The Geography of Educational Opportunity: Teacher Quality

Benjamin Feigenberg, Eric Hanushek, Samuel Mosley and Steven Rivkin

September 2024

Preliminary and Incomplete: Please do not Cite or Circulate[1]

I Introduction

Growing economic inequality in recent decades has coincided with a period of declining intergenerational mobility in the United States. In a series of seminal papers, economics researchers have sought to characterize the "geography of economic opportunity" by measuring intergenerational mobility at the commuting zone and county levels (Chetty and Hendren, 2018*a*; Chetty and Hendren, 2018*b*; Chetty et al., 2014). To understand what explains the geographic heterogeneity in intergenerational mobility that they identify, these authors have correlated summary measures of intergenerational mobility with area-level characteristics. Although this exercise identifies school quality (measured based on high school completion rates and test scores) as an important correlate of intergenerational mobility, the key question is whether these correlations reflect a causal effect of school quality or whether other family and neighborhood influences affect both education and economic outcomes and account for the correlations between the two (Gallagher, Kaestner and Persky, 2019). The evidence that measurable school characteristics including expenditures explain little of the variation in school quality, the fact that much of the most compelling evidence on the variation in teacher quality is based on within-school variation in test-score growth, and the growing recognition that school effects on cognitive skills capture only one channel through which schools affect longer-term economic outcomes complicate efforts to measure the contribution of schooling to geographic differences in economic mobility.

In this paper, we extend the approaches used to estimate teacher effectiveness to quantify the contributions of teachers to skill acquisition and social mobility, as measured by longer-term schooling and employment outcomes. Evidence shows that teachers constitute an important input into the education production process. Our empirical two-stage methodology first uses teachers who switch schools to estimate school-average teacher effectiveness at raising cognitive and noncognitive skills relative to other schools in the largest connected network. Comparisons of the effectiveness of teacher movers relative to peer-teachers in their origin and destination schools produce compelling estimates of school-average teacher effectiveness. The approach accounts for all school and neighborhood influences common to all students in a school at a point in time. We then aggregate

[1] The conclusions of this research do not necessarily reflect the opinions or official position of the Texas Education Agency, the Texas Higher Education Coordinating Board, or the State of Texas.

these estimates for each student to create measures of average teacher effectiveness over grades four through eight.

These average effectiveness measures become the key explanatory variables in longer-term outcome regressions. Although the teacher mover design produces cardinal rankings of average teacher effectiveness at raising skills, the possibility that students sort into schools based on factors related to teacher effectiveness has the potential to introduce omitted variables bias. In the present analyses, we address this concern by controlling for observable characteristics including race-ethnicity, gender, eligibility for a subsidized lunch, and prior test scores, absences and receipt of an out-of-school suspension. Looking forward, we will take a series of additional steps to account for unobserved heterogeneity. First, because the measures of prior skills may fail to account fully for unobserved differences related to both teacher effectiveness and longer-term outcomes, we intend to undertake sensitivity analyses that focus on student movers. Specifically, we will adopt an origin-destination school district design that uses the variation generated from those in the same cohort who move between the same two districts in different grades to identify the teacher quality effects on longer-term outcomes. This analysis relies on strong assumptions regarding the orthogonality of move timing with other outcome determinants, and we therefore use an alternative approach that restricts moves to those following school openings and closings. Because such moves typically occur within districts, this analysis compares cohorts exposed to such school moves in different grades to identify the effects of teacher quality on longer-term outcomes. Comparisons of estimates across specifications and extensive placebo testing and sensitivity analysis will provide information on the structure of school effects and influences of confounding factors.

We expect our findings to provide new evidence on variation in school-average teacher quality and the channels through which teachers affect longer-term outcomes and promote educational and economic opportunity. Comparisons across different levels of aggregation including district, commuter zone and time and across demographic groups including race and gender will illuminate the character of teacher quality effects given the existing institutional structures. The analyses included in this paper and the extensions described above represent part of a larger research agenda focused on the contributions of teachers and schools to longer-term educational attainment, labor market engagement, and involvement in the criminal justice system. By characterizing the degree to which geographic variation in teacher effectiveness and school resource effects can ultimately help to explain geographic variation in intergenerational mobility, this work will contribute to research on the determinants of economic and social opportunity and provide valuable information for the development of education and social policies.

II Data

This project links Texas administrative data from four agencies to produce panels that track students through elementary and secondary school, higher education, employment and engagement

with the criminal justice system as an adult. Data on elementary and secondary schooling come from the Texas Education Agency (TEA), data on higher education come from the Texas Higher Education Coordinating Board (THECB), data on earnings come from the Texas Workforce Commission (TWC), and computerized data on criminal histories come from the Texas Department of Public Safety (CCH). The merged data, housed at the University of Texas at Dallas Education Research Center, span almost thirty years (1994 to 2022).

Elementary and secondary school records include information on absences, eligibility for reduced price lunch, disciplinary infractions, out-of-school suspensions, math and reading test scores (available annually for grades 3-8), demographic information, including race, ethnicity, and gender, and program information, including participation in special education. TEA also provides panel data on educators and schools that include information on spending, class size, educator roles, educator characteristics (salary, experience, education), and grade and subject taught for teachers. From 2012 forward teachers can be matched to classrooms, enabling the estimation of teacher value-added measures at the classroom level. This component of the project focuses on these years.

Post-secondary schooling data collected by the THECB each semester include information on enrollment, credit hours, and class in public and private two- and four-year colleges and universities in Texas. Because the administrative data do not include students who attend post-secondary schooling outside of Texas, information on students who attended Texas secondary schools and subsequently attend a college or university outside of Texas included in the National Student Clearinghouse data are appended to the higher education administrative data. This mitigates outmigration induced measurement error for education outcomes.

In contrast to schooling, the workforce and criminal history data do not include information on those who have left Texas. The TWC data include quarterly earnings, employer and industry, while the CCH data include information on arrests that result in conviction as well as sentencing. These records include crime type and severity.

