. This suggests that the patterns we document may

[13]These relatively large college-going effects are consistent with Jackson (2014) finding particularly large college going responses among Latinx student to a college preparatory program in Texas.

reflect some underlying characteristics of the education production function that may generalize to other settings.

V.3 Distribution of Effectiveness by Advantage

Our results indicate that the least educationally advantaged students may benefit the most from attending more effective schools. As such, it is instructive to assess whether school effectiveness is evenly distributed by educational advantage. To this aim, we compute various percentiles of the school effectiveness index for students in each decile of the advantage index. This provides information about the extent of exposure to high-quality schools by educational advantage. We plot the percentiles for the deciles in Figure 5. One takeaway from this figure is that students of all educational advantage levels are exposed to schools that are both high and low on the effectiveness index. Indeed, the differences in school effectiveness within each decile (e.g., comparing the 5^{th} to the 95^{th} percentile of school effectiveness within a given educational advantage decile) are much larger than the differences in the same percentiles of effectiveness across educational advantage (e.g., comparing the 95^{th} percentile of school effectiveness for the top and bottom deciles of educational advantage). However, there *are* economically significant differences across deciles.

Looking across deciles of educational advantage, the 10^{th} decile (the most advantaged group) is exposed to higher levels of school effectiveness. Indeed, the 95^{th} percentile of school effectiveness for the bottom and top deciles are about 1.54 and 1.94 respectively. While this 0.41 SD difference is modest relative to the unconditional distribution of school effectiveness, it is economically significant. The estimates in Figure 3 indicate that a 0.41 SD increase in effectiveness would increase high school graduation by about 1.3 percentage points, college-going by over 1 percentage points, and reduce the likelihood of being arrested by about 0.9 percentage points. This represents the improvement in outcome student at the bottom of the educational advantage would enjoy if they attended school similar to those attended by the most advantaged. The differences at the median are slightly smaller, but similar (a difference of 0.35 SD) indicating some economically important difference for the top decile of educational advantage compared to others. While differences in school effectiveness do not account for most of the differences in outcomes across students with differing levels of educational advantage (see Figure 1), the potential gains to a more equitable distribution of students across schools are economically significant.

VI Conclusions

Recent research across several social sciences has shown that schools can have important and meaningful impacts on both short-run outcomes and longer-run outcomes. However, the extent to all students benefit similarly from attending better schools is not well understood. Moreover, the extent to which more or less advantaged students benefit differently from school quality in different dimensions (cognitive value added versus socio-emotional and behaviours value added) is unknown. We shed light on these issues by estimating the effect of attending a more effective school for students with very different likelihoods of graduating high school, or attending college. To shed light on the different dimensions of school quality that may matter, we (a) use an overall index of school quality that combines school value-added on cognitive tests, socio-emotional measures, and behaviours, and also (b) examine differential impacts for these different dimensions.

Reinforcing the importance of schools, all students benefit from attending effective schools. Interestingly, even those least likely to attend college experience sizable increases in college going from attending more effective schools. This is due to the least advantaged student receiving particularly large benefit from attending schools that improve socio-emotional development. Our analysis of school-based arrests also suggest large benefits to attending more effective schools particularly for those at the bottom of the educational advantage distribution. For arrests, school impacts on both socio-emotional development and behaviours are important for the least advantaged students. Overall we show that effective schools matter, and that they may matter even more for more fragile student populations. Our results reinforce the importance of soft skills, and suggest that if one were to use test-based measures of school quality alone, one would dramatically understate the benefits for students who need access the better schools the most.

References

Joshua D. Angrist, Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. Who benefits from kipp? *Journal of Policy Analysis and Management*, 31:837–860, 9 2012. ISSN 02768739. doi: 10.1002/pam.21647. URL http://doi.wiley.com/10.1002/pam.21647.

Joshua D. Angrist, Sarah R. Cohodes, Susan M. Dynarski, Parag A. Pathak, and Christopher R. Walters. Stand and deliver: Effects of boston's charter high schools on college preparation, entry, and choice. *Journal of Labor Economics*, 34:275–318, 4 2016. ISSN 0734306X. doi: 10.1086/683665. URL https://www.journals.uchicago.edu/doi/abs/10.1086/683665.

Joshua D. Angrist, Peter D. Hull, Parag A. Pathak, and Christopher R. Walters. Leveraging lotteries for school value-added: Testing and estimation. *Quarterly Journal of Economics*, 132:871–919, 5 2017. ISSN 15314650. doi: 10.1093/qje/qjx001. URL https://academic.oup.com/qje/article/132/2/871/2966019.

Lisa Barrow, Lauren Sartain, and Marisa de la Torre. Increasing access to selective high schools through place-based affirmative action: Unintended consequences. *American Economic Journal: Applied Economics*, 2020. ISSN 1945-7782. doi: 10.1257/APP.20170599.

Diether W. Beuermann and C. Kirabo Jackson. The Short and Long-Run Effects of Attending The Schools that Parents Prefer. *Journal of Human Resources*, pages 1019–10535R1, apr 2020. ISSN 0022-166X. doi: 10.3368/jhr.57.3.1019-10535r1. URL http://jhr.uwpress.org/content/early/2020/04/03/jhr.57.3.1019-10535R1http://jhr.uwpress.org/content/early/2020/04/03/jhr.57.3.1019-10535R1.abstract.

Lex Borghans, Bas ter Weel, and Bruce A. Weinberg. Interpersonal Styles and Labor Market Outcomes. *Journal of Human Resources*, 43(4):815–858, oct 2008.

Shannon T. Brady, Geoffrey L. Cohen, Shoshana N. Jarvis, and Gregory M. Walton. A brief social-belonging intervention in college improves adult outcomes for black americans. *Science Advances*, 6:eaay3689, 4 2020. ISSN 23752548. doi: 10.1126/sciadv.aay3689. URL http://advances.sciencemag.org/.

Jesse Bruhn. The consequences of sorting for understanding school quality, 2020.

Raj Chetty, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. How does your kindergarten classroom affect your earnings? evidence from project star. *Quarterly Journal of Economics*, 126:1593–1660, 11 2011. ISSN 00335533. doi: 10.1093/qje/qjr041. URL https://academic.oup.com/qje/article/126/4/1593/1923939.

Raj Chetty, John N. Friedman, and Jonah E. Rockoff. Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, 104(9):2593–2632, sep 2014.

Susana Claro, David Paunesku, and Carol S. Dweck. Growth mindset tempers the effects of poverty on academic achievement. *Proceedings of the National Academy of Sciences of the United States of America*, 113:8664–8668, 8 2016. ISSN 10916490. doi: 10.1073/pnas.1608207113. URL www.pnas.org/cgi/doi/10.1073/pnas.1608207113.

Sarah R. Cohodes, Elizabeth M. Setren, and Christopher R. Walters. Can successful schools replicate? scaling up boston039;s charter school sector. *American Economic Journal: Economic Policy*, 2020. ISSN 1945-7731. doi: 10.1257/POL.20190259.

Flavio Cunha, James J. Heckman, and Susanne M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78:883–931, 5 2010. ISSN 0012-9682. doi: 10.3982/ecta6551. URL https://onlinelibrary.wiley.com/doi/full/10.3982/ECTA6551https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA6551https://onlinelibrary.wiley.com/doi/10.3982/ECTA6551.

David J. Deming. Better Schools, Less Crime? *. *The Quarterly Journal of Economics*, 126(4):2063–2115, nov 2011.

David J. Deming, Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. School Choice, School Quality, and Postsecondary Attainment, 2014.

Angela L. Duckworth, Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly. Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6):1087–1101, jun 2007. ISSN 1939-1315. doi: 10.1037/0022-3514.92.6.1087.

Andrew Dustan, Alain de Janvry, and Elisabeth Sadoulet. Flourish or fail: The risky reward of elite high school admission in mexico city. *Journal of Human Resources*, 52:756–799, 6 2017. ISSN 15488004. doi: 10.3368/jhr.52.3.0215-6974R1. URL http://jhr.uwpress.org/content/52/3/756http://jhr.uwpress.org/content/52/3/756.abstract.

Carol S. Dweck. *Mindset : the new psychology of success*. Random House, 2006. ISBN 9780345472328.

Camille A. Farrington, Melissa Roderick, Elaine Allensworth, Jenny Nagaoka, Tasha Seneca Keyes, David W. Johnson, and Nicole O. Beechum. Teaching adolescents to become learners : the role of noncognitive factors in shaping school performance : a critical literature review. Technical report, 2012.

