Northwestern University

From the SelectedWorks of C. Kirabo Jackson

2020

School Effects on Socio-emotional Development, School-Based Arrests, and Educational Attainment

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

School Effects on Socio-emotional Development, School-Based Arrests, and Educational Attainment[*]

C. Kirabo Jackson
Northwestern University

Shanette C. Porter
Mindset Scholars Network

John Q. Easton
UChicago Consortium on School Research

Alyssa Blanchard
UChicago Consortium on School Research

Sebastián Kiguel
Northwestern University

February 12, 2020

Abstract

Using value-added models, we find that high schools impact students' self-reported socio-emotional development (SED) by enhancing social well-being and promoting hard work. Conditional on schools' test score impacts, schools that improve SED reduce school-based arrests, and increase high-school completion, college-going, and college persistence. Schools that improve social well-being have larger effects on attendance and behavioral infractions in high school, while those that promote hard work have larger effects on GPA. Importantly, school SED value-added is more predictive of school impacts on longer-run outcomes than school test-score value-added. As such, for the longer-run outcomes, using both SED and test score value-added more than doubles the variance of the explained school effect relative to using test score value-added alone. Results suggest that adolescence can be a formative period for socio-emotional growth, high-school impacts on SED can be captured using self-report surveys, and SED can be fostered by schools to improve longer-run outcomes. These findings are robust to tests for plausible forms of selection.

I Introduction

Literature in economics, psychology, and sociology documents that socio-emotional skills and mindsets, such as adaptability, grit, motivation, empathy, conflict resolution, problem-solving, and teamwork are strongly related to education and adult outcomes (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; Deming 2017). These skills and mindsets (also known as soft or non-cognitive) are distinct from the numeracy and literacy skills emphasized in most traditional education systems. In response to this growing knowledge base, many high schools are training teachers to attend to socio-emotional development (SED) and incorporating socio-emotional learning into their curriculums and self-report assessments.[1]

But, can high schools influence self-reports of SED and does it matter for long-run outcomes? Education policy and practice have preempted definitive evidence that SED can be meaningfully shaped in high school. For example, while intervention research has demonstrated that some socio-emotional factors are malleable, there is debate about whether and to what extent this is true for all socio-emotional factors (Revelle 2007; Rimfeld et al. 2016; Credé et al. 2017). Second, because the self-report measures that are typically used to assess SED in schools are susceptible to response biases, there is uncertainty regarding whether one can accurately measure impacts on these socio-emotional skills in ways that are informative for policy (e.g., West et al. 2016; Dweck and Yeager 2019). Finally, because most research on soft skills reflect correlations between measures of soft skills and long-run outcomes, evidence on the extent to which *school-generated* improvements on these self-reported skills *causally* improve subsequent outcomes is limited.[2]

To progress on these issues, we leverage a uniquely detailed data-set that links students to schools with self-reported survey measures of SED over time. 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. Using the SED measures in a value-added framework, we (a) estimate schools' causal impacts on self-reported SED, (b) establish the extent to which individual schools impact on SED (based on self-reports) persist over time, (c) explore the relationships among school impacts on different measures of SED, and (d) determine the extent to which attending a school that positively impacts self-reported SED leads to

---

[1]In 2004, Illinois was the first state to develop SEL standards and performance indicators. Since then, at the state level, Kansas, Michigan, Minnesota, New York, North Dakota, Tennessee, and Wisconsin have all incorporated measures of SEL into their curriculum (CASEL). There are also many individual school districts and charter school networks in other states that have implemented SEL learning.

[2]While some studies find that interventions at the primary (Alan et al., 2019) and middle school levels (Cohen et al. (2006) and Blackwell et al. (2007)) *can* improve soft skills and test scores in the short run, there is little evidence of long-run impacts. One exception is Dee and Penner (2019) who examine an intervention of which SEL training was a component. A meta-analysis of several growth mindset interventions, found small effects on academic achievement that *were not* mediated by self-reports of growth mindset (Sisk et al. 2018).

improved outcomes in high school and greater longer-run educational attainment.

School effects on SED cluster in two domains; promoting hard work and promoting social well-being. Accordingly, we compute leave-year-out estimates of school value-added (Chetty et al. 2014; Jackson 2018) on a hard work index and a social well-being index, and also standardized achievement tests. Using these leave-year-out estimates, we explore how attending a school that increases SED in other years (i.e., a high SED value-added school) improves both short- and longer-run outcomes. The standard deviation of estimated school effects on test scores and the SED measures are similar (between 0.06σ and 0.09σ), and these effects are all positively correlated with each-other.[3] However, conditional on test score value-added, high SED value-added schools improve attendance, reduce disciplinary incidents, improve course grades, reduce the number of school-based arrests, increase high school graduation, increase four-year college going, and increase college persistence. For this wide array of outcomes, using both SED and test score value-added more than *doubles* the variance of the explained school effect relative to using test score value-added alone. We can rule out most plausible sources of selection and we present several tests (such as within-sibling comparisons) that support a causal interpretation of our results.

We move beyond showing correlations between SED and long-run outcomes by documenting a wide array of short- and medium-run outcomes that are impacted by attending a school that *causally* improves SED. This work validates SED value-added as capturing school impacts on real skills and traits (as opposed to reporting biases). The analysis presents an important early step in our understanding of how schools may influence socio-emotional development of older adolescents, how it can be measured, and how this can be useful for policy.

II Data

We use administrative data from Chicago Public Schools (CPS). CPS is a large urban school district with 133 public (neighborhood /charter/ vocation/ magnet) high schools. CPS students are primarily African American (42%) and Latino (44%), and from families with disadvantaged economic backgrounds (86%). The main analysis data-set includes cohorts of 9th grade students who attended a neighborhood, charter, or magnet high school between 2011 and 2017 (*n*=157,630). When we examine longer-run outcomes, we focus on cohorts of 9th grade students between 2011 and 2014 (*n*=55,560) because these students are old enough to have attended college. Only first time 9th graders are included to eliminate sample selection biases due to grade repetition.

Measures. Our key variables are survey measures of SED[4]: interpersonal skills, school connectedness, academic engagement, grit, and study habits. Responses are collected by CPS on a

---

[3]These estimated magnitudes are in line with Loeb et al. (2018) and Fricke et al. (2019) who examine the variance of school effects on SEL growth. These important papers do not examine impacts on other outcomes.

[4]These measures were developed by the UChicago Consortium on School Research.

survey administered to students in 2008-09, and then every year from 2010-11 onward. Survey response rates were high on average (78%), however, nonresponse was higher for low-achievers (see Appendix Table S2). 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 for select years see Appendix Table S3).

Two of the SED survey measures relate to one's relationship with others in the school. The first of these is **Interpersonal skills**, which includes the following: *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.* The second such construct is **School Connectedness** which includes the following: *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.*

The other three SED survey measures capture students' orientation toward hard work. The first of these is **Academic Effort**, which 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.* The second construct is the perseverance facet of **Grit** which includes: *I finish whatever I begin. I am a hard worker. I continue steadily towards my goals. I don't give up easily.* The third construct is **Academic Engagement** which 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.*

We combine the social-related questions into a ***Social*** index and the hard-work-related questions into a ***Work Hard*** index. The construction of these indices was informed by conceptual frameworks for SEL. Appendix A shows that school effects on the individual survey constructs cluster into these two broader categories so that this categorization, in addition to being theory-driven, is justified by the data. To create each index we standardize each construct, compute the average of the included measures, and then standardize the index to be mean zero unit variance.

Test Scores: The "hard" skills measure in our data are standardized test scores.[5] 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 the ***Test Score*** index to be mean zero unit variance.

Our first longer run outcome is high school completion. About 79 percent of first time 9th

---

[5] 6th through 8th grade CPS students took the ISAT prior to 2014 and the NWEA or the PARCC thereafter. 9th graders took the EXPLORE assessments before 2014, and took the PARCC thereafter.

graders in CPS graduate high school. Our second key long run outcome is enrolling in college. Our college data come from the National Student Clearinghouse 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 (2010 through 2014 cohorts only). About 57 percent of first-time 9th graders enrolled in college. The data also include intermediate outcomes such as attendance, course grades, and discipline outcomes. The data are summarized in Table 1.