Our initial analyses focus on three key outcomes: (1) enrollment in college, (2) college persistence as measured by persistence to the second year of college, and (3) engagement in productive activities of school or work as measured by being employed or attending college in 6 of the 8 quarters in the two years following expected high school graduation.[2] Sample means for our measures of college enrollment, college persistence and engagement in productive activities are 0.39, 0.30, and 0.42, respectively, indicating that there is substantial variability in each of our key outcome measures. We are currently awaiting additional years of matched CCH data records that will allow us to investigate effects on involvement in the criminal justice system. For example, we can construct

[2] In future analyses that consider schoolwide contributions to longer-term outcomes, we will not rely on data matching teachers to classrooms (that is available only beginning in 2012). By incorporating earlier cohorts, we will be able to analyze a richer set of outcomes related to labor market participation and earnings, given that students from earlier cohorts will have labor market outcomes less affected by contemporaneous educational enrollment by 2022 (the last year for which longer-term labor market outcomes are available).

intensive margin measures related to the number and severity of prior arrests and aggregate time spent incarcerated. Detailed information on specific charging offenses and incarceration spells will also facilitate efforts to calculate the returns (in fiscal terms) to improvements in the cognitive and non-cognitive skills that we study.

As noted, the workforce and criminal justice data to which we have access only contains information on individuals residing in the State of Texas. To learn more about outmigration from Texas, Hanushek et al. (2024) uses 2000 Census data for those between 18 and 20 years old who were Texas middle school students in 1995 and compare schooling, work and institutional living arrangements for those living in Texas in 2000 versus those living in another state and as a function of possession of a high school degree. First, roughly 10 percent of all students and less than 6 percent of students who do not have a high school degree live outside of Texas in 2000. Second, rates of schooling, work and living in an institution are quite similar for those without high school degrees living in Texas and those living in another state; post-secondary school attendance is higher for those living outside of Texas, likely because some students left the state to attend school. All in all, the low rate of outmigration, particularly for those without a high school degree, similar rates of work and living in an institution by state of residence, and availability of National Student Clearinghouse data suggest that any migration-induced attrition bias is likely to be modest in magnitude.

III. Two-stage empirical approach

Random assignment of teachers to schools and students to classrooms and schools would enable the identification of teacher effects on longer-term outcomes with valid measures of teacher effectiveness. Of course, teachers and families actively choose places of residence and schools, and schools often engage in the purposeful assignment of students to classrooms. In the absence of such randomization, the selection on observables assumption that would have to be satisfied to draw a causal interpretation of the effects of teachers on longer-term outcomes would require that any unobserved determinants of the longer-term outcomes would be uncorrelated with teacher assignment. Prior research on schools (see, for instance, Chetty, Friedman and Rockoff, 2014*b*) has argued that unobserved student characteristics such as motivation and family resources are correlated with enrollment patterns. Moreover, any sorting of teachers by unobserved school or neighborhood factors potentially introduces additional sources of bias. By comparison, prior research (Kane and Staiger, 2008; Chetty, Friedman, and Rockoff, 2014a) suggests that the selection on observables assumption is generally satisfied in the estimation of teacher value added differences within schools. However, a focus on within-school variation in teacher value added controls for school and neighborhood differences at the cost of ignoring the contributions of between-school differences in teacher effectiveness to longer-term outcomes. We therefore extend the two-stage framework used by Chetty, Friedman, and Rockoff (2014b) and Jackson (2018) by using teacher movers to account for school and neighborhood heterogeneity in student cognitive and non-cognitive outcome trajectories.

III.a Estimation of school average teacher effectiveness

The purposeful sorting of teachers and students to neighborhoods and schools introduces serious impediments to the measurement of school-average teacher effectiveness at raising cognitive and noncognitive skills. We take advantage of the teacher and student panel data sets to control for unobserved student heterogeneity and differences in schools and neighborhoods that could contaminate the measurement of teacher effectiveness. We focus on four skill proxies. Specifically, we use math and reading scores to measure cognitive skill development, and we use student absent days and an indicator for any out-of-school suspension as proxies for non-cognitive skill development. Notice that for the two proxies for noncognitive skills, a more effective teacher will have a more negative value added.

A concern arises that these proxies use current students to measure achievement and behaviors. Consequently, the VA estimates may conflate school policies such as the suspension threshold with student behaviors or capture strategic teacher or school behavior such as teaching to the test. However, Hanushek et al. (2024) shows that the relationships between longer-term outcomes and VA estimates based on such contemporary proxies are quite similar to the relationships between longer-term outcomes and VA estimates based on proxies measured following student matriculation to the next grade and school.

The baseline model relies on prior test-scores, absences and suspensions to account for student heterogeneity and takes advantage of teacher movers to account for unobserved school and neighborhood differences; future extensions that leverage variation that results from student mobility following school openings and closings as well as student moves will serve as valuable sensitivity checks.

To guide thinking, consider the following linear skill production model in which skill **s** is a function of teacher, school, neighborhood and family effects:

$$(1)\ y^{s}_{ijlnt} = \alpha^{s}_{it} + \gamma^{s}_{jt} + \delta^{s}_{lt} + \theta^{s}_{nt} + \varepsilon^{s}_{ijlnt}$$

Here, skill **s** for student **i** with teacher **j** in school **l** in neighborhood **n** in year **t** is a function of time-varying student (**α**), teacher (**γ**), school (**δ**), and neighborhood (**θ**) effects and a random error (**ε**). The aforementioned sorting of teachers and students along with local variation in school finance introduces non-zero covariances among the student, teacher, school and neighborhood terms. By focusing on pairs of schools connected by a teacher who worked at each during the sample period, we can account for fixed and time-varying school and neighborhood factors and measure average teacher quality at one school relative to the other. Expanding this framework to the full network of schools connected by teacher movers enables us to estimate a cardinal index of school average teacher quality.

We regress the proxy for skill **s** on prior-year values of the four skill proxies (including cubics in prior test scores), demographic characteristics (race-ethnicity and gender fixed effects as well as an indicator for subsidized lunch eligibility), and a vector of teacher fixed effects, separately by year.

The teacher fixed effect coefficients serve as the estimates of teacher effectiveness at raising skill **s** in school **l** in year **t**. These estimates combine the permanent components of teacher effectiveness with the time-varying components related to experience and transitions between schools and grades. They also capture school and neighborhood influences common to all teachers in school **l** and neighborhood **n** in year **t**. Following the prior literature (Kane and Staiger, 2008; Chetty, Friedman, and Rockoff, 2014a), we maintain the assumption that the included controls capture systematic factors related to classroom assignment within schools.

