Elementary and Middle School and Principal Effects on Future Academic, Behavioral and Labor-Market Outcomes

by Eric A. Hanushek, Andrew Morgan, Steven G. Rivkin, Lauren Sartain, and Jeffrey C. Schiman[1]

December 2018

VERY PRELIMINARY. PLEASE DO NOT CITE

ABSTRACT

Evidence that teacher effects on both cognitive and non-cognitive skills contribute to longer-term academic, social and labor-market outcomes highlights potential limitations of a singular focus on achievement and the importance of measuring and identifying educator effects on the development of a range of skills. This likely holds even more for school leaders than for teachers, and in this study, we use administrative longitudinal data from the Chicago Public Schools (CPS) and the state of Texas to identify value-added of elementary and middle schools during the tenure of a single principal to cognitive and non-cognitive skills and longer-term outcomes. The panel data sets that span roughly twenty years each enable us to account for observed and unobserved influences by focusing on comparisons between students who attended the same school under the leadership of different principals and students from different elementary or middle schools who attend the same high school at the same time. The preliminary estimates reveal substantial variation in elementary and middle school effects on current achievement, achievement, absences and disciplinary infractions during high school, and on college attendance and persistence and the probability of being in college or employed following high school graduation. In the case of Texas middle schools, it appears that effects are much stronger for lower-achievement students. We also provide additional evidence on the character of principal principal contributions to these differences. Initial findings that are quite preliminary reveal similar differences in schools with a single long-serving principal as in schools where the cohorts attend school under different principals. This is not consistent with constant school quality under the leadership of a single principal. Rather they suggest long-serving principals can make major changes, though they are also consistent with little effect of principals. Regardless, the strong correlations effects on 9th grade absences and effect on the probability of being in school or employed following high school highlights the importance of considering and measuring school effects on both cognitive and non-cognitive skills.

---

[1] Haunshek: Stanford University, National Bureau of Economic Research, and University of Texas at Dallas; Morgan: University of Illinois at Chicago: Rivkin: University of Illinois at Chicago, National Bureau of Economic Research and University of Texas at Dallas; Sartain, University of Chicago; and Schiman: Georgia Southern University. This research has been supported by the Laura and John Arnold Foundation, the Smith Richardson Foundation, and CALDER. 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, the State of Texas, or the Chicago Public Schools.

I. Introduction

Although recent school accountability efforts including No Child Left Behind (NCLB) or Race to the Top (RTT) focus on shorter-term test score gains because of a desire to measure learning outcomes annually for elementary and secondary school grades, it is the impact of schooling on longer-term academic, behavioral and labor market outcomes that determine its value to individuals and external benefits to society. Research on teachers shows substantial effects on student cognitive and non-cognitive outcomes that may persist over time. For instance, the seminal work by Card and Krueger (1992) studying school resource effects, Carneiro and Heckman (2003) studying early schooling interventions, and Chetty et al. (2014) studying early teacher quality support the notion that improvements in early school quality may improve longer-term student outcomes.

Findings of a strong relationship between value added to test scores and future earnings lessens concerns that annual achievement measures used in accountability systems distort incentives away from practices and policies that would maximize longer-term outcomes. However, many have raised concerns that the introduction of high-stakes accountability has weakened the association between effects on test scores and those on the longer-term outcomes. For instance, Jackson (2018) finds evidence that a singular focus on test scores not only ignores important contributions to the development of non-cognitive skills, but also introduces substantial divergence between the ordering of teachers on the basis of test score value added and the ordering of teachers on the basis of their effects on longer-term academic attainment and potential earnings.

Increased emphasis on the quality of school leadership has accompanied the expansion of school accountability, and a nascent but growing body of research finds significant differences in principal productivity as measured by effects on achievement, teacher survey responses, and supervisor ratings (Brewer, 1979; Eberts and Stone, 1988; Dhuey and Smith, 2014; Branch, et al., 2018). This work largely parallels research on teacher value added, though the presence of a single school leader at each point in time and more complex dynamics introduced by the persistence of principal effects on school policies and the future stock of teachers complicate efforts to isolate the contribution of the principal. In that vein, Grissom and Loeb (forthcoming), raise questions about the measurement of principal effects. Recent work by Branch et al (2018) adopts a conservative estimation approach and finds strong evidence of significant variation in principal productivity that rises with the level of student poverty. This is consistent with Bloom et al (2015) that finds management practices in education significantly affect outcomes