III Methods

Our analysis involves two key steps. First, we aim to identify those schools that improve students' SED and test scores. With this information in hand, we then estimate the effects of attending schools that improve these measures. We discuss each step in turn.

Step 1: Identifying School Impacts on SED and Test Scores

We use value-added models to estimate schools causal impacts on 9th-grade SED and test scores. Our value-added model seeks to isolate the causal effects of individual schools on student measure $q \in Q = \{$test scores, social well-being, hard work$\}$ by comparing measures at the end of 9th grade to those of similar students (with the same incoming test scores, survey measures, course grades, discipline, attendance, and demographics, all at the end of 8th grade) at other schools. A school's value-added on a measure q captures how much that school increases that measure between 8th and 9th grade relative to the observed changes for similar students (based on the attributes listed above) at other schools. Formally, we model the 9th 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. 8th and 7th grade test scores, surveys, discipline, and attendance), gender, ethnicity, and free-lunch status, in addition to the socio-economic status of the student census block proxied by average occupation status and education levels. Our full model also includes 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., $\theta_{jt,q}^{VA}$), we estimate (1) below, where $\upsilon_{ijt,q} = \theta_{j,q}^{VA} + \varepsilon_{ijt,q}$.

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

The student-level residual from this regression is $u_{ijt,q}$. 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

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

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

When using value-added to *predict* outcomes for a particular cohort, we exclude data for that same cohort when estimating value-added to avoid mechanical correlation. As in Jackson (2014), 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, the correlations in Appendix B show that estimates for more temporally proximate years are more highly correlated with each-other. As such, following Chetty et al. (2014), to improve precision, we allow greater weight to value-added from years close to the prediction year and less weight to years that are farther away temporally. 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 q measure between 8th and 9th grade relative to the improvements of similar students at other schools. We use leave-year-out predictions for all analyses, but for brevity, we refer to them simply as a school's value-added.

Step 2: Estimating Effect of Value-Added on Outcomes

To quantify the effect of attending a school with one standard deviation higher predicted value-added on outcomes, we regress each outcome on the standardized predicted value-added for the different indexes (plus controls). Specifically, where Y_{ijt} is an outcome, and $\hat{\mu}_{jt,q}$ is the standardized out-of-sample predicted value added on measure $q \in Q = \{$test scores, social well-being, hard work$\}$, we estimate the following model by OLS.

$$Y_{ijt} = \sum_{q \in Q} \beta_q \hat{\mu}_{jt,q} + \beta_1 Z_{ijt} + \tau_t + \varepsilon_{ijt} \tag{4}$$

All variables are as defined above and τ_t is a year fixed-effect. Standard errors are adjusted for clustering at the school level.[6] In some models we report estimates using only a single value-added

---

[6]Note that individuals with missing 8th grade measures (i.e., surveys or test scores) are given imputed values based on all other observed pre-treatment covariates. We regress each survey measure or test score on all observed pre-8th grade covariates. We then obtain predicted 8th grade survey measures and test scores based on this regression. Those with missing test scores or surveys in 8th grade are given this predicted value. Note that all results are similar when missing values of 8th grade measures are not imputed.

predictor, while in others we include several at once. For each regression model, we compute the variation in the outcome that can be explained by the included value-added estimates. Specifically, after estimating equation (4) we compute $\hat{F} = \sum_{q \in Q} \hat{\beta}_q \hat{\mu}_{jt,q}$. This is the impact of attending school j based on the linear relationship between the value-added estimates for that school j and the outcome. $Var(\hat{F})$ is therefore the variance of the predictable impact of schools based on the value-added estimates. By comparing the explained variance in models that include only test score value-added, only SED value added, and all the value-addeds, we can assess how much additional predictive power there is in each value-added measure over the others.

To take the estimated effect of value-added 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-value-added schools. To assess this, we estimate models like equation (4) predicting each observed covariate (when no covariates are included as controls). We find no discernible differences in the observed characteristics of those assigned to high and low value-added schools for any skill measure (See Appendix C) – indicating that this condition is satisfied.

IV   The Impact of School Value-Added on SED and Test-Scores

Here we establish that schools' SED value-addeds do, in fact, predict school impacts on SED. The coefficients in the top panel of Table 2 represent the effect of attending a school with one-standard deviation higher value-added (for each measure) on self-reported 9th grade social well-being. For brevity, we will refer to social well-being value-added as social value-added. As expected, social value-added is highly predictive of school impacts on social well-being. Using only social value-added, the coefficient of 0.0895 (p-value<0.01) in column (1) indicates that attending a school that has one standard deviation higher predicted social value-added (i.e., going from a school at the 85th percentile of the social value-added distribution versus one at the median) would improve social well-being by 8.9 percent of a standard deviation– compelling evidence that schools can, and do, impact reported social well-being and that these impacts are persistent over time.

While work hard value-added predicts 9th grade social well-being when on its own, in models that include both dimensions of SED value-added (column 4), work hard value-added has little additional explanatory power. In column 3, we explore the extent to which school impacts on test scores predict social well-being. In models with test score value-added only, the coefficient on test score impacts is 0.0347. This is much smaller than the predictive power of the social well-being value-added. In models that include test score value-added and both SED value-addeds (column 5), test score value-added does have some independent explanatory power. However, relative to using the SED value-addeds, adding test score value-added increases the explained variance by only 1.9 percent. That is, virtually all of the detectable variation in school impacts on self-reported social well-being (using all three value-addeds) is captured by social value-added.

7

We now turn to school impacts on the self-reported work hard dimension in middle panel of Table 2. In models with work hard value-added only, attending a school that has one standard deviation higher work hard value-added improves self-reported work hard in 9th grade by 6.3 percent of a standard deviation (p-value<0.01). Models that use social value-added only are similar to those that use work hard value-added only. In models that include school value-added on both SED measures simultaneously, the coefficient on work hard is the largest (0.0456) but that for social well-being is statistically significant (0.025). Relative to work hard value- added alone, adding social value-added increases the explained variance by a modest 8.9 percent. In models with test score value-added only (column 3), the coefficient on test score value-added is 0.0276 (p-value<0.01). This is much smaller than the predictive power of work hard value-added. In models including all three value-added measures (column 5), test score value-added has little independent explanatory power. Indeed, relative to using the SED value-addeds, adding test score value-added increases the explained variance by only 1.7 percent. In sum, the best predictor of a school's impact on work hard is school value-added on work hard – social well-being value-added has a small amount of independent predictive power for impacts on the work hard dimension, while test score value-added has no independent predictive power.

We also conduct similar analyses for 9th-grade test scores (lower panel of Table 2). In models that use test score value-added only (column 3), attending a school with one standard deviation higher predicted test score value-added increase 9th-grade test scores by 6.27 percent of a standard deviation (p-value<0.01).[7] Interestingly, value-added on SED measures are almost as good predictors of impacts on test scores as test score value-added. In models with both the SED and test score value-addeds, each measure independently predicts test scores in 9th grade. Relative to using test score value-added only, adding the SED value-addeds increase the explained variance by 42 percent. This stands in stark contrast to the pattern for SED measures– where the vast majority of a schools effect on SED is captured by the SED value-addeds. Remarkably, value-added on SED contains considerable independent explanatory power in explaining school impacts on test scores – suggesting that SED may be foundational for academic success. We now explore how independent variation in SED value-added matters for other outcomes that may mediate long-run impacts.