Peter Fredriksson, Björn Öckert, and Hessel Oosterbeek. Long-term effects of class size. *Quarterly Journal of Economics*, 128:249–285, 2 2013. ISSN 00335533. doi: 10.1093/qje/qjs048. URL https://academic.oup.com/qje/article/128/1/249/1839904.

De Leon L. Gray, Elan C. Hope, and Jamaal S. Matthews. Black and belonging at school: A case for interpersonal, instructional, and institutional opportunity structures. *Educational Psychologist*, 53:97–113, 4 2018. ISSN 00461520. doi: 10.1080/00461520.2017.1421466. URL https://www.tandfonline.com/doi/abs/10.1080/00461520.2017.1421466.

James Heckman, Rodrigo Pinto, and Peter Savelyev. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103:2052–2086, 10 2013. ISSN 00028282. doi: 10.1257/aer.103.6.2052.

James J Heckman and Yona Rubinstein. The Importance of Noncognitive Skills: Lessons from the GED Testing Program. *American Economic Review*, 91(2):145–149, may 2001. ISSN 0002-8282. doi: 10.1257/aer.91.2.145.

C. Kirabo Jackson. Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago*. *The Economic Journal*, 120(549):1399–1429, dec 2010. ISSN 00130133. doi: 10.1111/j.1468-0297.2010.02371.x.

C. Kirabo Jackson. Do college-preparatory programs improve long-term outcomes? *Economic Inquiry*, 52:72–99, 1 2014. ISSN 00952583. doi: 10.1111/ecin.12040. URL http://doi.wiley.com/10.1111/ecin.12040.

C. Kirabo Jackson. What do test scores miss? the importance of teacher effects on non-test score outcomes. *Journal of Political Economy*, 5 2018. doi: 10.3386/w22226.

C. Kirabo Jackson, Shanette C. Porter, John Q. Easton, Alyssa Blanchard, and Sebastián Kiguel. School effects on socio-emotional development, school-based arrests, and educational attainment. *American Economic Review: Insights*, 2020. doi: 10.1257/AERI.20200029.

Brian A. Jacob. Where the boys aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education Review*, 21:589–598, 12 2002. ISSN 02727757. doi: 10.1016/S0272-7757(01)00051-6.

Mariame Kaba and Frank Edwards. Policing Chicago Public Schools:A Gateway to the School-to-Prison Pipeline. Technical report, 2012. URL http://cpdincps.com/fullreport.

Tim Kautz, James J. Heckman, Ron Diris, Bas ter Weel, and Lex Borghans. Fostering and Measuring Skills. *National Bureau of Economic Research,*, No. w19656, nov 2014.

Naveen Kumar. Public school quality and student outcomes: Evidence from model schools in india, 2020. URL https://drive.google.com/file/d/1-Va6T6WLZzIi1H0UpTBxR0WUo04takXz/view.

Erik Lindqvist and Roine Vestman. The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment. *American Economic Journal: Applied Economics*, 3(1):101–128, jan 2011. ISSN 1945-7782.

Airan Liu. Non-Cognitive skills and the growing achievement Gap. *Research in Social Stratification and Mobility*, page 100546, sep 2020. ISSN 02765624. doi: 10.1016/j.rssm.2020.100546. URL https://linkinghub.elsevier.com/retrieve/pii/S0276562420300822.

Jing Liu and Susanna Loeb. Engaging teachers: Measuring the impact of teachers on student attendance in secondary school. *Journal of Human Resources*, pages 1216–8430R3, 7 2019. ISSN 0022-166X. doi: 10.3368/jhr.56.2.1216-8430r3. URL http://jhr.uwpress.org/content/early/2019/07/02/jhr.56.2.1216-8430R3http://jhr.uwpress.org/content/early/2019/07/02/jhr.56.2.1216-8430R3.abstract.

Mary C. Murphy, Maithreyi Gopalan, Evelyn R. Carter, Katherine T.U. Emerson, Bette L. Bottoms, and Gregory M. Walton. A customized belonging intervention improves retention of socially disadvantaged students at a broad-access university. *Science Advances*, 6:eaba4677, 7 2020. ISSN 23752548. doi: 10.1126/sciadv.aba4677. URL http://advances.sciencemag.org/.

Hessel Oosterbeek, Nienke Ruijs, and Inge de Wolf. Using admission lotteries to estimate heterogeneous effects of elite schools, 2020. URL https://www.dropbox.com/s/1p4t9cxhosctyrs/eliteschools2020march.pdf?dl=0.

Nathan Petek and Nolan G Pope. The multidimensional impact of teachers on students, 2020.

Tim R. Sass, Ron W. Zimmer, Brian P. Gill, and T. Kevin Booker. Charter High Schools' Effects on Long-Term Attainment and Earnings. *Journal of Policy Analysis and Management*, 35(3):683–706, jun 2016.

Ying Shi. Who benefits from selective education? evidence from elite boarding school admissions. *Economics of Education Review*, 74:101907, 2 2020. ISSN 02727757. doi: 10.1016/j.econedurev.2019.07.001.

Victoria F. Sisk, Alexander P. Burgoyne, Jingze Sun, Jennifer L. Butler, and Brooke N. Macnamara. To what extent and under which circumstances are growth mind-sets important to academic achievement? two meta-analyses. *Psychological Science*, 29:549–571, 4 2018. ISSN 0956-7976. doi: 10.1177/0956797617739704. URL http://journals.sagepub.com/doi/10.1177/0956797617739704.

Glen R. Waddell. LABOR-MARKET CONSEQUENCES OF POOR ATTITUDE AND LOW SELF-ESTEEM IN YOUTH. *Economic Inquiry*, 44(1):69–97, jan 2006.

Christopher R. Walters. The demand for effective charter schools. *Journal of Political Economy*, 126:2179–2223, 12 2018. ISSN 1537534X. doi: 10.1086/699980. URL https://www.journals.uchicago.edu/doi/abs/10.1086/699980.

Gregory M. Walton and Geoffrey L. Cohen. A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92:82–96, 1 2007. ISSN 00223514. doi: 10.1037/0022-3514.92.1.82.

Gregory M. Walton and Geoffrey L. Cohen. A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, 331:1447–1451, 3 2011. ISSN 00368075. doi: 10.1126/science.1198364. URL https://pubmed.ncbi.nlm.nih.gov/21415354/.

Tables and Figures

Table 1: Summary Statistics

	Analytic Sample		Bottom Decile of Educational Advantage Index		Top Decile of Educational Advantage Index	
	mean	SD	mean	SD	mean	SD
			Demographics			
Female	0.4916	0.4999	0.2431	0.4290	0.6977	0.4593
Special education (IEP)	0.1834	0.3870	0.4526	0.4978	0.0560	0.2300
Free lunch	0.7879	0.4088	0.9525	0.2127	0.4327	0.4955
Reduced-price lunch	0.0734	0.2608	0.0188	0.1360	0.1534	0.3603
Census Block SES	-0.4616	0.8658	-0.5887	0.8148	-0.1104	0.9039
White	0.0847	0.2784	0.0408	0.1979	0.2312	0.4216
Black	0.4121	0.4922	0.5232	0.4995	0.2427	0.4287
Native American	0.0017	0.0417	0.0020	0.0444	0.0030	0.0546
Asian/Pacific Islander	0.0325	0.1772	0.0013	0.0365	0.1856	0.3888
Latino	0.4589	0.4983	0.4306	0.4952	0.3303	0.4703
			9th grade Intemediate Outcomes			
Test Scores in 9th Grade	-0.0276	0.9834	-0.9980	0.6447	1.4077	0.7164
Work Hard in 9th Grade	0.1795	0.9874	-0.0807	1.0225	0.5116	0.9697
Social in 9th Grade	-0.0026	0.9988	-0.2386	1.0452	0.3376	0.9782
Surveys in 9th Grade	0.1718	0.9523	-0.0994	0.9935	0.5367	0.9287
Behavior in 9th Grade	0.1688	0.7620	-0.5201	1.4190	0.4537	0.1951
Days Absent in 9th Grade	15.1211	18.7236	34.4276	27.7463	5.5405	7.5782
Days Suspended in 9th grade	0.8183	3.3172	2.9514	6.6062	0.0621	0.6685
Diciplinary Incidents in 9th Grade	0.0782	0.4218	0.2922	0.8717	0.0068	0.0942
On Track in 9th Grade	0.8462	0.3607	0.5738	0.4945	0.9804	0.1385
			8th Grade Measures			
Math in 8th Grade	0.1908	0.9377	-0.8607	0.6109	1.7885	0.7468
ELA in 8th Grade	0.1959	0.9355	-0.8514	0.8203	1.6046	0.7917
Emotional Health in 8th Grade	0.0673	0.8972	-0.1994	0.8923	0.3224	0.9298
Academic Engagement in 8th Grade	0.2691	0.9137	0.1333	0.8621	0.3582	1.0133
Grit in 8th Grade	0.0440	0.8373	-0.3160	0.8772	0.4330	0.8148
School Connectedness in 8th Grade	0.1393	0.9015	-0.0320	0.8701	0.4391	0.9836
Study Habits in 8th Grade	0.1497	0.8904	-0.2341	0.8541	0.6797	0.9568
Absences in 8th Grade	8.7303	8.6344	19.7363	12.4395	4.5914	3.8855
GPA in 8th Grade	2.7899	0.7795	2.0326	0.7810	3.6003	0.4908
Days Suspended in 8th Grade	0.4479	1.8229	2.2757	4.4857	0.0231	0.2611
Incidents in 8th Grade	0.0655	0.3359	0.3824	0.8485	0.0011	0.0349
			Long-term Outcomes			
Any school-Based arrest	0.0377	0.1905	0.1260	0.3319	0.0044	0.0660
Graduation	0.7392	0.4391	0.4278	0.4948	0.9370	0.2429
Enrolled in any college within 2 years	0.5288	0.4992	0.1742	0.3793	0.8733	0.3326
Enrolled in a 4 year college within 2 years	0.3386	0.4732	0.0596	0.2368	0.7790	0.4150
Enrolled in a 2 year college within 2 years	0.2764	0.4472	0.1283	0.3345	0.2501	0.4331
N	157027		15703		15702	