In this study, we use administrative longitudinal data from the Chicago Public Schools (CPS) and Texas to estimate the relationship between elementary and middle-school principal value added to achievement and academic, behavioral, and labor market outcomes. The panel data from the Chicago Public School district (CPS) and Texas public schools enable us to account for observed and unobserved influences that may confound estimates of principal effects. Both longitudinal datasets span many years, allowing us to account for elementary or middle-school fixed effects while focusing on comparisons between two principals who led the same school at different times. The fact that students from different elementary and middle schools feed in to the same high school enables the inclusion of high school or even high school by year fixed effects to account for the direct effects of high school quality and disciplinary policies and unobserved factors related to high school choice. The data sets complement one another: the different settings including elementary grade structure provides evidence on the generalizability of the findings; data advantages in one data set can be used to fill in the gaps of the other; and the availability of different outcome measures enable us to expand outcomes under study.

Importantly, focusing on students who have only a single principal between grades 4 and 8 in the CPS, where schools span grades K to 8, or during middle school in Texas circumvents the need to specify the dynamics of skill acquisition during these years. The need to control for unobserved heterogeneity leads to the focus on the higher elementary grades in CPS, and the estimates therefore understate the variation in elementary-school principal effectiveness.

The elementary and high school by year fixed effect estimates reveal substantial variation in principal effects on current achievement, achievement, absences and disciplinary infractions during high school, and on college attendance, persistence and the probability of being in college or employed following high school graduation. There are similarities but also differences in the patterns observed for CPS and Texas. Perhaps most striking are the much stronger and robust results for initially lower-achievement students in Texas middle schools where principal effects on both cognitive and non-cognitive skills are strongly related to post-secondary schooling and employment.

These results provide compelling evidence of within school differences in school effects associated with different principals, but the attribution to the principals as the sources of these effects remains questionable. Other factors not under the control of principals including aspects of changes in teacher composition and district curricular reforms contribute to the variation in principal fixed effects, and these cannot be measured directly. However, patterns of outcome variation within and across principal spells can illuminate the channels through which principals may influence school quality.

We therefore divide long principal spells into two periods and compare within-school differences in value-added for different cohorts in schools with a single principal throughout with differences for cohorts in schools with principal turnover. One might expect larger differences in schools with turnover if principals exert substantial influence. However, if it takes time for principals to make meaningful changes there might be substantial variation within the spell of a single principal. Therefore, we will describe differences in teacher turnover between schools that experience a principal transition and those that do not and compare turnover with achievement change.

The next section describes the Texas and CPS administrative data. Section 3 presents the model used to identify principal effects, and Section 4 reports the estimates of the variances in principal effects and their associations with one another. Because baseline achievement differences likely influence future goals and challenges during school, we examine heterogeneity by baseline achievement scores. Finally, Section 5 summarizes the analysis and considers implications for the measurement of principal productivity.

II. CPS and Texas data

The panel data sets for CPS and Texas underlie the analysis of principal effects in each of the jurisdictions. Data for each span over twenty years and follow children as they progress through elementary, secondary, and post-secondary schooling as well as into the workforce in Texas. We now describe the data sources used to construct the CPS and Texas samples.

II.a. CPS data

We use administrative data from the University of Chicago Consortium on School Research. The CPS data span the period 1993-4 to 2013-4 and contain extensive information on educators and students. The student data include math and reading test scores in grades 3 through 10, math and reading GPA, attendance, disciplinary infractions, demographic characteristics, special education status, eligibility for a subsidized or free lunch, school attended, grade, and school characteristics. To characterize early principal effectiveness at the elementary school level, we use results from the Iowa Test of Basic Skills (offered 1994 to 2005) and the Illinois Standards Achievement Test (offered 2006 to 2014). The ITBS and ISAT are standardized achievement tests used to measure school performance and determine accountability ratings. At the high school level, students were administered the ACT Plan test at the beginning of their 10th grade year from 1993-4 to 2010-1, and the ACT Explore test in the spring of their 9th grade year from 2011-2012 to 2013-2014.

Data on CPS staff contain information about the current position, allowing us to observe the principal of record in each year. To create linkages over time, we used principal name as the primary variable, and took extensive efforts to account for name changes resulting from changes in naming conventions or marriage. The final data set includes 380 principals that are merged to schools by the school ID. To supplement the administrative data we incorporate survey information from teachers and students.

II.b. Texas data

The Texas administrative data, assembled by the Texas Schools Project at the University of Texas at Dallas, include information from the Texas Education Agency (TEA), which combines different data sources to create matched panel data sets of students, teachers, and principals. The panels include all Texas public school teachers, administrators, staff, and students in each year, permitting accurate descriptions of the schools for each principal. Note that in the Texas data, administrators are assigned a unique ID, which facilitating the construction of the principal panel data set.