V Impacts on Potentially Mediating Outcomes

On Track: The first other outcome we explore is an "on track" indicator. This indicator identifies students as on-track if they earn at least five full-year course credits and no more than one semester F in a core course in their first year of high school. Students who are on-track in Chicago at the end of 9th grade are more than three times more likely to graduate high school in four years than

---

[7]These estimates are in line with Jackson (2013) that finds that the standard deviations of school effects in North Carolina are about 9 and 6 percent of a student standard deviation for math and English, respectively.

off-track students. Importantly this is a more accurate predictor of graduation than achievement test scores or background characteristics (Allensworth and Easton, 2005). We report the estimated impacts of school value-added in Table 3. Using each of the value-added measures individually, (columns 1 through 3) each value-added measure individually predicts being on-track. However, the estimated effects are larger for the SED measures. Specifically, attending a school with one standard deviation higher social value-added leads to a 1.9 percentage point increase in the likelihood of being on track (p-value<0.01), that for work-hard value-added is 2.07 percentage points (p-value<0.01), and that for test score value-added is 1.26 percentage points (p-value<0.01). Relative to a model with test score value-added only, the explained variance using both test score and SED value-added is about 5 times lager – indicating that (a) much of what schools may do to keep students on track to graduate high school is largely unmeasured by impacts on standardized tests, and (b) school impacts on self-reported survey measures capture much more of a school's impact on staying on track than impacts on test scores.

Course Grades: The second panel of Table 3 reports impacts on 9th grade GPA. In models with the individual value-addeds, the coefficients on social value-added is 0.0335 (p-value<0.1), that on work hard value-added is 0.0446 (p-value<0.05), and that for test score value-added is 0.0206 (p-value<0.05). The work hard dimension is more predictive of GPA than the social well-being dimension. In models that include all value-added measures simultaneously, none is statistically significant. However, the explained variance in the model with all three measures is 5 times as large as the test score value-added model, and 1.5 times as large as the models using only the SED value-added – reinforcing the importance of having measures of school impacts beyond test scores.

Attendance: Impacts on 9th grade absences are in the third panel of Table 3. The point estimates in column (1) through (3) indicate that each value-added measure individually predicts better attendance in 9th grade. However, social value-added explains more variance than the other two. A one standard deviation increase in test score value-added leads to 0.674 fewer absences. By comparison, a one standard deviation increase in social value-added reduces absences by 1.23 days (about twice as large) – an effect size of roughly 0.06σ or an 8.2 percent reduction compared to the average. Relative to the test score value-added only model, adding the SED value-addeds increases the explained variance by a factor of roughly 3. The fact that social well-being value-added is the most predictive of reduced absences suggests that more well-adjusted students, who feel a greater sense of belonging, are more likely to attend school (see Walton and Brady 2017 for a review of belonging research in academic contexts).

Discipline: Next we examine impacts on the number of disciplinary incidents. Most disciplinary incidents occur in grades 9 and 10, so we focus on 9th grade. The fourth panel of Table 3 reveals that social well-being value-added and test score value-added predict fewer incidents, while work hard value added does not. A one standard deviation increase in social value-added reduces

9

the number of incidents by 0.009 compared to only 0.0064 for test score value-added. The ratio of the explained variance using all three value-added measures to the variance explained using social value-added alone or test score value-added alone is 1.3 and 2.1, respectively. Consistent with this, in the combined model, only social value-added predicts impacts on the number of incidents.

School-Based Arrests: A key medium-run outcome that we examine is having a school-related arrest among those who are old enough to have graduated high school (bottom panel of Table 3). These are arrests for any activities conducted on school grounds, during off-campus school activities (including while taking school transportation), or due to a referral by a school official (link). During our sample period 4.1 percent of all students had a school-based arrests, 5.3 percent of males, and 7.9 percent of African American males. While disciplinary outcomes are correlated with being arrested ($\rho = 0.325$) one is not a very strong predictor of the other, and the long-term implications are very different. Indeed, roughly 20 percent of juvenile arrests in 2010 were school-based arrests (Kaba and Edwards, 2012), so that these have important long-term implications.

In models with each value-added individually, increasing social value-added reduces the likelihood of an arrest by 0.65 percentage points (*p*-value<0.01), increasing work hard value-added reduces the likelihood of an arrest by 0.746 percentage points (*p*-value<0.01) and increasing test-score value-added reduces the likelihood of an arrest by 0.451 percentage points (*p*-value<0.01). The SED value-added measures have greater predictive power than test score value-added. The ratio of the explained variance using all three value-addeds relative to using test score value-added alone is 2.76 – indicating that one can capture more than twice the variability in school impacts on arrests using the test score and SED value-addeds together than using test score value-added alone. Using all three, the standard deviation of the predicted school effect is 0.8 of a percentage point. Compared to the average, this represents a relative risk decrease of about twenty percent. Note that because these value-addeds make students more likely to remain in school (shown below), the documented reductions in school-based arrest rates are a lower bound. Looking among the two SED measures, work hard value-added has the strongest predictive power, but not greatly so.

In sum, schools that raise test scores are not always those that improve SED and *vice versa*. Schools' SED value-addeds are stronger predictors of impacts on all outcomes (excluding test scores) than test score value-added. Among the SED dimensions, some schools improve social well-being, while others promote working hard. Consistent with them impacting two different socio-emotional dimensions, work hard value-added is a stronger predictor of the academic outcomes (such as on-track and GPA), while social value-added is more predictive of impacts on non-academic behaviours (such as attendance and disciplinary incidents). Value-added on both SED dimensions are much more predictive of effects on school-based arrests than test score value-added. Because school-based arrests likely have important longer-run implications beyond academic outcomes, these results underscore the importance of evaluating schools effects on non-academic out-

10

comes (Beuermann et al., 2018).

VI Predicting Educational Attainment Outcomes

We report impacts on longer-run educational attainment outcomes in Table 4. In models that include a single value-added measure at a time, a one standard deviation increase in test score value-added (i.e., going from an average school to one at the 85th percentile of the effectiveness distribution) increases the likelihood of high school graduation by about 1 percentage point (column 3), that for social value-added is 1.47 percentage points (column 1), and that for hard work value-added is largest at 1.57 percentage points (column 2). The variance of school impacts on high school graduation explained by work hard value-added is 87 percent larger than that explained using test score value-added. In the model with all three value-added measures (column 5), the standard deviation of the predicted school impacts is 0.0185. Using the SED value-added increases the explained variance by over 160 percent relative to using test score value-added alone. Because much of the predictable school impacts on high school completion are captured by the surveys, using the test score value-added increases the explained variance by only 38 percent relative to using work hard value-added alone, and by only 25 percent using only the two SED value-addeds.

The second panel of Table 4 reports the impact on enrolling in college within 2 years of expected high-school graduation. In models that include a single value-added measure, a one standard deviation increase in test score value-added increases college-going by 1.36 percentage points, that for social value-added is 1.53 percentage points, and that for hard work value-added is 1.92 percentage points. If one uses all three value-added measures, the standard deviation of the predicted school impacts is 2.2 percentage points. Using the SED value-addeds increases the explained variance by 113 percent relative to using test score value-added alone. In contrast, using the test-score and SED value-addeds increases the explained variance by only 34 percent relative to using work hard value-added alone, and by 30 percent using only the two SED value addeds. In sum, relative to using test score value-added, the SED value addeds increases our ability to identify school impacts on longer-run outcomes considerably. We summarize the magnitude of the predicted school effect using the different value-addeds and using all three visually in Figure 1.

To delve deeper into the college results, we explore impacts on attending a 2-year or a 4-year college. We find no effects on 2-year college going (not shown). In contrast, we find large effects on four-year college going (third panel of Table 4). In the single value-added models, a one standard deviation increase in test score value-added increases 4-year college going by 1.83 percentage points, that for social value-added is larger at 2.67 percentage points, and that for work hard value-added is largest at 3.13 percentage points. A one standard deviation increased in predicted school impacts (based on all three value-added measures) would increase four-year college going by 3.48 percentage points. Using the SED value-addeds increases the explained variance by

over 180 percent relative to using test score value-added alone. In contrast, using all three measures versus using the SED value-addeds increases the explained variance by only 18 percent – reinforcing the relative importance of school impacts on socio-emotional development. The fact that the impacts on four-year college going are larger than those for high school completion shows that the college-going effect is not driven entirely by increased high school completion. Indeed, if all of the high school graduation effect lead to increased college-going, then it could explain *at most* $0.0185/0.0222 = 83\%$ of the overall college-going effect and $0.0185/0.0348 = 53\%$ of the four-year college going effect. This suggests that much of the increased college going is due to students who would have graduated high school being more likely to attend college.