Notes: Number of observations may vary by variable due to missingness and variation in cohorts for which a variable was collected.

Table 2: Temporal Stability of Value-Added and Correlations Across Value-Added

Correlations of Value-Added Within Outcomes Across Time

	Test scores Value-added	Work Hard Value-added	Social Value-added	Behaviours Value-added
t+1	.417	.274	.36	.708
t+2	.285	.192	.166	.571
t+3	.109	.103	.09	.506
t+4	.18	.188	.212	.372

Correlations of Average School-Level Value-Added Across Outcomes (143 Schools)

Test Score Value Added	1			
Work Hard Value-Added	0.4449	1		
Social Value-Added	0.4795	0.6486	1	
Behaviours Value-Added	0.1468	0.0205	0.0746	1

Notes: All reported results are restricted to school-year cells with at least 10 respondents. The **top panel** reports, for each 9^{th} grade measure measure, the correlations between a schools value-added in year t and value-added for years t+1, t+2, t+3, and t+4. The **bottom panel** reports, the correlations between the value-addeds (estimates across all years) for the 9^{th} grade measures.

Table 3: Factor Analysis

	Variance	Difference	Proportion
Factor 1	1.41463	1.38108	1.1886
Factor 2	0.03355	.	0.0282

Rotated factor loadings (pattern matrix) and unique variances

	Factor1	Factor2	Uniqueness
Work Hard Value-Added	0.7922	-0.0144	0.3723
Social Value-Added	0.7954	0.0142	0.3671
Test Scores Value-Added	0.3351	-0.1141	0.8747
Behaviours Value-Added	0.2052	0.1419	0.9378

Method: principal factors

Rotation: orthogonal varimax (Kaiser off)

Table 4: Average Impacts of Value-Added and School Effectiveness

	1 Test scores 9th Grade	2 Surveys 9th Grade	3 Behaviors 9th Grade	4 HS Graduation	5 Enrolled in Any College Within 2 Years	6 School-Based Arrests
School Effectiveness Index	0.0668***	0.0790***	0.0428***	0.0189***	0.0217***	-0.00786***
	(0.0115)	(0.00892)	(0.0103)	(0.00372)	(0.00593)	(0.00203)
Socioemotional Value-Added	0.0602***	0.0772***	0.0328***	0.0170***	0.0186***	-0.00691***
	(0.0117)	(0.00996)	(0.0105)	(0.00371)	(0.00585)	(0.00205)
Test-Score Value-Added	0.0639***	0.0336***	0.0199***	0.0114***	0.0154***	-0.00373**
	(0.0114)	(0.00577)	(0.00729)	(0.00249)	(0.00497)	(0.00152)
Behavior Value-added	0.0214**	0.0247**	0.173***	0.00919**	0.0171***	-0.0123***
	(0.0107)	(0.0101)	(0.00848)	(0.00455)	(0.00496)	(0.00270)
Observations	102,200	124,833	157,027	82,092	55,509	82,092

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Results are based on regression of outcomes on a single measure of out-of-sample school impacts (overall effectiveness, test score value-added, socio-emotional value-added, or behaviour value-added). All models include individual demographic controls (race / ethnicity, free and reduced price lunch, and gender), 8th grade lags (math and ELA test scores, survey measures, absences, and discipline), and school-level averages for all the demographics and lagged measures, as well as year fixed effects. We also include the socio-economic status of the student census block proxied by average occupation status and education levels. Missing 8^{th} grade measures were imputed using 7^{th} grade measures and demographic characteristics. For the longer-run college outcomes, the sample includes first time 9^{th} grade students between 2011 and 2014. *For the longer-run high-school outcomes, the sample includes first time* 9^{th} *grade students between 2011 and 2015.* For the measures, the sample includes first time 9^{th} grade students between 2011 and 2017. **Note:** Sample sizes may differ across outcomes due to some missingness in 9th grade test scores and surveys.