Equation (2) presents the difference between the average estimated teacher effect on skill **s** in school **l** in year **t,** where J_{lt} represents the number of teachers in school **l** in year **t**, and the estimated "permanent" effectiveness of moving teacher **m** on skill **s**, also estimated in year **t**:

$$(2)\ \frac{1}{J_{lt}}\sum_{j=1}^{J_{lt}}\hat{\gamma}^{s}_{j,lt} - \widehat{\gamma'}^{s}_{mt} =$$

$$(3)\ \left(\frac{1}{I_{lt}}\sum_{i=1}^{I_{lt}}\alpha^{s}_{i,lt} + \overline{\gamma^{s}_{lt}} + \delta^{s}_{lt} + \theta^{s}_{nt} + \frac{1}{I_{lt}}\sum_{i=1}^{I_{lt}}\widehat{\varepsilon^{s}_{ijlnt}}\right) - \left(\frac{1}{I_{mlt}}\sum_{i=1}^{I_{mlt}}\alpha^{s}_{i,lt} + \gamma'^{s}_{m} + \delta^{s}_{lt} + \theta^{s}_{nt} + \frac{1}{I_{mlt}}\sum_{i=1}^{I_{mlt}}\widehat{\varepsilon^{s}_{ijlnt}}\right) =$$

$$(4)\ \left(\overline{\gamma^{s}_{lt}} - \gamma'^{s}_{m}\right) + \left(\frac{1}{I_{lt}}\sum_{i=1}^{I_{lt}}\alpha^{s}_{i,lt} - \frac{1}{I_{mlt}}\sum_{i=1}^{I_{mlt}}\alpha^{s}_{i,lt}\right)$$

As Equations 3 and 4 show, differencing out the estimated effectiveness of teacher **m** accomplishes two objectives. First, since the estimated teacher effects are based on data for year **t**, differencing removes the school ($\boldsymbol{\delta}$), neighborhood ($\boldsymbol{\theta}$) and school average student ($\boldsymbol{\alpha}$) influences that are common to all teachers in school **l** in year **t**. Note that the difference between the school average student effect and teacher **m** average student effect (terms in second set of parentheses in Equation 4) remains but is zero on average in the absence of classroom sorting on unobservables. Second, it enables a meaningful comparison of average teacher effectiveness in school **l** and school **k**, the other school in which teacher **m** works, by comparing the average effectiveness of teachers in each school with the effectiveness of the same teacher **m** during the year in which the mover works in that school.

Using the moving teacher to benchmark school average teacher effectiveness requires that we isolate the permanent components of teacher effectiveness (designated by $\boldsymbol{\gamma'}$) that do not differ systematically across years and schools. To do so, we run auxiliary regressions of the estimated school-by-year-level teacher effects on teacher fixed effects as well as a set of single-year experience dummies for values 1 to 3, an indicator for a school transition prior to that year, and an indicator for teaching in a different grade than the prior year. We then construct our measure of permanent teacher effectiveness as predicted teacher-by-school year value added, net of experience, school transition and grade transition effects. In the next iteration we will also account for the presence of a student classified as emotionally disturbed and class size to even more comprehensively capture differences in classroom settings.

Many teachers work in multiple schools during the sample period (which begins in 2012). To obtain estimates of school average teacher effectiveness for the largest connected network, we regress the dependent variable shown in equation (2) on vectors of school and teacher fixed effects for the full sample of teachers who work in multiple schools. The teacher fixed effects restrict comparisons to schools with at least one teacher in common, and the coefficients on the school fixed effects represent the index of school average teacher effectiveness. Note that the largest network of schools connected by a teacher mover includes 89% of all schools in the sample. Because the school fixed effect estimates include sampling error, we use empirical Bayes methods to shrink the estimates toward their mean.

III.b Estimation of teacher effects on longer-term outcomes

In the second stage we regress longer-term outcomes on indexes of teacher effectiveness at improving each of the four skill proxies averaged over grades four through eight. The baseline specifications also include the same demographic controls as the first-stage skill regressions, as well as grade three test scores, absences and an indicator for any suspension.[3] These models rely on the selection on observables assumption, meaning that the controls for unobserved student heterogeneity including the prior skill measures account for any purposeful sorting of students by average teacher quality.

Because of the possibility that assumption could be violated, we investigate the sensitivity of the estimates to alternative specifications. In the currently included analyses, we re-aggregate the average teacher effectiveness measures using the initial elementary school a student attends and the middle school(s) they are predicted to subsequently attend, rather than using the actual school attended for all grades. This produces intent-to-treat estimates not affected by endogenous school changes based on realized teacher quality.

Moving forward, as an alternative robustness check, we will limit the second-stage sample to students who attend schools where school openings or closings affect attendance patterns. By controlling for initial school attended in the second stage regressions, we will restrict comparisons to differences in teacher effectiveness related to the timing of the closing or opening. Cohorts affected by openings or closings in different grades will vary in their exposure to the closing or opening school.

A final issue concerns the selection of teachers used in the absence and out-of-school suspension regressions. At present, we leverage only the contributions of math teachers to student absences and suspensions when producing our cardinal measures of school-level average teacher effectiveness at raising non-cognitive skills. While math teachers may plausibly drive year-on-year changes in student math test scores, all a student's teachers likely affect the development of behavioral skills. Consequently, the current measures of school-average teacher effectiveness at

[3] As detailed in Section V, current estimates are derived from a sample including only students expected to graduate high school in Spring 2019. Future models that incorporate students from other cohorts and additional longer-run outcome measures will add indicators for cohort and year.

reducing absences and out-of-school suspensions are measured with error. Despite this measurement challenge, we find evidence that average teacher effectiveness along these domains, particularly that captured by out-of-school suspensions, is a robust driver of longer-run outcomes.

IV. Variation in teacher effectiveness

The distributions of school-average teacher effectiveness at raising cognitive and noncognitive skills illuminate the potential contribution of teachers to the variation in longer-term outcomes including differences across geographies. We apply Bayesian shrinkage methods to dampen the influence of measurement error and present shrunk estimates of mean teacher effectiveness. In future analyses, we will decompose the variance in teacher effectiveness to understand the extent of variation within versus across alternative geographies, including commuting zones and school districts. We will also investigate correlations between average teacher effectiveness measures and both school-level average student characteristics and neighborhood attributes.