Finally, the results in the bottom panel show estimated impacts on persisting in college beyond the first year. Each of the value-added estimates predict positive impacts on college persistence (i.e., being observed in a second year of college). The standard deviation of the predicted school impacts (using value-added on all three measures) on persistence is 1.7 percentage points. The same figure for college going is 2.2 percentage points, suggesting a persistence rate of roughly 1.7/2.2=0.77. This suggests that about 77 percent of the effects on college going remains one year after college entry. Given average persistence rates of about 72 nationally, these impacts suggest that the marginal college goers have one-year persistence rates that are similar to the average student.

Additional Testing For Selection

We have demonstrated that there is no selection to schools on observables. To show that our results are not driven by selection in **unobserved** dimensions we implement two additional tests (see Appendix C). First we show that students whose homes are zoned for schools with higher SED value-added attend schools with higher value-added and have better outcomes. Specifically, in a two-stage least squares (2SLS) framework using the value-added of the residentially zoned high school as an instrument for the value-added of the attended high school, we find similar effects of SED value-added on outcomes as the OLS estimates. This shows that our result are not driven by better students (in unobserved dimensions) selecting to better schools outside their attendance zones. Second, to rule out that the 2SLS results are driven by better families selecting to better neighborhoods, we show that our results are similar (albeit less precise) even when making comparisons among the subset of siblings that we can identify in our data. The 2SLS model is robust to selection within families, while the siblings model is robust to selection across families. As such, while none of these tests is dispositive in isolation, together they are compelling evidence that our estimated school impacts on SED and the main results reflect true causal impacts.

VII   Discussion and Conclusions

We identify persistent school impacts on two distinct dimensions, promoting social well-being and hard work, indicating that surveys can be used to reliably measure SED (net reporting biases) and that schools can foster SED beyond elementary school. This is the first paper to validate school impacts on SED out-of-sample, provide evidence that the estimates are causal, document potentially instrumental mediating outcomes (e.g., attendance, GPA), and show that both crime and longer-run educational attainment outcomes are influenced by attending a high school that improves SED. Moving beyond correlational evidence, this is the first paper to link schools' *causal* impacts on self-reported SED to longer-run outcomes. Our finding that school impacts on SED have larger effects on short- and long-run outcomes than schools' test score impacts has important implications for how policy-makers measure school quality. The analysis presents an important early step in our understanding of how schools may influence socio-emotional development. To better inform policy, further investigation is needed into what school practices, policies, and conditions tend to promote SED, and how those may differ from those that promote test score growth.

References

Sule Alan, Teodora Boneva, and Seda Ertac. Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit*. *The Quarterly Journal of Economics*, apr 2019.

Elaine M Allensworth and John Q Easton. The on-track indicator as a predictor of high school graduation, 2005.

Diether W. Beuermann, Kirabo C. Jackson, Laia Navarro-Sola, and Francisco Pardo. What is a Good School, and Can Parents Tell? The Multidimensionality of School Output and Parental Preferences. *National Bureau of Economic Research*, 2018. doi: 10.3386/w25342.

Lisa S. Blackwell, Kali H. Trzesniewski, and Carol Sorich Dweck. Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78(1):246–263, jan 2007. ISSN 00093920.

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.

CASEL. CSI Resources: Frameworks & Competencies. URL https://casel.org/csi-resources-frameworks-competencies/.

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.

Geoffrey L Cohen, Julio Garcia, Nancy Apfel, and Allison Master. Supporting online material for reducing the racial achievement gap: A social-psychological intervention. *Science*, 313:1307–1310, sep 2006. ISSN 1095-9203.

Marcus Credé, Michael C. Tynan, and Peter D. Harms. Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*, 113(3):492–511, sep 2017.

Thomas Dee and Emily Penner. My Brother's Keeper? The Impact of Targeted Educational Supports. Technical report, 2019.

David J. Deming. The Growing Importance of Social Skills in the Labor Market*. *The Quarterly Journal of Economics*, 132(4):1593–1640, nov 2017.

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.

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

Carol S Dweck and David S Yeager. Mindsets: A View From Two Eras. *Perspectives on psychological science : a journal of the Association for Psychological Science*, 14(3):481–496, 2019. ISSN 1745-6924. doi: 10.1177/1745691618804166.

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.

Hans Fricke, Susanna Loeb, Robert Meyer, Andrew Rice, and Libby Pier. Measuring School Contributions to Growth in Social-Emotional Learning — Policy Analysis for California Education, 2019. URL https://edpolicyinca.org/publications/stability-school-contributions-student-social-emotional-learning-gains.

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.

Geoff Hing and Smith Richards Jenniver. Chicago school choice in charts - Chicago Tribune. URL https://www.chicagotribune.com/ct-chicago-school-neighborhood-enrollment-charts-20160106-htmlstory.html.

C. Kirabo Jackson. Match quality, worker productivity, and worker mobility: Direct evidence from teachers. *Review of Economics and Statistics*, 95(4):1096–1116, 2013.

C. Kirabo Jackson. Teacher Quality at the High School Level: The Importance of Accounting for Tracks. *Journal of Labor Economics*, 2014. ISSN 0734-306X.

C. Kirabo Jackson. What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes. *Journal of Political Economy*, may 2018. doi: 10.3386/w22226.

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.

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.

Susanna Loeb, Michael S Christian, Heather J Hough, Robert H Meyer, Andrew B Rice, and Martin R West. School Effects on Social-Emotional Learning: Findings from the First Large-Scale Panel Survey of Students. Technical report, 2018. URL https://www.edpolicyinca.org/sites/default/files/SEL{_}School{_}Effects{_}May-2018.pdf.

William Revelle. Experimental approaches to the study of personality, 2007.

Kaili Rimfeld, Yulia Kovas, Philip S. Dale, and Robert Plomin. True grit and genetics: Predicting academic achievement from personality. *Journal of Personality and Social Psychology*, 111(5):780–789, nov 2016. ISSN 00223514. doi: 10.1037/pspp0000089.

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(4):549–571, apr 2018.

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

Gregory M. Walton and Shannon T. Brady. The many questions of belonging. In *Handbook of Competence and Motivation: Theory and Application*, pages 272–293. 2017. ISBN 9781462536030.

Martin R. West, Matthew A. Kraft, Amy S. Finn, Rebecca E. Martin, Angela L. Duckworth, Christopher F. O. Gabrieli, and John D. E. Gabrieli. Promise and Paradox. *Educational Evaluation and Policy Analysis*, 38(1):148–170, mar 2016.

Tables and Figures

Table 1: Summary Statistics

	Analytic Sample - Short Term (2011 -2017)		Analytic Sample - Long Term (2011-2014)	
	mean	SD	mean	SD
Demographics				
Female	0.502	0.500	0.500	0.500
Special education (IEP)	0.182	0.386	0.169	0.375
Free lunch	0.777	0.416	0.769	0.422
Reduced-price lunch	0.0753	0.264	0.0856	0.280
White	0.0898	0.286	0.0876	0.283
Black	0.403	0.490	0.427	0.495
Native American	0.00169	0.0411	0.00182	0.0426
Asian/Pacific Islander	0.0327	0.178	0.0327	0.178
Latino	0.460	0.498	0.440	0.496
9th *grade Intermediate Outcomes*				
Days Absent in 9th Grade	14.999	18.633	17.584	21.194
GPA in 9th Grade	2.426	1.007	2.264	1.043
Diciplinary Incidents in 9th Grade	0.0769	0.419	0.106	0.452
Long-term Outcomes				
Any school-Based arrest			0.041	0.197
Graduation			0.743	0.437
Enrolled in any college within 2 years			0.532	0.499
Enrolled in a 2 year college within 2 years			0.276	0.447
Enrolled in a 4 year college within 2 years			0.343	0.475
Observations	157630		55560	