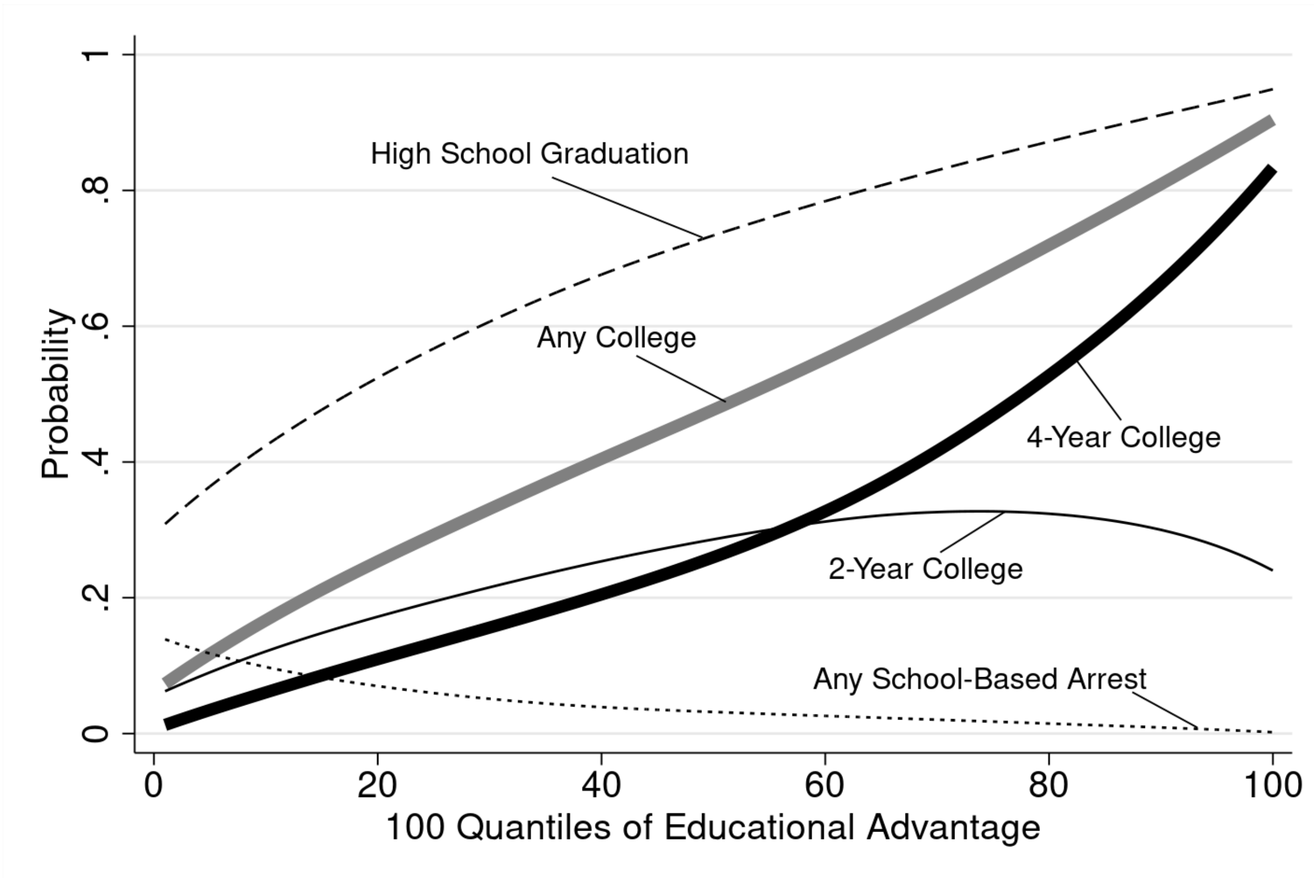
Table 5: Testing for Selection

	Intermediate Outcomes				Long-Run Outcomes			
	1	2	3	4	5	6	7	8
	9th Grade Test Scores			Predicted	HS Graduation			Predicted
School Effectiveness Index	0.0668***	0.0462***	0.0459***	-0.000119	0.0189***	0.0316***	0.0143***	4.05e-05
	(0.0115)	(0.0111)	(0.00975)	(0.000165)	(0.00372)	(0.00748)	(0.00530)	(8.90e-05)
Observations	102,200	99,649	16,384	102,200	82,092	79,498	8,188	82,092
F-statistic on First Stage		549.9				827.7		
	9th Grade Survey Measures			Predicted	Enrolled in College within 2 Years			Predicted
School Effectiveness Index	0.0790***	0.0928***	0.0457***	-3.66e-05	0.0217***	0.0283***	0.00943	-4.25e-07
	(0.00892)	(0.0106)	(0.0138)	(5.22e-05)	(0.00593)	(0.0104)	(0.0125)	(7.06e-05)
Observations	124,833	122,071	28,800	124,833	55,509	53,190	3,399	55,509
F-statistic on First Stage		506.9				683.3		
	9th Grade Behaviors			Predicted	In-school Arrests			Predicted
School Effectiveness Index	0.0428***	0.0833***	0.0169**	0.000199	-0.00786***	-0.0152***	-0.00754*	-2.55e-05
	(0.0103)	(0.0116)	(0.00765)	(0.000184)	(0.00203)	(0.00352)	(0.00450)	(1.62e-05)
Observations	157,027	153,928	41,709	157,588	82,092	79,498	8,188	82,092
F-statistic on First Stage		557				827.7		
Sibling FE			X				X	
School Assignment IV		X				X		

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

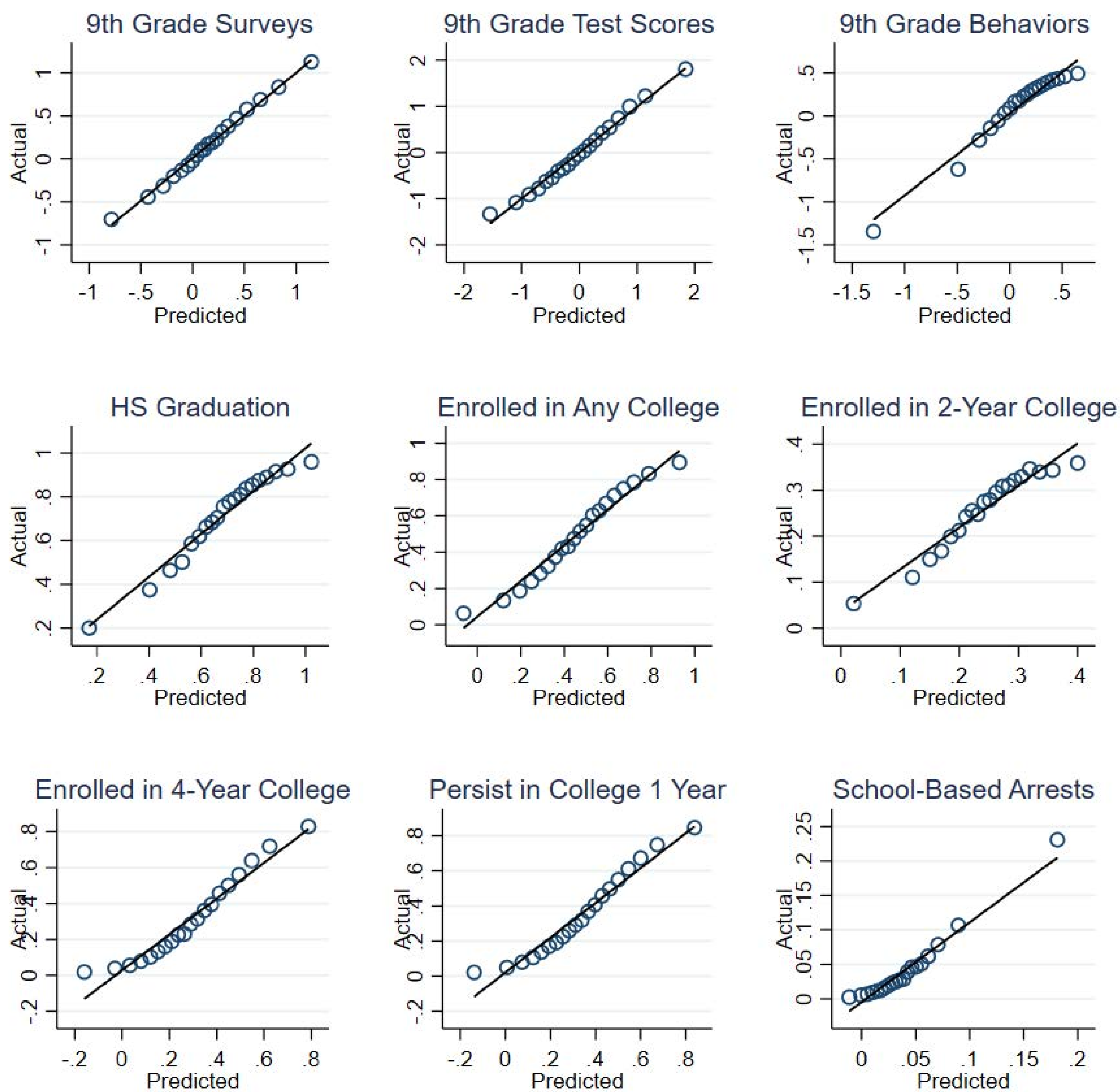
Results are based on regression of outcomes on out-of-sample school effectiveness. All models include individual demographic controls (race / ethnicity, free and reduced price lunch, and gender), 8th grade lags (math and ELA test scores, survey measures, absences, and discipline), and school-level averages for all the demographics and lagged measures, as well as year fixed effects. We also include the socio-economic status of the student census block proxied by average occupation status and education levels. Missing 8^{th} grade measures were imputed using 7^{th} grade measures and demographic characteristics. For the longer-run outcomes, the sample includes first time 9^{th} grade students between 2011 and 2014. For the measures, the sample includes first time 9^{th} grade students between 2011 and 2017. **Columns 4 and 8:** Predicted outcomes are fitted values from a linear regression of said outcome on *all* observed controls. The predictors include lagged measures (i.e., 8^{th} grade test scores, surveys, behaviours), gender, ethnicity, free-lunch status, and the socio-economic status of the student's census block. To avoid mechanical correlation, we use leave-year out predicted outcomes (i.e., predicted outcomes based on the relationship between the outcome and covariates in *other* years). The reported point estimates are those on predicted outcomes on the value-addeds with no controls. **Note:** Sample sizes may differ across outcomes due to some missingness in 9th grade test scores and surveys.