The Public Education Information Management System (PEIMS), the TEA's statewide educational database, reports key demographic data including race, ethnicity, and gender for students and school personnel as well as student eligibility for a subsidized lunch. PEIMS also contains detailed annual information on teacher and administrator experience, salary, education, class size, grade, population served, and subject. Importantly, this database is merged with information on student achievement by campus, grade, and year. Students in Texas were tested in math and reading each year, taking the Texas Assessment of Academic Skills (offered 1993-2002), the Texas Assessment of Knowledge and Skills (offered 2003-2012), or the State of Texas Assessments of Academic Readiness (offered 2012-2016). Each test was administered each spring to eligible students enrolled in grades three through eight.[2] These criterion

[2] Many special education and limited English proficient students are exempted from the tests. In each year roughly 15 percent of students do not take the tests, either because of an exemption or because of repeated absences on testing days.

referenced tests, which assess student mastery of grade-specific subject matter, are merged with the student and personnel information. Reading and math tests each contain approximately 50 questions, although the number of questions and average percent correctly answered varies across time and grades. We transform all test results into standardized scores with a mean of zero and variance equal to one for each subject, grade, and year, implying that our achievement measures describe students by their relative position in the overall state performance distribution.

One of the strengths of our sample is the large number of principals and schools that are observed. During the 1995-2016 sample period we observe 1491 unique principals in 714 different schools. This facilitates precise information of principal effects and supports an investigation of heterogeneity by initial achievement.

III. Empirical Framework

To estimate principal effectiveness, we build on existing methods described in a growing literature that aims to identify the effects of principals on test scores.[3] As highlighted in Branch et al. (2018), estimation of principal value-added must address many of the same but also some very different issues as estimation of teacher value added. On the one hand, family sorting into neighborhoods introduces potentially non-random variation in student composition among schools which must be addressed when estimating both principal and teacher value added. On the other hand, issues arising from the purposeful allocation of students into classrooms and test measurement error is

[3] These include including (Clark and Martorell 2009; Branch, Hanushek, and Rivkin 2012; Chiang, Lipscomb, and Gill 2012; Coelli and Green 2012; Hochbein and Cunningham 2013; Dhuey and Smith 2014; and Grissom, Kalogrides, and Loeb 2015).

mitigated when studying principals since performance is measured at the school level and principals oversee many more students than do individual teachers.[4] A unique challenge faced when estimating principal value added is the fact that principals may influence school quality even after leaving, given involvement in teacher hiring and establishing curriculum and school culture. Moreover, there exists no comparison principal at a single point in time, ruling out within school-year comparisons. In short, while the measurement of principal value added avoids some of the complications involved in the estimation of teacher value-added, it also presents unique challenges. In Section V we return to the question of whether within-school variation in school effects across cohorts can credibly be attributed to the principal. Until then we refer to principal effects which are equivalent to school effects during a principal's tenure.

To lessen the influences of prior principals, we eliminate any overlap in principals for students with different principals in grades 4 to 8. For example, if Ms. Smith served as principal in school A between 1995 and 2000 Ms. Jones served as principal in School A between 2001 and 2015, students who completed 8th grade in 1999 or 2000 would be included in the sample as Ms. Smith's students, but students who completed 8th grade in any year between 2001 and 2008 would be excluded because they would have attended the school under the leadership of both Ms. Smith and Ms. Jones. Inclusion of only students who completed 8th grade in 2009 or later ensures that the outcomes of Ms. Smith's students are compared with the outcomes of students who never had her as principal. Because CPS has relatively few schools with two principals who serve the five years necessary to have students in grades 4 to 8 and meet these sample conditions, in

[4]Kane et al. (2013), Chetty, Friedman, and Rockoff (2014), Rothstein (2010), and Guarino et al. (2014) investigate the presence and magnitude of biases introduced by nonrandom assignment to classrooms.

schools with one principal with tenure of at least five years, we create a second 'long-serving principal' by combining two or potentially three principal spells together. This enables us to compare elementary school students with other cohorts from the same school.

The estimation of effects on longer-term outcomes introduces additional complications associated with confounding influences in the years subsequent to completion of elementary school. These complications include differences in high-school quality, labor-market conditions, college tuition and quality, and the community environment. To mitigate bias introduced by these complications, we include high school by year fixed effects in some specifications to control for the myriad factors that influence high school and post-secondary outcomes. High-school classmates who attended different elementary schools experience the same high school principal and thus are exposed to a principal of the same effectiveness, similar local labor market conditions, and similar structure of college prices and opportunities following high school graduation. Because disciplinary and grading practices and policies vary across high schools, the focus on within high school by year variation restricts comparisons to those that are meaningful.