Table 2: Effects on 9th Grade Measures

	1	2	3	4	5
	Outcome = Social Well-Being in 9th Grade (124,685 Obs.)				
Social Value-Added	0.0895***			0.0921***	0.0901***
	(0.00871)			(0.00846)	(0.00859)
Work Hard Value-Added		0.0614***		-0.00354	-0.00610
		(0.00931)		(0.0106)	(0.0109)
Test Scores Value-Added			0.0347***		0.0123**
			(0.00589)		(0.00574)
Predicted Variance	0.00787	0.00374	0.00151	0.00789	0.00804
	Outcome = Work Hard in 9th Grade (124,487 Obs.)				
Social Value-Added	0.0597***			0.0257***	0.0243***
	(0.00833)			(0.00661)	(0.00670)
Work Hard Value-Added		0.0637***		0.0456***	0.0438***
		(0.00911)		(0.0113)	(0.0118)
Test Scores Value-Added			0.0276***		0.00846
			(0.00554)		(0.00544)
Predicted Variance	0.00351	0.00402	0.000957	0.00438	0.00445
	Outcome = Test Scores in 9th Grade (102,235 Obs.)				
Social Value-Added	0.0577***			0.0359***	0.0296***
	(0.0107)			(0.0120)	(0.00982)
Work Hard Value-Added		0.0563***		0.0300**	0.0218*
		(0.0113)		(0.0140)	(0.0122)
Test Scores Value-Added			0.0627***		0.0504***
			(0.0115)		(0.0112)
Predicted Variance	0.00327	0.00314	0.00493	0.00369	0.00703

Robust standard errors adjusted for clustering at the school-level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Results are based on regression of 9th grade measures on out-of-sample Social, Work hard, and Test Score 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 the all demographics and lagged measures, as well as year fixed effects. Missing 8th grade measures were imputed using 7th grade measures and demographic characteristics. Sample size varies by outcome due to missingness. Results do not change if we restrict results to a balanced sample.

Table 3: Effects on Other Outcomes and Behaviors

	1	2	3	4	5
	On Track in 9th Grade (obs. 114,512)				
Social Value-Added	0.0190***			0.0102	0.00924
	(0.00636)			(0.00661)	(0.00661)
Work Hard Value-Added		0.0208***		0.0129	0.0115
		(0.00758)		(0.00874)	(0.00866)
Test Scores Value-Added			0.0126***		0.00911**
			(0.00379)		(0.00374)
Predicted Variance	0.000354	0.000427	0.000198	0.000453	0.000589
	GPA in 9th Grade (obs. 122,948)				
Social Value-Added	0.0335*			0.00647	0.00506
	(0.0185)			(0.0166)	(0.0165)
Work Hard Value-Added		0.0446**		0.0396*	0.0375
		(0.0221)		(0.0232)	(0.0228)
Test Scores Value-Added			0.0206**		0.0133
			(0.00990)		(0.00860)
Predicted Variance	0.00110	0.00198	0.000533	0.00196	0.00227
	Days Absent in 9th Grade (Obs. 157,630)				
Social Value-Added	-1.232***			-0.979***	-0.922***
	(0.269)			(0.306)	(0.300)
Work Hard Value-Added		-1.044***		-0.346	-0.273
		(0.270)		(0.307)	(0.302)
Test Scores Value-Added			-0.674***		-0.359*
			(0.219)		(0.213)
Predicted Variance	1.493	1.081	0.568	1.543	1.690
	Number of Disciplinary Incidents in 9th Grade (obs. 160,148)				
Social Value-Added	-0.00927**			-0.0102**	-0.00950*
	(0.00421)			(0.00511)	(0.00509)
Work Hard Value-Added		-0.00598		0.00129	0.00224
		(0.00412)		(0.00469)	(0.00464)
Test Scores Value-Added			-0.00650**		-0.00468
			(0.00322)		(0.00312)
Predicted Variance	8.45e-05	3.54e-05	5.28e-05	8.54e-05	0.000111
	Ever Received School Based Arrest (obs. 55,560)				
Social Value-Added	-0.00650***			-0.00295	-0.00284
	(0.00242)			(0.00305)	(0.00303)
Work Hard Value-Added		-0.00746***		-0.00536*	-0.00450
		(0.00240)		(0.00310)	(0.00301)
Test Scores Value-Added			-0.00451**		-0.00295
			(0.00187)		(0.00197)
Predicted Variance	4.16e-05	5.51e-05	2.55e-05	5.96e-05	7.06e-05

Robust standard errors adjusted for clustering at the school-level in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Results are based on regression of behaviors on Social, Work hard, and Test Score 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 the all demographics and lagged measures, as well as year fixed effects.

Table 4: Effects on Longer-Run Educational Attainment

	1	2	3	4	5
	Graduate High-School (obs. 82,146)				
Social Value-Added	0.0147***			0.00758*	0.00667
	(0.00354)			(0.00425)	(0.00428)
Work Hard Value-Added		0.0157***		0.0103**	0.00914*
		(0.00378)		(0.00480)	(0.00496)
Test Scores Value-Added			0.0102***		0.00702***
			(0.00248)		(0.00240)
Predicted Variance	0.000211	0.000246	0.000131	0.000274	0.000344
	Enroll in College within 2-Years of Expected high School Graduation (obs. 55,560)				
Social Value-Added	0.0153***			0.00483	0.00448
	(0.00506)			(0.00445)	(0.00440)
Work Hard Value-Added		0.0192***		0.0158**	0.0129**
		(0.00589)		(0.00641)	(0.00573)
Test Scores Value-Added			0.0136***		0.00980**
			(0.00465)		(0.00384)
Predicted Variance	0.000230	0.000366	0.000232	0.000378	0.000494
	Enroll in 4- Year College within 2-Years of Expected high School Graduation (obs. 55,560)				
Social Value-Added	0.0267***			0.0113	0.0108
	(0.00818)			(0.00704)	(0.00719)
Work Hard Value-Added		0.0313***		0.0233**	0.0198**
		(0.00940)		(0.0103)	(0.00972)
Test Scores Value-Added			0.0183***		0.0117***
			(0.00598)		(0.00435)
Predicted Variance	0.000700	0.000969	0.000420	0.00103	0.00121
	Persist in College within 3-Years of Expected high School Graduation (obs. 55,560)				
Social Value-Added	0.0113***			0.00314	0.00283
	(0.00417)			(0.00402)	(0.00403)
Work Hard Value-Added		0.0145***		0.0122**	0.00970*
		(0.00494)		(0.00567)	(0.00517)
Test Scores Value-Added			0.0115***		0.00868**
			(0.00424)		(0.00361)
Predicted Variance	0.000125	0.000208	0.000164	0.000213	0.000303

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

Notes: Results are based on regression of educational attainment on out-of-sample Social, Work hard, and Test Scores 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 the all demographics and lagged measures, as well as year fixed effects.