Figure 1. *Average Outcomes: By Estimated Educational Advantage*



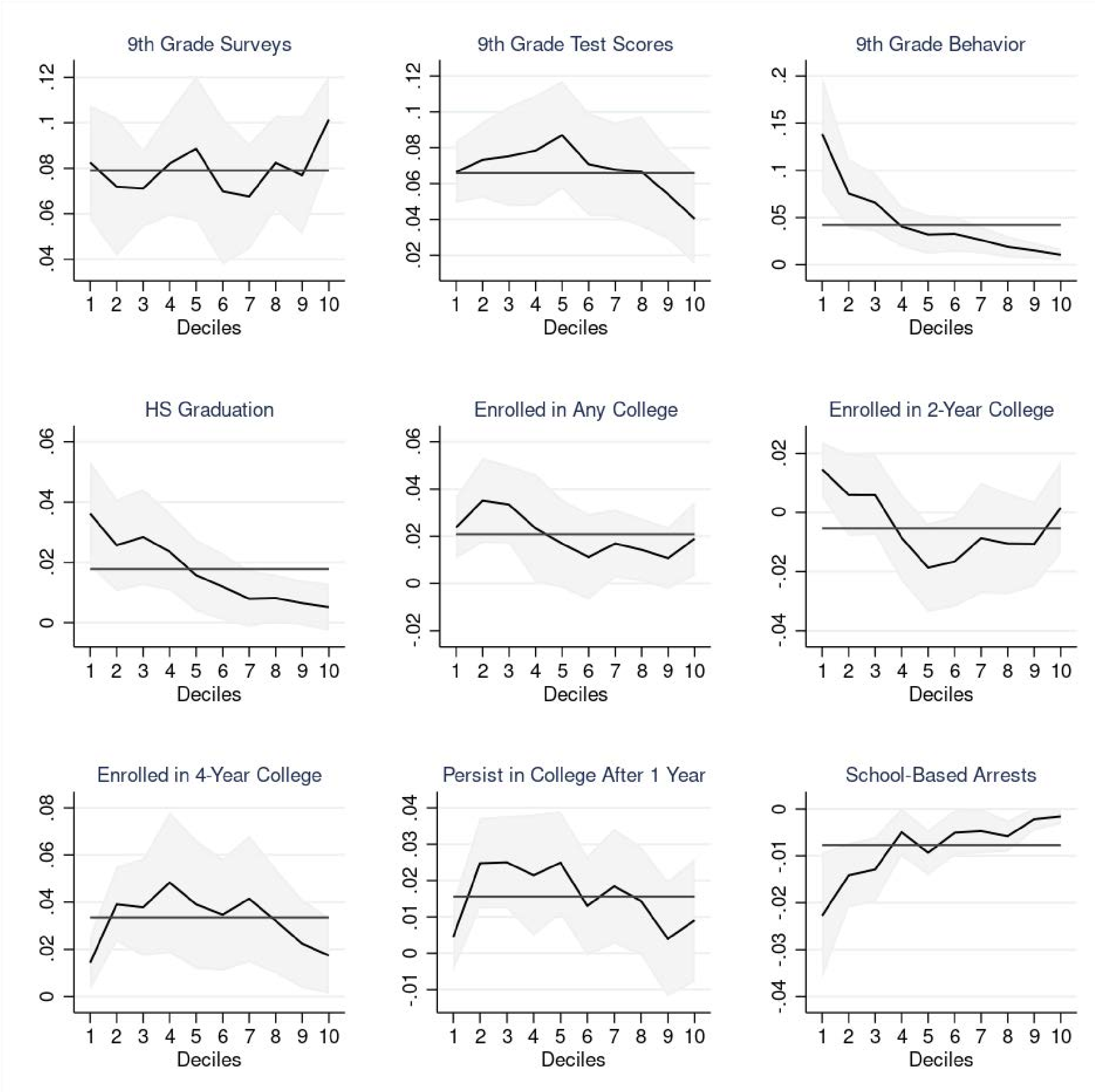
Notes: This figures plots the average of each outcome for different percentiles of the estimated educational advantage distribution. The predicted educational advantage is the fitted value from an ordered probit model predicting the level of education attained based on all 8^{th} grade measures and demographics (*in all other years*). We present the coefficient estimates from the ordered probit model for the full sample in Appendix Table A3.

Figure 2. *Actual Outcome by Predicted Outcome*



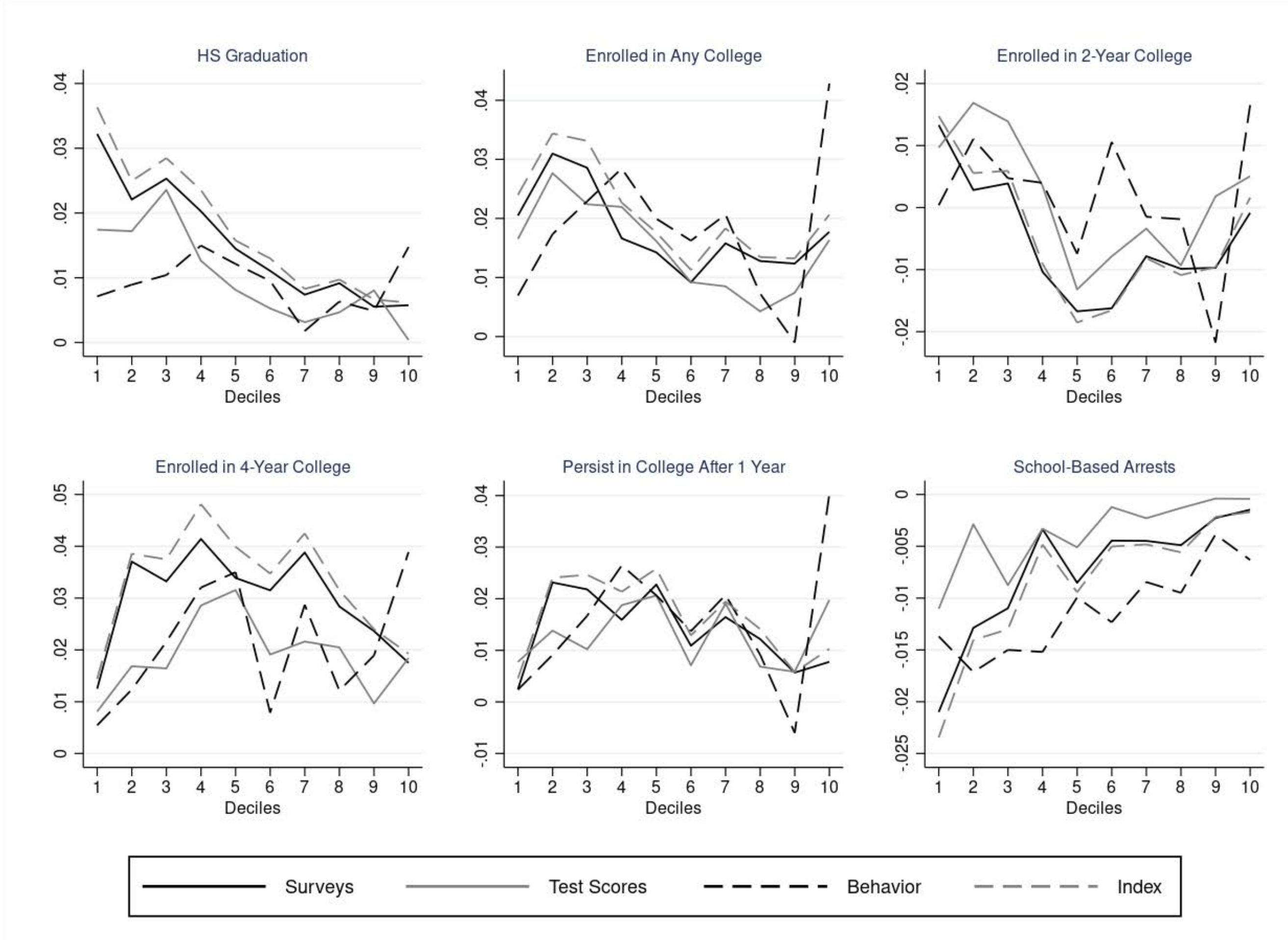
Notes: Each graph presents the average of the actual outcome for different groups of students by predicted outcome. The predicted outcomes are the fitted values from a regression of each outcome on all observed demographics and 8^{th} grade measures based on students in *other* years. The predictors include lagged measures (i.e., 8^{th} grade test scores, surveys, behaviours), gender, ethnicity, free-lunch status, and the socio-economic status of the student's census block.