In Table 1, we report estimates of the distributions of both school-average teacher effectiveness and average teacher effectiveness experienced by students from grades 4-8, based on elementary and middle school averages of teacher effectiveness. The latter is roughly two-thirds of the former and illustrative of the variation in the quality of instruction across elementary- and middle-school grades 4-8. Importantly, these estimates do not incorporate early grade differences in school quality because of the need to account for unobserved heterogeneity with baseline skill measures. Nevertheless, the fact that most elementary schools include grades k-3 and 4-5 suggests that differences in the upper elementary grades provide information on the likely variation in the early grades.

a. Distribution of school average teacher effects

Table 1 reports estimates of standard deviations of school average teacher effectiveness and the averages for students over grades 4 to 8 for the 89 percent of schools in the largest connected network. To highlight the impacts of averaging to the school level, controlling for school and neighborhood effects and random error, Table 1 also reports estimates of teacher VA, school average teacher VA prior to removing school and neighborhood effects or shrinking, and school average teacher VA prior to shrinking.

The estimates in Columns 4 and 5 reveal modest but educationally meaningful variation in teacher effects that decline by around one third when the effects are averaged over grades 4-8 for each student. Given that we hold school average teacher effects constant across the period and the majority of students attend only one elementary school above grade 3 and one middle school between grades 6 (or 7) and 8, the decreases between Columns 4 and 5 are consistent with low correlations between elementary and middle school effects. The Column 4 magnitudes of 0.07 and 0.05 standard deviations in math and reading, respectively, are in the range of 50 percent as large as typical

estimates of the within-school variation in teacher effectiveness, and the standard deviations of the student averages are even smaller.

Moving from Column 1 to Column 5 shows the effects of aggregation, removal of school and neighborhood influences, shrinkage, and aggregation across all grades attended on the variations in VA. Estimates of VA variability decline by roughly 85 percent, on average, with marginally greater declines for estimates of VA to proxies for noncognitive skills. Aggregation from the teacher to the school average (moving from Column 1 to Column 2) cuts the VA to math and reading achievement roughly in half, while it has little effect on VA to absences or suspensions. In contrast, removal of the school and neighborhood influences increases the variation in the math and reading VA estimates but substantially decreases the variation of the VA to noncognitive skill proxies. This suggests that much of the variation in absences and suspensions occurs across schools, while there is much more variation in achievement VA within schools. Shrinkage reduces the variation of all VA measures by more than 50 percent, while averaging within students and across grades decreases the variation by roughly an additional one-third.

Figures 1-4 present kernel density plots of the unshrunk and shrunk school average teacher effects, as well as plots of average teacher effectiveness experienced by students from grades 4-8. These exhibits reveal substantial variation for all four VA effects. Table 2 shows that the interquartile ranges equal roughly one standard deviation, and the differences between the 10th and 90th percentiles tend to hover around 2 standard deviations. These differences introduce the possibility that teacher quality contributes meaningfully to post-secondary schooling and productivity differences. We turn now to the results of the estimation of longer-term outcomes on the value-added effects, and their magnitudes determine the impacts of the variation in school average teacher effectiveness on schooling and productivity outcomes.

V. Effects of teacher effectiveness on longer-term outcomes

We next present estimates from regressions of longer-term outcomes on the measures of school-average teacher VA to each of the four skill proxies, aggregated across grades 4 to 8 for each student. In Table 3, we restrict the sample to students in grade 3 in 2009-2010, who have grade 3 information on achievement, absences and suspensions, who attend a school in the largest connected network in grades 4 through 8 and who progress one grade per year. These restrictions reduce the sample size by roughly 45% of the total number of students in the 2009-2010 3rd grade cohort. Given that post-secondary enrollment data are currently only available to us through the 2020-2021 academic year, this cohort of students, who are expected to graduate from high school in spring 2019, is the youngest cohort for whom we can construct the three longer-term outcomes of interest (which require two years of data after the expected high school graduation year). The absence of student-teacher matches prior to 2012 precludes the inclusion of earlier cohorts.

Across each of the three longer-term outcomes analyzed, we arrive at qualitatively similar conclusions regarding the impacts of teacher effectiveness with respect to test scores and non-

cognitive skills. The coefficients on math VA and suspension VA are significant at the 1 percent level with the exception of the effect of suspension VA on the probability of engaging in a productive activity that is significant at the 5 percent level. By comparison, the coefficient on absence VA is significant at the 1 percent level in the college attendance specification, and the coefficient on reading VA is significant at the 10 percent level in the college persistence specification. Although the other reading and absence VA coefficients have the expected signs, they are not significant at any conventional level.

To facilitate interpretation, we next multiply estimated treatment effects from Table 3 by the standard deviation of the relevant measure of teacher effectiveness (presented in Column 4 of Table 1). This product identifies the estimated effect of a one standard deviation increase in school-average teacher effectiveness on the longer-term outcome of interest. In the case of math VA, given that the standard deviation of average teacher effectiveness in raising math test scores is 0.071, a 1 SD increase in average teacher effectiveness on this margin translates into a 0.9 percentage point increase in college attendance (a 2.4% increase relative to a mean attendance rate of 38.7%). Turning to the impacts of average teacher effectiveness at raising math scores on college persistence and on our measure of productive engagement in school or work, we find that a 1 SD increase in average effectiveness is associated with increases of 1.3 and 0.7 percentage points, respectively, in these two longer-term outcomes (corresponding to percentagewise effects of 4.3% and 1.7% given baseline rates of college persistence and productive engagement of 30.1% and 41.8%). Not only is the effect of school average reading value added insignificant for two of the specifications, it is smaller in the college persistence specification in which it is marginally significant: a 1 SD increase in average teacher effectiveness in raising reading scores leads to a 0.4 percentage point or a roughly 1.3 percent increase in the probability of college persistence.

Similar magnitudes emerge for the non-cognitive skill proxies. As above, in Table 4, we re-scale estimated treatment effects from Table 3 by the standard deviation of the relevant measure of teacher effectiveness. Focusing first on average effectiveness at reducing student absences, point estimates imply that a 1 SD increase in teacher effectiveness leads to increases in college attendance, college persistence, and productive engagement of 1.0, 0.3 and 0.4 percentage points (alternatively, 2.6%, 1.0%, and 1.0% effects in percentage terms). Only the estimated impact of teacher effectiveness at reducing absences on college attendance is significant at standard confidence levels. Finally, we find that increased average teacher effectiveness at reducing suspensions has a strong positive impact on longer-run outcomes, with effect sizes comparable to those associated with variation in average teacher effectiveness at raising math test scores. Specifically, a 1 SD reduction in suspensions value added (corresponding to a decline in suspensions) leads to increases in college attendance, college persistence, and productive engagement of 1.1, 0.9 and 0.7 percentage points (alternatively, 2.8%, 3.0%, and 1.7% effects in percentage terms).