Figure 1. Size of Predicted Impacts on Longer run Outcomes: by Model

Standard Deviation of Explained School Impacts on Longer-run Outcomes Using Test Score Value-Added and SED Value-Added
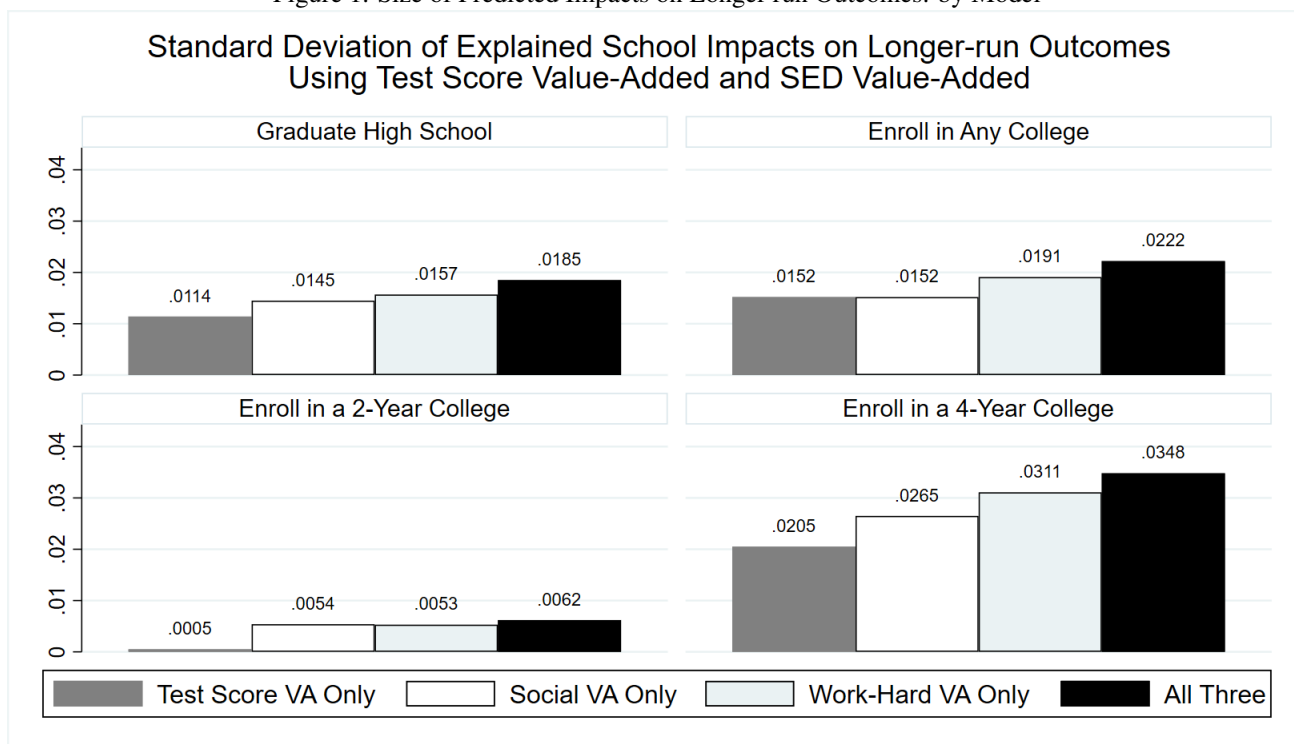
Notes: For each model, we take the fitted values based only on the value-added estimates as the predicable portion of school impacts on each outcome. We then compute the standard deviation of this predicable variation under each model. Note that because the school value-addeds are standardized to be unit variance across all schools (as opposed to across all students in the regression samples) the standard deviations of the predicted effects are not *exactly* equal to the reported effects of increasing value-added by one standard deviation.

Supplementary Material

Table S1: Testing Selection on Observables

	1	2	3	4	5	6	7	8	9	10	11
	Math 8	ELA 8	Emotional Health 8	Academic Engagement 8	Grit 8	School Connectedness 8	Study Habits 8	Absences 8	GPA 8	Incidents 8	Suspensions 8
Test Scores Value Added	0.00972	0.0150	0.00851	0.00637	0.0168	0.00783	0.0123	-0.0668	-0.00911	0.00180	0.0282
	(0.0325)	(0.0306)	(0.00766)	(0.00641)	(0.0123)	(0.0110)	(0.0103)	(0.398)	(0.0188)	(0.00361)	(0.0296)
Work Hard Value Added	0.0327	0.0183	0.00614	0.00430	0.00947	0.0226	0.0130	0.397	0.0192	-0.00158	0.00764
	(0.0597)	(0.0587)	(0.0119)	(0.0126)	(0.0236)	(0.0179)	(0.0162)	(0.846)	(0.0416)	(0.00667)	(0.0568)
Social Value Added	-0.00240	0.00743	-0.00627	0.00719	0.00490	-0.00467	0.00552	-0.631	0.00782	0.000395	-0.0276
	(0.0610)	(0.0577)	(0.0112)	(0.0127)	(0.0180)	(0.0153)	(0.0147)	(0.651)	(0.0383)	(0.00680)	(0.0581)

	12	13	14	15	16	17	18	19	20	21	22
	Female	Special Ed	Free Lunch	Reduced Lunch	White	Black	Native	Asian/Pacific Islander	Latinx	Multiracial	Asian
Test Scores Value Added	0.00157	-8.18e-05	-0.0119	8.71e-05	0.00665	0.0257	-8.69e-05	-0.00108	-0.0308	0.000260	-0.000418
	(0.00475)	(0.00313)	(0.00993)	(0.00240)	(0.00658)	(0.0198)	(0.000156)	(0.00253)	(0.0197)	(0.000173)	(0.000793)
Work Hard Value Added	-0.000223	-0.00169	-0.00763	-0.00212	0.0122	-0.00453	-3.66e-07	0.00909	-0.0189	0.000181	0.00112
	(0.0115)	(0.00710)	(0.0201)	(0.00520)	(0.0169)	(0.0449)	(0.000350)	(0.00577)	(0.0391)	(0.000352)	(0.00211)
Social Value Added	0.00315	0.00631	-0.000798	0.00170	-0.00439	0.00434	0.000239	-0.00677	0.0102	-0.000184	-0.00140
	(0.0112)	(0.00582)	(0.0220)	(0.00404)	(0.0161)	(0.0388)	(0.000292)	(0.00561)	(0.0332)	(0.000343)	(0.00192)

	23	24	25	26	27	28	29	30	31	32	33
	Math 7	ELA 7	Emotional Health 7	Academic Engagement 7	Grit 7	School Connectedness 7	Study Habits 7	Absences 7	GPA 7	Incidents 7	Suspensions 7
Test Scores Value Added	0.0103	0.0160	0.00192	0.00238	0.0131	0.00607	0.00978	0.0884	5.45e-05	0.00189	0.0185
	(0.0321)	(0.0311)	(0.00501)	(0.00403)	(0.00912)	(0.00549)	(0.00818)	(0.449)	(0.0221)	(0.00262)	(0.0275)
Work Hard Value Added	0.0275	0.0143	0.0108	0.000928	0.00448	0.0136	0.0116	0.118	0.0192	0.00261	0.0154
	(0.0612)	(0.0612)	(0.00899)	(0.00820)	(0.0168)	(0.00955)	(0.0116)	(0.987)	(0.0484)	(0.00533)	(0.0563)
Social Value Added	-0.00135	0.00615	-0.00986	-0.00150	0.00522	-0.00577	-0.00413	-0.435	0.00142	-0.00333	-0.0379
	(0.0624)	(0.0605)	(0.00831)	(0.00774)	(0.0138)	(0.00857)	(0.0106)	(0.866)	(0.0454)	(0.00550)	(0.0569)

Robust standard errors adjusted for clustering at the school-level in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: Results are based on regression of each covariate on out-of-sample Social, Work hard, and Test Scores Value Added, and year fixed effects. No controls are included in these models.
Missing 8th grade measures were imputed using 7th grade measures and demographic characteristics. The sample includes all students in neighborhood, charter, and magnet schools, between 2011 and 2017.