Figure 3. *Impacts on Outcomes: By Estimated Educational Advantage*



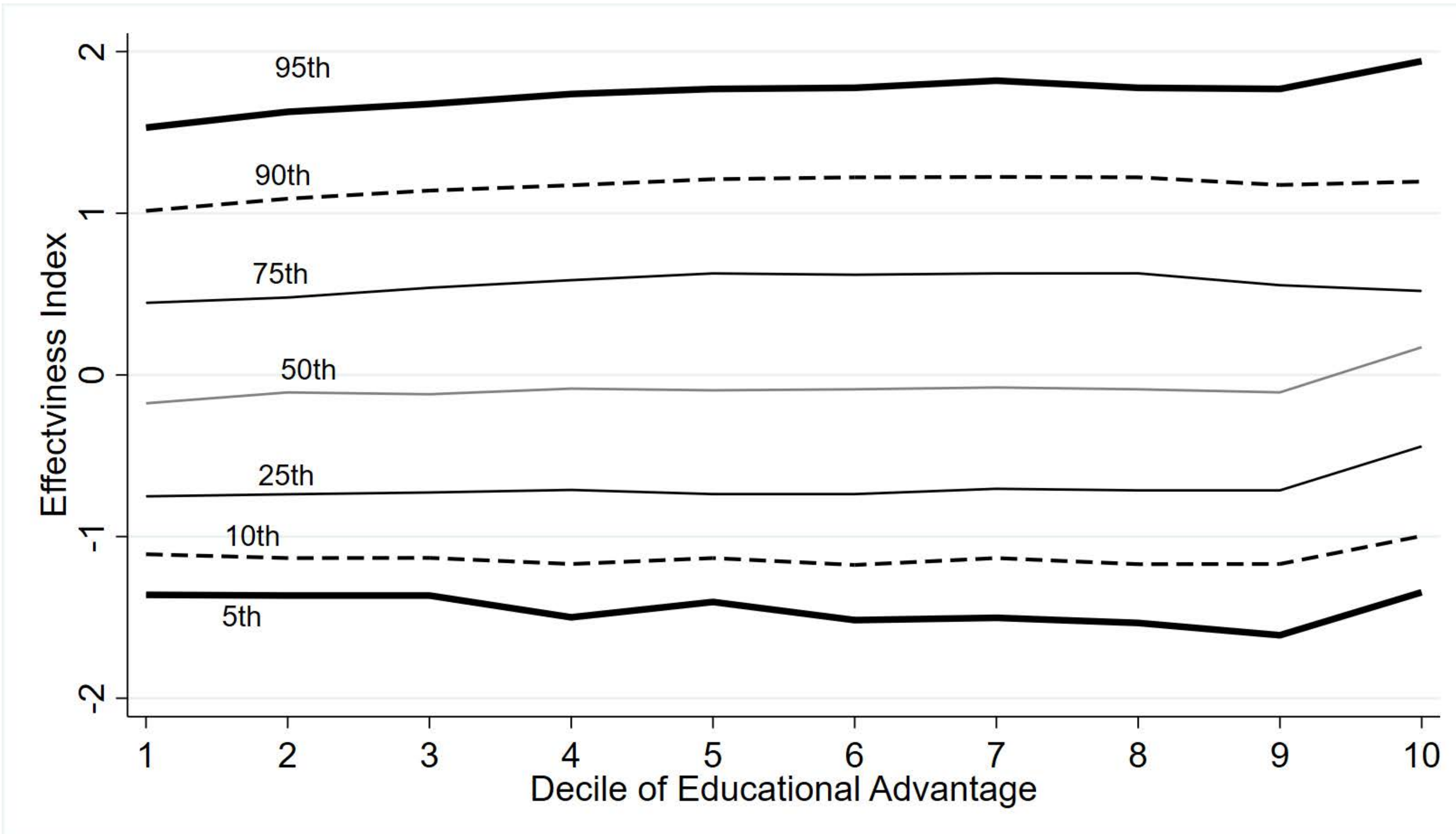
Notes: Each graph represents the marginal impacts of a 1 standard deviation increase in overall school effectiveness for different deciles of the educational advantage distribution for a single outcome. Each panel presents the results of 10 separate regressions each defined as in Equation (5). The dashed black horizontal line in each panel depicts the the average marginal impacts as defined in Table 4.

Figure 4. *Impacts on Long-Run Outcomes: By Quality Dimension and Educational Advantage*



Notes: Each of the 6 panels represents the marginal impacts of a 1 standard deviation increase in school impacts (effectiveness index, SED value-added, test score value-added, behaviours value added) for different deciles of the educational advantage distribution for a single outcome. As such, each panel represents the results of 40 separate regressions. Each regression model controls for the same covariates as in Equation (5).

Figure 5. *Percentiles of Effectiveness Index: By Estimated Educational Advantage*



Notes: This plot various percentiles of the overall effectiveness index for student with different levels of educational advantage.

VII Appendix

Table A1: Summary Statistics for Survey Completers and Non-Completers

	Analytic Sample		Completed the Surveys		Did not complete Surveys	
	mean	SD	mean	SD	mean	SD
Demographics						
Female	0.4916	0.4999	0.502669	0.499995	0.458725	0.4983
Special education (IEP)	0.1834	0.3870	0.158572	0.365278	0.258087	0.437588
Free lunch	0.7879	0.4088	0.780976	0.413587	0.808699	0.393331
Reduced-price lunch	0.0734	0.2608	0.077028	0.266637	0.0625	0.242065
Census Block SES	-0.4616	0.8658	-0.46797	-0.87357	-0.44255	-0.84185
White	0.0847	0.2784	0.089733	0.2858	0.069668	0.254591
Black	0.4121	0.4922	0.382196	0.485926	0.50176	0.500003
Native American	0.0017	0.0417	0.001672	0.040855	0.001939	0.043989
Asian/Pacific Islander	0.0325	0.1772	0.036384	0.187244	0.020714	0.142428
Latino	0.4589	0.4983	0.480416	0.499618	0.394413	0.488731
9th grade Intemediate Outcomes						
Test Scores in 9th Grade	-0.0276	0.9834	0.029924	-0.96869	-0.21285	-1.00746
Work Hard in 9th Grade	0.1795	0.9874	0.186834	-0.98092	-0.02135	-1.13492
Social in 9th Grade	-0.0026	0.9988	0.003019	-0.99426	-0.15024	-1.09945
Surveys in 9th Grade	0.1718	0.9523	0.179159	-0.94566	-0.01489	-1.08968
Behavior in 9th Grade	0.1688	0.7620	0.233114	-0.64323	-0.029	-1.02141
Days Absent in 9th Grade	15.1211	18.7236	12.99633	15.59982	21.65436	24.95769
Days Suspended in 9th grade	0.8183	3.3172	0.644835	2.795615	1.337551	4.495297
Diciplinary Incidents in 9th Grade	0.0782	0.4218	0.061845	0.3596	0.127143	0.566646
On Track in 9th Grade	0.8462	0.3607	0.870445	0.335815	0.757014	0.428896
8th Grade Measures						
Math in 8th Grade	0.1908	0.9377	0.25101	-0.93307	0.010372	-0.92871
ELA in 8th Grade	0.1959	0.9355	0.257527	-0.91814	0.010957	-0.96243
Emotional Health in 8th Grade	0.0673	0.8972	0.079809	-0.90438	0.029781	-0.87456
Academic Engagement in 8th Grade	0.2691	0.9137	0.275683	-0.92486	0.249519	-0.87944
Grit in 8th Grade	0.0440	0.8373	0.052673	-0.84616	0.017878	-0.81006
School Connectedness in 8th Grade	0.1393	0.9015	0.143819	-0.91049	0.125375	-0.87399
Study Habits in 8th Grade	0.1497	0.8904	0.16246	-0.90448	0.111173	-0.84576
Absences in 8th Grade	8.7303	8.6344	8.113539	7.758448	10.57956	10.63503
GPA in 8th Grade	2.7899	0.7795	2.837592	0.772915	2.646929	0.781757
Days Suspended in 8th Grade	0.4479	1.8229	0.360557	1.558484	0.709553	-2.43069
Incidents in 8th Grade	0.0655	0.3359	0.053284	0.287962	0.102219	0.448415
Long-Run Outcomes						
Any school-Based arrest	0.0377	0.1905	0.031782	0.175422	0.053904	0.225834
Graduation	0.7392	0.4391	0.777252	0.416094	0.63654	0.481007
Enrolled in any college within 2 years	0.5288	0.4992	0.573986	0.494502	0.405241	0.490956
Enrolled in a 4 year college within 2 years	0.3386	0.4732	0.373327	0.483694	0.243449	0.429179
Enrolled in a 2 year college within 2 years	0.2764	0.4472	0.29627	0.456617	0.222361	0.415846
N	157027		117827		39200	

Notes: Survey completers are students who have 9^{th}-grade data for emotional health, academic engagement, grit, school connectedness, and study habits. As such, we report averages for some measures even among non-completers because many noncompleters are missing some data but not others.