The estimates in Table 3 come from a sample of students who follow standard grade progression patterns and attend a school in the largest connected network in grades 4 through 8. To

investigate the importance of endogenous student exit decisions out of a connected network school, we predict the average teacher effectiveness to which students would be exposed based on their third-grade school. First, we assume that students remain enrolled in their original elementary school through the maximum grade level offered by that school. Second, we assign students to a middle school based on the patterns of matriculation to middle school for each elementary school. Specifically, we identify the two middle schools in which graduates from that elementary school are most likely to enroll. Then for each grade level after predicted elementary school graduation, we construct a measure of predicted average teacher effectiveness that is a weighted average of the effectiveness of these two destination middle schools, with weights defined by the shares of elementary school graduates who enrolled in each middle school (we rescale the weights to sum to one). We restrict the sample to include only students for whom average teacher effectiveness measures are available for the origin elementary school and the two destination middle schools. While this restriction reduces the sample size to some degree, regression specifications that relate longer-run outcomes to predicted average teacher effectiveness nonetheless include 20%-30% more students than those specifications in Table 3 which leverage actual student exposure.

Table 5 shows that the basic patterns for these intent-to-treat estimates are quite similar to those presented in Table 3 based on actual schools attended for the restricted sample of students who attend a school in the largest connected network in all grades. Not surprisingly, almost all of the intent-to-treat estimates are smaller in magnitude than the corresponding estimates shown in Table 3, and the differences tend to be largest in the college attendance specifications. These findings suggest that endogenous selection out of the largest connected network of schools does not introduce substantial bias.

VI. Discussion and conclusion

This paper provides a roadmap for constructing cross-school comparisons of average teacher effectiveness that difference out the contributions of school, neighborhood, and student factors to the short-run cognitive and noncognitive measures of interest. We demonstrate the practical applicability of this approach, which allows for comparisons to be constructed across the vast majority of public elementary and middle schools operating in the State of Texas. Initial estimates, which do not yet include measures of criminal justice system engagement in early adulthood and are based on only a single cohort of students, further demonstrate that the measures of average teacher effectiveness we construct are meaningful and drive important differences in students' future educational and labor market engagement. Moreover, the findings in Hanushek et al. (2024) suggest that the VA to non-cognitive skills is likely to be an even more important determinant of involvement in the criminal justice system. Next steps include understanding the extent to which differences in average teacher effectiveness have the potential to contribute meaningfully to measured differences

across geography in economic opportunity, and to what degree aggregate measures of geography-level intergenerational mobility may miss important within-geography variability on the basis of average teacher effectiveness.

Moving forward, we will also investigate the stability of average teacher effectiveness measures across cohorts, and the bias-coverage tradeoff associated with alternative approaches that rely on subsamples of years to construct teacher move networks and school-level measures of average teacher effectiveness. To provide additional support for a causal interpretation of the estimated association between average teacher effectiveness and longer-run outcomes, we are now working to assemble data on school openings and closings that will allow us to exploit variation in student-level predicted average teacher effectiveness in the absence of student relocations that would make it difficult to disentangle the contributions of teacher effectiveness to longer-run outcomes from the contributions of other unobserved neighborhood factors. Although our analyses have focused on average teacher effectiveness in elementary and middle schools due to endogeneity in the specific tests taken by high school students (which vary based on course enrollment), we also intend to leverage high school records to investigate the contributions of average elementary/middle school teacher effectiveness on intermediate outcomes, including absences and suspensions in high school as well as measures of course selectivity and grade progression.

References

Austin, Wes, David Figlio, Dan Goldhaber, Eric Hanushek, Tara Kilbride, Cory Koedel, Jaeseok Lee, Jin Lou, Umut Ozek, Eric Parsons, Steven Rivkin, Tim Sass, and Katharine Strunk. 2022. "Academic Mobility in U.S. Public Schools: Evidence from Nearly 3 Million Students." *Journal of Public Economics,* forthcoming.

Carrell, Scott, and James West. 2010. "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors." *Journal of Political Economy*, 118(3): 409–432.

Chetty, Raj, and Nathaniel Hendren. 2018*a*. "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects." *Quarterly Journal of Economics*, 133(3).

Chetty, Raj, and Nathaniel Hendren. 2018*b*. "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." *Quarterly Journal of Economics*, 133(3).

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

Chetty, Raj, John Friedman, and Jonah Rockoff. 2014*b*. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review*, 104(9): 2633–2679.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics*, 129(4): 1553–1623.

Dinerstein, Michael, and Isaac Opper. 2022. "Screening with Multitasking: Theory and Empirical Evidence from Teacher Tenure Reform." NBER Working Paper No.30310.

Feigenberg, Benjamin, and Conrad Miller. 2021. "Racial Divisions and Criminal Justice: Evidence from Southern State Courts." *American Economic Journal: Economic Policy*, 13(2).

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of Geographic Variation in Health Care: Evidence from Patient Migration." *Quarterly Journal of Economics*, 141(4): 1681–1726.

Gallagher, Ryan, Robert Kaestner, and Joseph Persky. 2019. "The Geography of Family Differences and Intergenerational Mobility." *Journal of Economic Geography*, 19: 589–618.

Gilraine, Michael, and Nolan Pope. 2021. "Making Teaching Last: Long-Run Value-Added." NBER Working Paper No.29555.

Hanushek, Eric A., Andrew Morgan, Steven G. Rivkin, Jeffrey C. Schiman, Ayman Shakeel, and Lauren Sartain. 2024. "The Lasting Impacts of Middle School Principals." NBER Working Paper No.32642.

Jackson, C. Kirabo. 2018. "What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes." *Journal of Political Economy*, 126(5): 2072–2107.

Kane, Thomas J., and Douglas O. Staiger. 2008. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." NBER Working Paper No. 14607.

Rivkin, Steven, Eric Hanushek, and John Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica*, 73(2): 417–458.