Table S2: Summary Statistics for Survey Completers and Non-Completers

Variables	Full Sample		Completed Surveys in 9th Grade		Did Not Complete Surveys in 9th Grade	
	mean	SD	mean	SD	mean	SD
9th grade						
Math in 9th Grade	0.0101	(0.991)	0.0630	(0.975)	-0.162	(1.023)
English in 9th Grade	0.00459	(0.995)	0.0605	(0.986)	-0.184	(1.001)
Emotional health in 9th Grade	-0.0529	(0.989)	-0.0473	(0.987)	-0.219	(1.043)
Academic engagement in 9th Grade	0.113	(1.061)	0.113	(1.060)	0.138	(1.111)
Grit in 9th Grade	-0.0122	(0.950)	-0.00633	(0.946)	-0.199	(1.035)
School connectness in 9th Grade	-0.0351	(0.956)	-0.0313	(0.953)	-0.138	(1.042)
Study habits in 9th Grade	-0.0178	(1.004)	-0.0132	(1.003)	-0.175	(1.045)
Absences in 9th Grade	14.998	(18.633)	12.9	(15.538)	21.49	(24.875)
GPA in 9th Grade	2.426	(1.007)	2.503	(0.974)	2.154	(1.071)
Days Suspended in 9th Grade	0.804	(3.293)	0.633	(2.772)	1.334	(4.501)
Incidents in 9th Grade	0.0769	(0.419)	0.0607	(0.356)	0.127	(0.568)
Indices						
Test Scores in 9th Grade	-0.00746	(0.992)	0.0485	(0.977)	-0.191	(1.018)
Workhard in 9th Grade	0.182	(0.988)	0.190	(0.981)	-0.0221	(1.128)
Social in 9th Grade	0.00430	(0.998)	0.00975	(0.994)	-0.141	(1.095)
8th grade						
Math in 8th Grade	0.202	(0.938)	0.261	(0.930)	0.0202	(0.939)
English in 8th Grade	0.207	(0.937)	0.267	(0.916)	0.0197	(0.974)
Emotional health in 8th Grade	0.0704	(0.896)	0.0827	(0.900)	0.0321	(0.882)
Academic engagement in 8th Grade	0.267	(0.913)	0.272	(0.921)	0.252	(0.888)
Grit in 8th Grade	0.0442	(0.836)	0.0525	(0.842)	0.0183	(0.817)
School connectness in 8th Grade	0.140	(0.900)	0.144	(0.906)	0.124	(0.882)
Study habits in 8th Grade	0.152	(0.889)	0.165	(0.900)	0.113	(0.853)
Absences in 8th Grade	8.606	(8.003)	7.972	(8.003)	10.629	(11.27)
GPA in 8th Grade	2.803	(0.781)	2.850	(0.772)	2.659	(0.791)
Days Suspended in 8th Grade	0.431	(1.812)	0.348	(1.546)	0.690	(2.442)
Incidents in 8th Grade	0.0636	(0.334)	0.0517	(0.286)	0.100	(0.451)
Demographics						
Female	0.502	(0.500)	0.512	(0.500)	0.469	(0.499)
Special education (IEP)	0.182	(0.386)	0.157	(0.364)	0.258	(0.438)
Free lunch	0.777	(0.416)	0.768	(0.422)	0.803	(0.397)
Reduced-price lunch	0.0753	(0.264)	0.0786	(0.269)	0.0650	(0.247)
White	0.0898	(0.286)	0.0951	(0.293)	0.0734	(0.261)
Black	0.403	(0.490)	0.374	(0.484)	0.491	(0.500)
Native	0.00169	(0.0411)	0.00163	(0.0403)	0.00190	(0.0435)
Asian/Pacific Islander	0.0327	(0.178)	0.0364	(0.187)	0.0215	(0.145)
Latino	0.460	(0.498)	0.480	(0.500)	0.398	(0.489)
Multiracial	0.00147	(0.0383)	0.00149	(0.0386)	0.00138	(0.0371)
Long-term						
Ever arrested in school	0.041	(0.197)	0.039	(0.194)	0.048	(0.213)
HS Graduation	0.745	(0.436)	0.781	(0.414)	0.646	(0.478)
Enrolled in any college within 2 years	0.536	(0.499)	0.580	(0.494)	0.414	(0.493)
Enrolled in a 2 year college within 2 years	0.278	(0.448)	0.297	(0.457)	0.225	(0.418)
Enrolled in a 4 year college within 2 years	0.346	(0.476)	0.379	(0.485)	0.251	(0.434)
N	157630		119133		38497	

Table S3: Psychometric Properties of SED measures: 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. 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.

A Correlations Across Measures

Given that we have school impacts on several skill measures, it if helpful to see if they are related to each-other. To explore this, we report the correlations between the school impacts on the various skill measures in 9th grade. These are reported in the lower panel of Table S4. The estimated school impacts (across the 133 schools) are all positive. That is, schools that improve one skill measure tend to improve the others. While all the correlations are positive, some variable are more closely related to others. For example, the correlation between math test score impacts and English test score impacts is 0.657. This is consistent with other studies. Interestingly, the correlation between test score impacts and impacts on the SED measures are reasonably large. The correlation between math value-added and the value-added on the individual surveys constructs are between 0.31 and 0.46. The correlations are similar for school impacts on English scores and the various surveys. To explore the extent to which schools tend to cluster in their impacts, we also conduct exploratory factor analysis. The factor ladings are reported in Table S5. The factors models suggests three distinct underlying factors. The first factor is most strongly related to the two test score impacts. the second factor is most strongly related to the survey measures relating to academic motivation and effort (grit, study habits, and academic engagement). The third factor is most strongly related to the two survey measures that relate to social well being (belonging, and social skills). While there is not complete separation of variables in the model, it clearly identified three district dimensions of school output.

Based on these results, we create a social index by combining the two social survey questions, we create a hard working index by combining the academic engagement questions, and we create a test score index by combining the two test scores. We also create a an overall survey index that combines all the survey questions. To create each index we (1) standardize each measure, then we (2) compute the arithmetic mean across the included measures, and (3) standardize the combined index to be mean zero unit variance. The correlations across school impacts on the indexes are in the lower rows on Table S4. The two survey dimensions are highly correlated, but not perfectly so. The correlations between school impacts on the social index and the hard work index is 0.6. As we will show below, while there is much shared variation, there is explanatory power in the independent variation in each index. The correlation between school impacts on test scores and impacts on the summary survey indexes are about 0.47. This suggests that any single index could likely predict improved outcomes in an average sense. However, the policy relevant question is whether the school impacts on surveys can provide additional information, and the extent to which school impacts on self-reported survey measures reflect improvement in crease skills.

Table S4: Correlations of School Value Added: Over time and Across Outcomes

	Correlations of Value-Added Within Outcomes Across Time										
lag	Math Value Added	ELA Value Added	Grit Value Added	Emotional Health Value Added	Academic Engagement Value Added	School Connectedness Value Added	Study Habits Value Added	Test Scores Value Added	Surveys Value Added	Social Value Added	Workhard Value Added
t+1	.43	.326	.265	.203	.234	.433	.198	.417	.334	.36	.274
t+2	.281	.201	.103	.06	.176	.243	.178	.284	.172	.166	.192
t+3	.113	.048	.045	.113	.11	.104	.039	.109	.098	.09	.103
t+4	.105	.154	.118	.167	.132	.239	.209	.18	.213	.212	.188

Correlations of Average SchoolLevel Value-Added Across Outcomes											
Math Value Added	1.										
ELA Value Added	.657	1.									
Grit Value Added	.308	.308	1.								
Emotional Health Value Added	.359	.283	.605	1.							
Academic Engagement Value Added	.306	.131	.425	.386	1.						
School Connectedness Value Added	.456	.431	.487	.675	.378	1.					
Study Habits Value Added	.424	.442	.731	.482	.516	.49	1.				
Test Scores Value Added	.919	.896	.328	.342	.238	.478	.464	1.			
Surveys Value Added	.482	.409	.8	.792	.71	.793	.808	.478	1.		
Social Value Added	.45	.398	.585	.886	.415	.94	.528	.455	.864	1.	
Workhard Value Added	.419	.343	.813	.574	.815	.535	.881	.409	.919	.6	1.

Table S5: Factor Loading for different Value- Added Measures

	Factor 1: Test Scores	Factor 2: Academic Motivation	Factor 3: Social Health
Math Value-Added	**0.7149**	0.2051	0.1878
ELA Value-Added	**0.7154**	0.1859	0.1583
Bellonging Value-Added	0.3774	0.4281	**0.5518**
Social Skills Value-Added	0.1825	0.4135	**0.651**
Academic Engagement Value-Added	0.2806	**0.5513**	0.1583
Grit Value-Added	0.1189	**0.729**	0.374
Academic Effort Value-Added	0.2713	**0.7489**	0.2557

These estimates are based on an exploratory factor analysis using the Bartlett Method. The rotated factor loadings are reported.

B Stability of School Effects Over Time

The first test of whether schools have systematic impacts on survey measures is whether school impacts in one year are correlated with impacts in other years. As pointed out in Jackson (2014) in the presence of transitory shocks, standard analysis of variance may lead one to conclude that schools have systematic impacts when in fact they do not. As such, an alternate approach is to explore the extent to which school value-added is persistent over time. The basic logic is that if a school is able to systematically improve students' SED, it should be able to do so in multiple years. To test this, for each of the SED or test score measures in 9th grade, we computed our annual value-added estimates from (4) and correlated them for the same schools over time. We report the correlations between a school's value-added in year t and in years $t+1$ through $t+5$ in the top panel of Table S4.