Table A2: Psychometric Properties of SED measures (*as reported by the University of Chicago Consortium on School Research*): 2011 through 2013

Measure	School Year	Separation	Reliability	Item Infits	Item Outfits
Grit	2010-11	1.68	0.74	0.84, 0.76, 0.71, 1.24	0.85, 0.76, 0.71, 1.19
Social Skills	2010-11	1.69	0.74	1.08, 1.36, 1.41, 1.11	1.05, 1.33, 1.44, 1.15
Academic Effort	2010-11	1.74	0.75	0.85, 1.22, 1.1, 0.91	0.82, 1.17, 1.12, 0.94
Academic Engagement	2010-11	1.59	0.7	0.49, 0.56, 0.71, 0.56	0.49, 0.57, 0.72, 0.58
Belonging	2010-11	2.07	0.81	0.93, 1.02, 0.99, 0.96, 1.29	0.91, 0.97, 0.99, 0.93, 1.33
Grit	2011-12	1.54	0.7	0.8, 0.73, 0.68, 1.19	0.81, 0.57, 0.6, 0.42
Social Skills	2011-12	1.68	0.74	1.37, 1.36, 1.28, 1.06	1.68, 1.24, 1.18, 0.95
Academic Effort	2011-12	1.75	0.75	0.85, 1.22, 1.08, 0.92	0.82, 1.17, 1.1, 0.96
Academic Engagement	2011-12	1.56	0.71	0.54, 0.53, 0.47, 0.69	0.56, 0.55, 0.48, 0.71
Belonging	2011-12	2.13	0.82	0.98, 1.28, 0.91, 1.02, 0.97	0.97, 1.32, 0.89, 0.97, 0.94
Grit	2012-13	1.55	0.71	0.77, 0.69, 0.63, 1.13	0.79, 0.7, 0.63, 1.1
Social Skills	2012-13	1.67	0.74	1.3, 1.37, 1.23, 1.04	1.55, 1.25, 1.12, 0.94
Academic Effort	2012-13	1.77	0.76	0.86, 1.2, 1.13, 0.94	0.83, 1.15, 1.15, 0.97
Academic Engagement	2012-13	1.57	0.71	0.55, 0.54, 0.47, 0.69	0.57, 0.56, 0.48, 0.70
Belonging	2012-13	2.14	0.82	0.95, 1.28, 0.90, 1.03, 0.96	0.95, 1.31, 0.87, 0.98, 0.93

Notes. The reported statistics are from internal documentation at the University of Chicago Consortium on School Research where Rasch analysis was performed on individual survey items. All measures are anchored to 2010-11 step and item difficulties. Infit and outfit measures greater than 1 indicate underfit to the Rasch model and values lower than 1 indicate overfit. Generally, infit and outfit values in the range of 0.6-1.4 are considered reasonable for survey measures. Reliability represents individual reliability and includes extreme people. The patterns are very similar for years 2013 through 2018.

Table A3: Ordered Probit Parameter Estimates

	longterm		cont'd
8th Grade Math	0.296***	Native	-0.487**
	(0.0104)		(0.212)
8th Grade Math Squared	0.00605	Asian	-0.0206
	(0.00585)		(0.165)
8th Grade ELA	0.170***	Latinx	-0.359**
	(0.00932)		(0.160)
8th Grade ELA Squared	0.0148***	Other Race	-0.0315
	(0.00483)		(0.347)
Emotional Health in 8th Grade	-0.0110	Female	0.0409
	(0.00736)		(0.151)
Academic Engagement in 8th Grade	-0.00877	Female*White	0.104
	(0.00582)		(0.158)
Grit in 8th Grade	0.0451***	Female*Black	0.304**
	(0.00433)		(0.155)
School Connectedness in 8th Grade	-0.0186***	Female*Native	0.319*
	(0.00708)		(0.190)
Study Habits in 8th Grade	0.106***	Female*Asian	0.154
	(0.00830)		(0.156)
8th Grade top 25% Absenses	-0.593***	Female*Latinx	0.197
	(0.0147)		(0.149)
Serious Incidents in 8th Grade	-0.391***	Female*Other Race	-0.517
	(0.0267)		(0.423)
Receive Free Lunch	-0.199***	/cut1	-1.224***
	(0.0455)		(0.182)
Receive Reduced Price Lunch	0.0397	/cut2	-0.535***
	(0.0456)		(0.187)
White	-0.288**	/cut3	0.0536
	(0.146)		(0.194)
Black	-0.372**		
	(0.175)	Observations	115,381

Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
Note that the sample size is larger than the analytic sample used for the main outcome analysis. This is prediction model uses all available data, which include observation for individuals who attend schools that do not have valid value-added estimates. the results are very similar is we restrict the prediction to only those same individual in the main analytic long term sample.

Table A4: Effect of SED Value-Added on Average Intermediate and Long-Term Student Outcomes

	1 Surveys 9th Grade	2 Test scores 9th Grade	3 Behaviors 9th Grade	4 HS Graduation	5 Enrolled in Any College Within 2 Years	6 School-Based Arrests
Workhard Value-Added	0.0681***	0.0564***	0.0256**	0.0157***	0.0192***	-0.00681***
	(0.00962)	(0.0113)	(0.0104)	(0.00377)	(0.00593)	(0.00205)
Social Value-Added	0.0785***	0.0577***	0.0370***	0.0149***	0.0155***	-0.00572***
	(0.00884)	(0.0108)	(0.0106)	(0.00354)	(0.00512)	(0.00206)
Observations	124,833	102,200	157,027	82,092	55,509	82,091

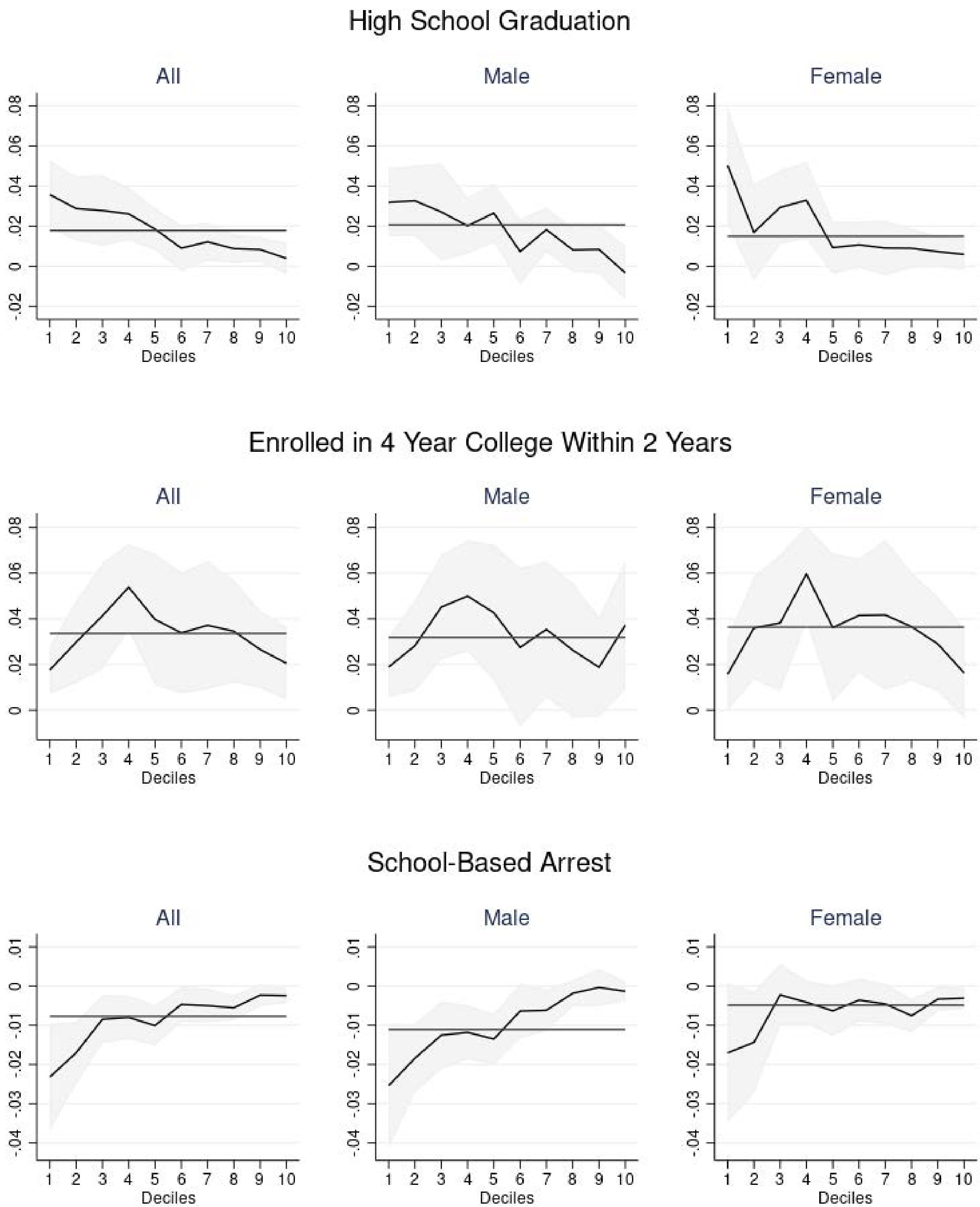
Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: Each point estimate comes from a separate regression.