Table 1. Estimated Standard Deviations of Teacher Effectiveness at Raising Cognitive and Non-Cognitive Skills Relative to Other Schools in the Largest Connected Network, by Aggregation Level and Controls for School and Neighborhood Influences

Aggregation level	Teacher	School	School	School	Student average over grades 4-8
	(1)	(2)	(3)	(4)	(5)
Controls for school and neighborhood influences	No	No	Yes	Yes	Yes
Shrunk	No	No	No	Yes	Yes
Math VA	0.283	0.134	0.183	0.071	0.045
Reading VA	0.222	0.119	0.140	0.052	0.032
Absence VA	2.447	2.376	1.332	0.439	0.296
Suspension VA	0.093	0.080	0.051	0.018	0.012
Observations	98,450	6,132	6,131	6,131	232,802

Notes: VA estimates in each column include only teachers/schools observed in the largest connected network.

Table 2. Percentile Distributions of School Average Teacher Effectiveness at Raising Cognitive and Non-Cognitive Skills Relative to Other Schools in the Largest Connected Network

Percentile	10th (1)	25th (2)	50th (3)	75th (4)	90th (5)	Mean (6)
Math VA	0.138	0.180	0.223	0.268	0.310	0.223
Reading VA	-0.187	-0.156	-0.126	-0.096	-0.067	-0.127
Absence VA	-1.607	-1.378	-1.147	-0.916	-0.698	-1.143
Suspension VA	-0.021	-0.012	-0.005	0.003	0.013	-0.004

Notes: VA estimates are shrunk using Bayesian methods and are constructed relative to other schools in the largest connected network. Estimates are adjusted to account for school and neighborhood influences.

Table 3. Estimated Effects of School Average Teacher Effectiveness at Raising Cognitive and Non-Cognitive Skills Aggregated over Grades 4 to 8 on College Attendance, College Persistence, and the Probability of Engaging in the Productive Activities of School or Work after High School

	College attendance (1)	College persistence (2)	Engagement in a productive activity (3)
Math VA	0.131	0.183	0.101
	(0.042)	(0.035)	(0.038)
Reading VA	0.047	0.086	0.041
	(0.060)	(0.052)	(0.056)
Absence VA	-0.022	-0.006	-0.008
	(0.007)	(0.006)	(0.006)
Suspension VA	-0.624	-0.488	-0.379
	(0.162)	(0.152)	(0.152)
Observations	199,147	196,149	196,149

Notes: VA estimates are shrunk using Bayesian methods and are constructed relative to other schools in the largest connected network. Estimates are adjusted to account for school and neighborhood influences. College persistence is measured by persistence to the second year of college. Students are considered to be engaging in a productive activity if they are attending school or working in 6 out of the 8 quarters during the two-year period following the expected date of high school graduation based on the date of 8th grade completion. The sample includes students expected to graduate high school in Spring 2019 and is restricted to students who progressed one grade level each year between grades 4 and 8 and who have available average teacher effectiveness measures in each year.

Table 4. Simulated Effects of One Standard Deviation Increases in School Average Teacher Effectiveness at Raising Cognitive and Non-cognitive Skills on College Attendance, College Persistence, and the Probability of Engaging in the Productive Activities of School or Work after High School

	College attendance	College persistence	Engagement in a productive activity
	(1)	(2)	(3)
Math VA	0.009	0.013	0.007
Reading VA	0.002	0.004	0.002
Absence VA	-0.010	-0.003	-0.004
Suspension VA	-0.011	-0.009	-0.007

Notes: Each estimate is constructed as (1) the estimated treatment effect from Table 3 multiplied by (2) the standard deviation of the relevant measure of teacher effectiveness (presented in Column 4 of Table 1). This product identifies the estimated effect of a one standard deviation increase in school average teacher effectiveness on the longer-term outcome of interest. College persistence is measured by persistence to the second year of college. Students are considered to be engaging in a productive activity if they are attending school or working in 6 out of the 8 quarters during the two-year period following the expected date of high school graduation based on the date of 8th grade completion.

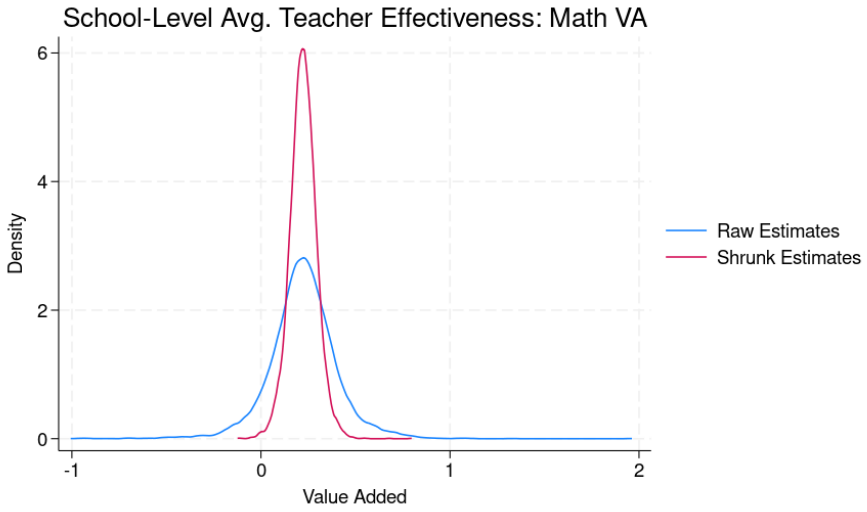
Table 5. Estimated Intent-to-Treat Effects of School Average Teacher Effectiveness at Raising Cognitive and Non-Cognitive Skills Aggregated over Grades 4 to 8 on College Attendance, College Persistence, and the Probability of Engaging in the Productive Activities of School or Work after High School

	College attendance (1)	College persistence (2)	Engagement in a productive activity (3)
Predicted Math VA	0.117	0.182	0.122
	(0.036)	(0.032)	(0.035)
Predicted Reading VA	0.009	0.041	0.014
	(0.046)	(0.042)	(0.045)
Predicted Absence VA	-0.020	-0.01	-0.01
	(0.005)	(0.005)	(0.005)
Predicted Suspension VA	-0.369	-0.403	-0.324
	(0.127)	(0.121)	(0.124)
Observations	262,705	237,261	237,261