We start with the test score impacts. Looking at math scores, the correlation between a schools impact in adjacent years (i.e., years t and $t+1$) on math scores is 0.43. Specifically, a school that is above average at raising math scores in one year is likely to be above average in the following year. However, the relationship is far from one-for-one. Specifically, a school that was at the 98th percentile of math score value-added in one year, would be expect to be at the 81th percentile the following year. Looking at relationships two years apart, the correlation falls to 0.281. This means that a school that was at the 98th percentile of math score value-added in one year, would be expect to be around the 70th percentile two years later. Looking at more than three years out, the correlation stabilizes at around 0.12. This suggests that a school that was at the 98th percentile of math score value-added in one year, would be expect to be around the 61st percentile three to five years later. This is clear evidence that school impacts do persist over time, but that the extent of the persistence is modest. It is important to note that school value-added is estimated with error so that these correlations are a lower bound on what we would observe if school impacts were perfectly measured (i.e., with no errors).

The basic pattern of persistent effects that are stronger for adjacent years is observed for all the SED and test score measures. For grit, the one-year correlation is 0.265. This persistence level is smaller than that for math, suggesting that school impacts on grit are nosier than those in math test scores. This one-year correlation is similar for the other survey measures other than school connectedness (which has a correlation of 0.433). To put these correlations into perspective, a correlation of 0.265 implies that a school that was at the 98th percentile of grit, effort, academic engagement, or social skills value-added in one year, would be expect to be at the 70th percentile the following later. As with the test score impacts, the persistence is lower with greater temporal distance. For all the survey measures, the correlation between impacts in year t and year $t+3$ is about 0.12. This implies that a school that was at the 98th percentile of value-added for any of the

27

survey measures in one year, would be expect to be at the 60th percentile three later. While these persistence effects are modest, they are statistically significantly different from zero – compelling evidence that schools do have systematic impacts on self-reported measures of socio-emotional development.

In sum, the results indicate that (a) both SED and test-score value-added in one year is predictive of value added in the next, and (b) value added estimates for temporally close years will be a better predictor of value added than years that are temporally distant. In essence, this reveals that school value added (on both test scores and SED) has some persistence over time but does exhibit some "drift". We will exploit this fact when predicting a schools impact on hard or soft skills based on impact on other years.

C   Testing For Selection

Because students are not randomly assigned to schools, there is a concern that our estimated value-addeds are related to unobserved predictors of outcomes so that our estimates are biased. While there is no way to prove that the value-addeds of the attended schools are unrelated to unobserved determinants of outcomes, we present several test to show that this is likely satisfied in our setting.

No Selection on Observables

First, to show that our school SED value-addeds are not biased, we show that they are unrelated to observed determinants of student outcomes. That is, similar to a test for random assignment, we show that there is balance of covariates between high and low value-added schools. To show this, we estimate the following model by Ordinary Least Squares (OLS).

$$Z_{ijt} = \sum_{q \in Q} \pi_q \hat{\mu}_{jt,q} + \tau_t + \varepsilon_{ijt} \tag{5}$$

The parameter estimates of π_q provide a test of whether the observed covariate (Z_{ijt}) is correlated with the value-added on dimension q. If strong observable predictors of the outcomes are unrelated to our school value-added estimates (i.e. $\pi_q = 0$ for all covariates), then it is plausible that *unobservable* predictors are also unrelated to our value-added estimates so that our estimates are unbiased. We show evidence of this empirically in Table S1. If our estimated school value-addeds were correlated with student characteristics, then the coefficient on the value-added predictors would be significantly different from zero. We estimate this model with all three predictors across 33 pre-treatment student characteristics – resulting in 99 estimates. None of these point estimates is statistically significant at the 5 percent level. Remarkably, this is also true across the 99 models that have each value-added dimension individually – indicating no selection on observables. 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. In Chicago, almost two-thirds of children attend schools other than their zoned school (Hing and Jenniver), so that this is a potential concern. To show that this does not bias our results, we propose an instrumental variables approach that instruments for the value-added of the school attended with the value-added of the residentialy assigned school. This approach eliminates all selection to non-zoned school that could

have led to bias.

The first stage regressions are strong for all value-addeds – yielding first stage F-statistics above 20. The two-stage-least-squares (2SLS) regressions are reported in Table S6. Looking at work hard value-added (middle panel), a comparison of columns 1 and 2 reveals that the OLS and 2SLS models for work hard in 9th grade are similar and not statistically distinguishable from one-another. Looking at columns 4 and 5, and then 7 and 8, reveals that the OLS and 2SLS models of work hard value-added are also very similar for high school completion and college-going. Owing to much larger standard errors in the IV model, the point estimate on college-going is not significant, but it is very similar to the OLS estimate. Note however that the effect on high school graduation is significant at the five percent level in both the OLS the 2SLS models. The lower panel present a very similar pattern for social well-being value-added.[8] In sum, our SED value-added measures do 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 value-added. We address this possibility.

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 are able to 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 11,640 families in which more than one sibling is observed in 9th grade. Of these that have multiple children old enough to have graduated from high school we have 3352 such families. For those old enough to have enrolled in college, this number falls to 1393 families. 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. 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 (4). The within-family estimates are presented in Table S6. The sibling models are presented in columns (3), (6) and (9). As one can see, in the lower two panels, the estimated impacts of SED value-added are robust to the inclusion of the family fixed effects. That is, while the standard errors are much larger in the family fixed effects models, the point estimates for impacts on both work hard and social well-being value added on high school completion and college going

---

[8]The top panel reports impacts for test score value-added. Unlike the SED value-addeds, these models are less consistent across specification. However, test score value-added is not the focus of our study.

are very similar to the OLS models.[9] This indicates that selection of families does not drive the estimates.

Taken together, we show that (a) our value-added estimates are unrelated to all 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.

---

[9]The top panel reports impacts for test score value-added. Unlike the SEL value-added, these models are less consisting across specification. However, test score value-added is not the focus of our study.

Table S6: Selection on Unobservables Models

	1	2	3	4	5	6	7	8	9
	Test Scores			Graduate High School			Enroll in College		
Test Score Value-Added	0.0605***	0.0489***	0.0275**	0.00940***	0.00460	0.00631	0.0134**	-3.19e-05	0.00232
	(0.0123)	(0.00501)	(0.0131)	(0.00293)	(0.00292)	(0.00487)	(0.00538)	(0.00399)	(0.00520)
Observations	95,206	91,237	14,341	76,352	72,737	7,153	51,791	48,832	2,968
OLS	X			X			X		
School Assignment IV		X			X			X	
First-Stage F-statistic		1941			2055			2083	
Sibling Fixed Effects			X			X			X
Number of Families			6666			3352			1393
	Work Hard			Graduate High School			Enroll in College		
Work Hard Value-Added	0.0663***	0.0765***	0.0261*	0.0170***	0.0269***	0.0168**	0.0220***	0.0145	0.0181
	(0.00993)	(0.0107)	(0.0144)	(0.00420)	(0.00735)	(0.00718)	(0.00722)	(0.00977)	(0.0161)
Observations	116,021	111,993	25,292	76,208	72,616	7,136	51,647	48,711	2,958
OLS	X			X			X		
School Assignment IV		X			X			X	
First-Stage F-statistic		727.7			842.8			758.9	
Sibling Fixed Effects			X			X			X
Number of Families			11601			3344			1388
	Social Well-Being			Graduate High School			Enroll in College		
Social Value-Added	0.0904***	0.0811***	0.0624***	0.0180***	0.0349***	0.0113*	0.0206***	0.0320***	0.0171
	(0.00916)	(0.0123)	(0.0149)	(0.00389)	(0.00808)	(0.00642)	(0.00608)	(0.0105)	(0.0110)
Observations	116,210	112,168	25,383	76,208	72,616	7,136	51,647	48,711	2,958
OLS	X			X			X		
School Assignment IV		X			X			X	
First-Stage F-statistic		612.6			845.4			686.8	
Sibling Fixed Effects			X			X			X
Number of Families			11640			3344			1388

Robust standard errors adjusted for clustering at the school-level in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$