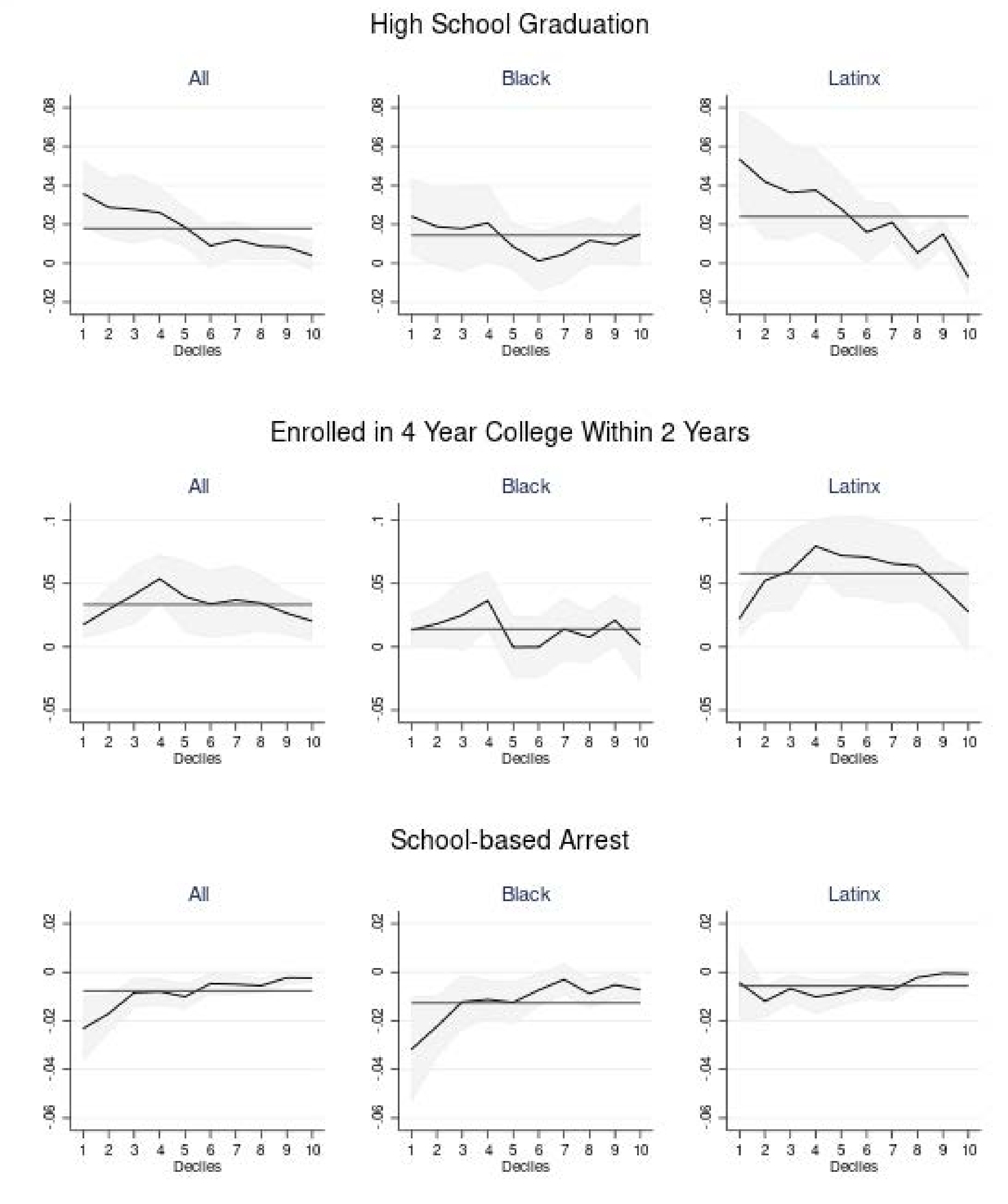
Note: Sample sizes may differ across outcomes due to some missingness in 9th grade test scores and surveys.

Figure A1. *Impacts on Outcomes: By Estimated Educational Advantage and Sex*



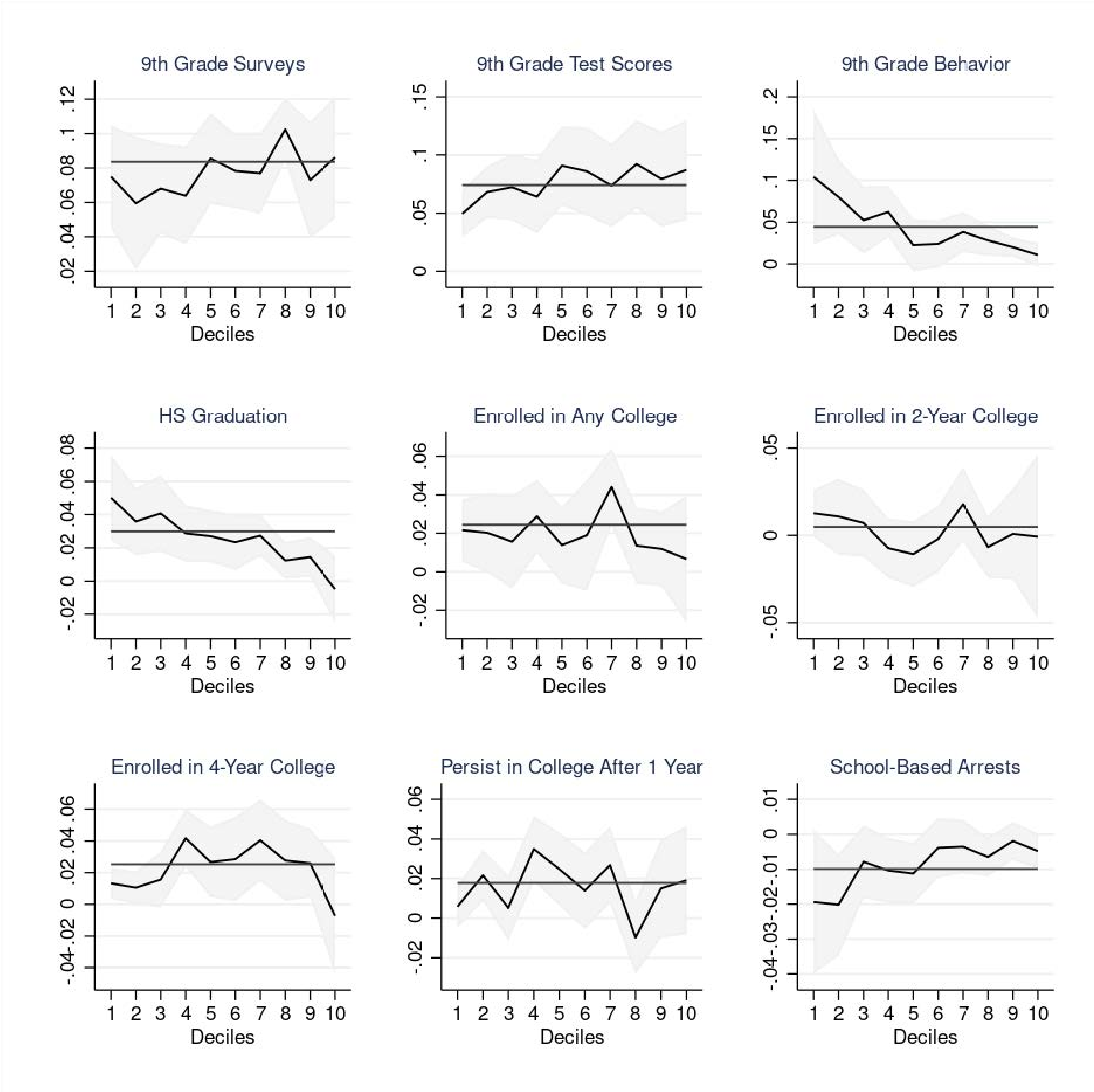
Notes: Each graph represents the marginal impacts of a 1 standard deviation increase in overall school effectiveness for different deciles of the educational advantage distribution for a single outcome by sex. Each panel presents the results of 10 separate regressions each defined as in Equation (5).

Figure A2. *Impacts on Outcomes: By Estimated Educational Advantage and Race*



Notes: Each graph represents the marginal impacts of a 1 standard deviation increase in overall school effectiveness for different deciles of the educational advantage distribution for a single outcome by reported race. Each panel presents the results of 10 separate regressions each defined as in Equation (5).

Figure A3. *Impacts on Outcomes: By Educational Advantage (neighborhood schools only)*



Notes: This graph is based only on the sample of students who attend traditional neighborhood schools. This excludes charter schools, selective enrolment schools, and magnet schools. Each graph represents the marginal impacts of a 1 standard deviation increase in overall school effectiveness for different deciles of the educational advantage distribution for a single outcome. Each panel presents the results of 10 separate regressions each defined as in Equation (5) but only on the sample of traditional public school students.