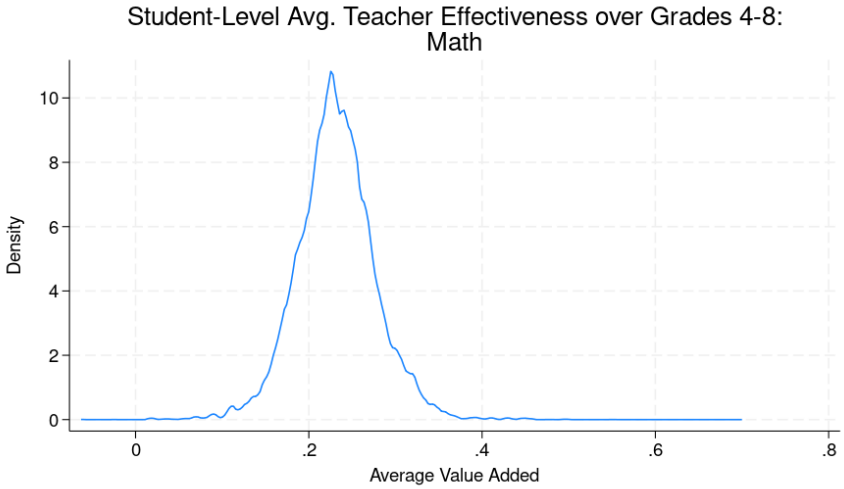
Notes: School-level VA estimates are shrunk using Bayesian methods and are constructed relative to other schools in the largest connected network. Estimates are adjusted to account for school and neighborhood influences. Predicted VA measures are then constructed by assuming that students remain enrolled in their original elementary school through the maximum grade level offered by that school and by assigning students to a weighted average of the effectiveness levels of potential destination middle schools, with weights constructed based on peer enrollment patterns. College persistence is measured by persistence to the second year of college. Students are considered to be engaging in a productive activity if they are attending school or working in 6 out of the 8 quarters during the two-year period following the expected date of high school graduation based on the date of 8th grade completion. The sample includes students expected to graduate high school in Spring 2019.

Figure 1. Kernel Density Plots: Math Value Added

Panel (a)



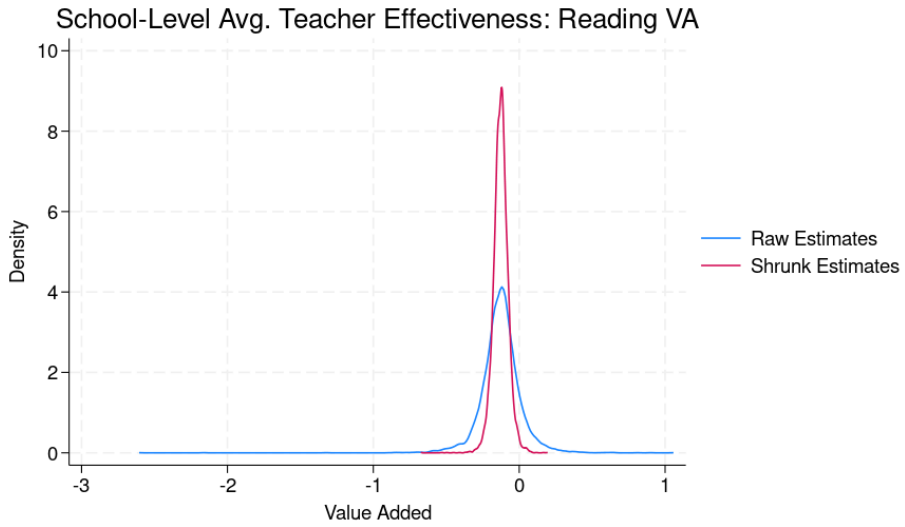
Panel (b)



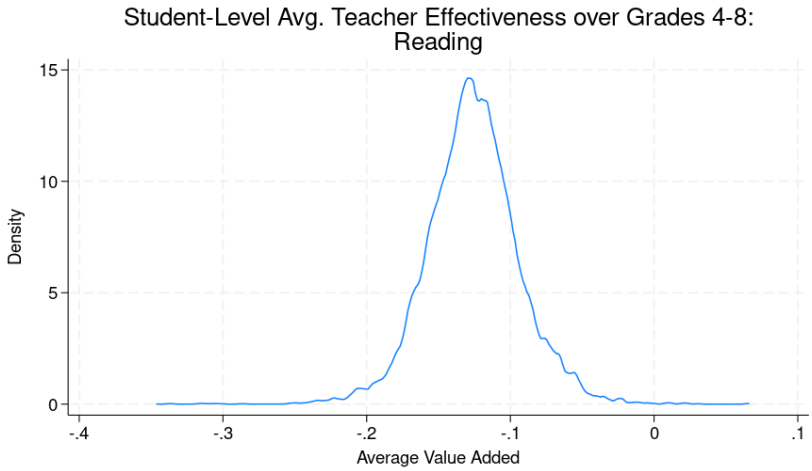
Notes: School-level VA estimates in Panel (a) are constructed relative to other schools in the largest connected network. Estimates are adjusted to account for school and neighborhood influences. We present raw/unshrunk estimates and estimates that are shrunk using Bayesian methods. In Panel (b), we present the distribution of student-level averages over grades 4-8 of the shrunk school-level estimates.

Figure 2. Kernel Density Plots: Reading Value Added

Panel (a)



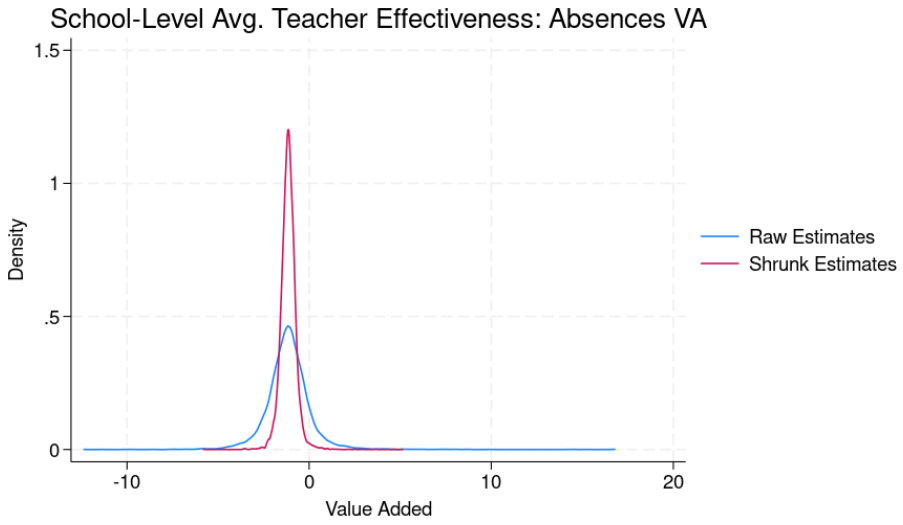
Panel (b)



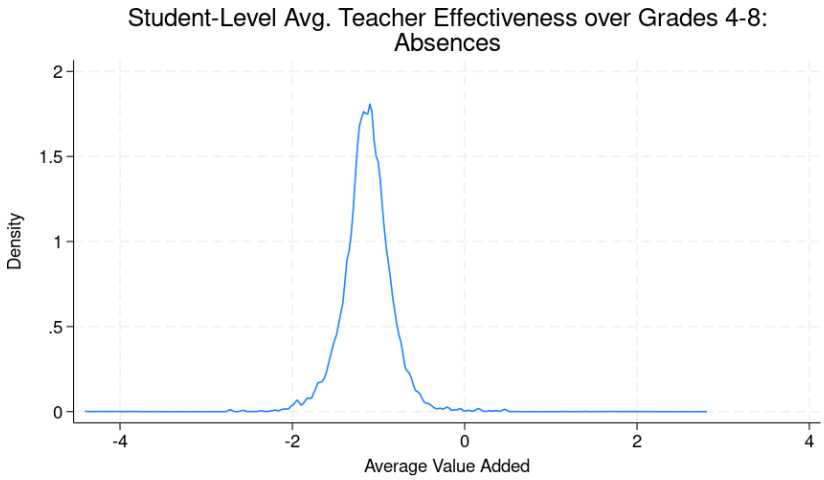
Notes: School-level VA estimates in Panel (a) are constructed relative to other schools in the largest connected network. Estimates are adjusted to account for school and neighborhood influences. We present raw/unshrunk estimates and estimates that are shrunk using Bayesian methods. In Panel (b), we present the distribution of student-level averages over grades 4-8 of the shrunk school-level estimates.

Figure 3. Kernel Density Plots: Absences Value Added

Panel (a)



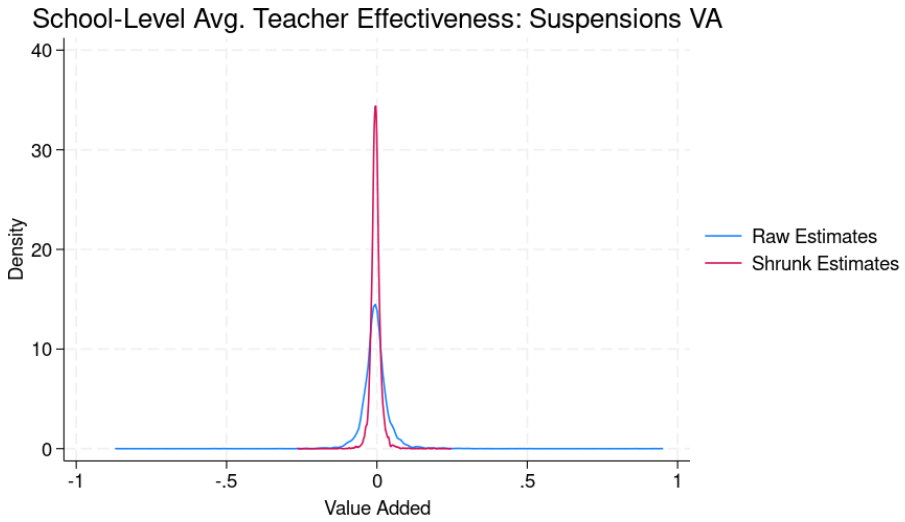
Panel (b)



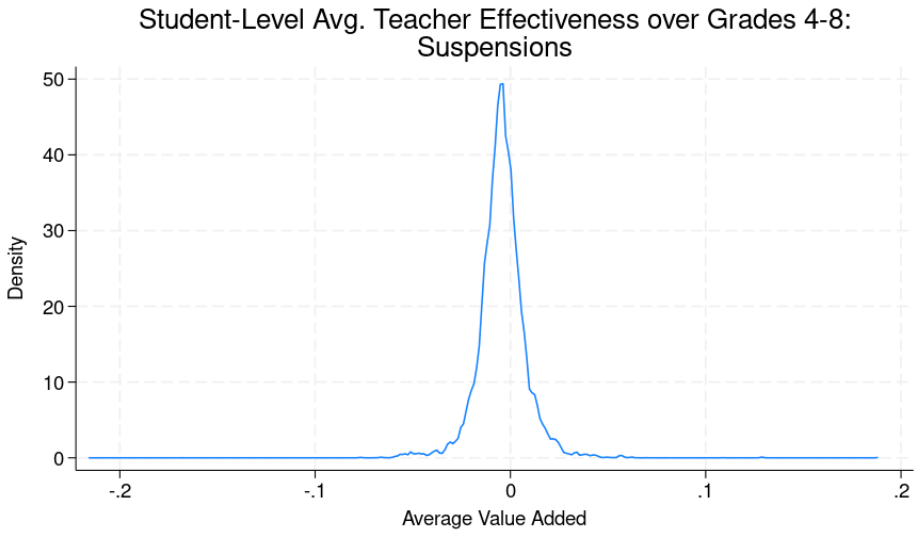
Notes: School-level VA estimates in Panel (a) are constructed relative to other schools in the largest connected network. Estimates are adjusted to account for school and neighborhood influences. We present raw/unshrunk estimates and estimates that are shrunk using Bayesian methods. In Panel (b), we present the distribution of student-level averages over grades 4-8 of the shrunk school-level estimates.

Figure 4. Kernel Density Plots: Suspension Value Added

Panel (a)



Panel (b)



Notes: School-level VA estimates in Panel (a) are constructed relative to other schools in the largest connected network. Estimates are adjusted to account for school and neighborhood influences. We present raw/unshrunk estimates and estimates that are shrunk using Bayesian methods. In Panel (b), we present the distribution of student-level averages over grades 4-8 of the shrunk school-level estimates.