Estimates of principal effectiveness are generated from the following specification:

$$(1) \qquad A_{ishpt} = f(A_{ist-1}) + X_{ist} + S_{st} + \delta_s + \eta_{ht} + \theta_{ps} + \varepsilon_{ishpt}$$

Equation (1) models outcome (A) for student i in school s in year t with principal p and who attends high school h as a cubic function of prior achievement, student (X) and average school-year (S) controls, a series of error components including elementary (or middle) school (δ) and high school by year (η) fixed effects, a principal by school fixed effect (θ) that serves as the measure of principal effectiveness, and a random error (ε).[5] The high school by year fixed effect absorbs all time-invariant differences in high schools and all shocks with identical effects across the state. Although the CPS data do not contain elementary-school absences for most of the period, that is not the case for Texas. The absence of a control for prior non-cognitive skills may introduce bias, but we are able to examine the sensitivity of the Texas estimates for various outcomes and samples to the inclusion of an absence control.

Unobserved heterogeneity constitutes the primary threat to this empirical approach for estimating principal effects. Consider the possibility that students sort to high schools on the basis of 8th grade skills. If there are two elementary schools with identical student bodies at entry but the principal in School 1 is far more effective than the principal in School 2 at raising skills during elementary school, students in School 1 will tend to matriculate to more competitive high schools than students in School 2. However, the substantial variation in skills within elementary schools means that each high school would contain students from all the elementary schools, and within each high school students from Schools 1 and 2 would tend to have roughly equal skills. In a model with high school by year fixed effects, the focus on within high-school differences in outcomes for students from different elementary schools would ignore the existence of substantial differences in elementary school principal quality due to the sorting by skill among high schools. More generally, this type of sorting would bias downward the variance of elementary school principal productivity. Therefore, a model with high school

[5]The fixed-effect approach follows Bertrand and Schoar (2003), Grissom et al. (2015), Cannon, Figlio, and Sass (2013), and Branch, Hanushek, and Rivkin (2012).

by year fixed effects would tend to understate differences in principal productivity. Although this suggests the exclusion of these fixed effects, they do control for unobserved student factors related to the choice of high school. Moreover, GPA and disciplinary infraction comparisons are only meaningful among students in the same high school and year. Finally, random shocks to schools may inflate the variances, and we address this issue by applying Bayesian shrinkage methods.

IV. Results

We estimate a series of principal effects on academic, behavioral, and labor market outcomes based on equation 1. The academic outcomes include 8th grade mathematics and reading test scores, 9th grade mathematics and reading test scores and GPA by subject, 9th grade absences, receipt of any disciplinary infraction in 9th grade, college attendance and persistence, and the probability of being employed or in college following high school graduationWe define college persistence as being enrolled in a college for three consecutive terms. In Texas, we know the class standing for students and define persistence alternately as advancing at least one level, e.g. freshman to sophomore, or using the same definition as Chicago; the estimates are not sensitive to the definition. We set the indicator for being in work or in school equal to one if the student either has positive earnings or attends college or another educational institution in the year following high school graduation. We estimate four specifications for each outcome that differ by whether elementary and high school by year fixed effects are included. Although most outcomes are estimated over both samples, data limitations mean that some outcomes are only available in one of the two states.

The tables below report the variances of the raw principal fixed effects and their correlations. In a subsequent version of this manuscript, we will provide variance estimates based on a Bayesian shrinkage procedure. In addition to estimates based on all students in schools, we divide the sample on the basis of prior mathematics achievement (above or below the median for CPS or Texas).

IV.a. Estimated Principal Effectiveness and the Correlation with longer-term outcomes

In Table 1, we present the standard deviations of principal effects on each outcome, separately for Texas and Chicago. In both settings, the estimates reveal substantial variation across principals over each dimension of performance. The magnitudes remain largely unchanged by the addition of high-school by year fixed effects but fall substantially following the inclusion of elementary or middle school fixed effects, particularly in Texas. This is consistent with the persistent selection into Texas middle schools. Finally, a comparison of columns 4 and 5 in the bottom panel illustrates that only the standard deviation of the effect on 9th grade absences declines (by roughly 15 percent) following the inclusion of controls for prior absences; .

Focusing on specifications including both high-school by year fixed effects and elementary school fixed effects (columns 4 and 8), the results suggest that a one standard deviation improvement in early principal quality is associated with a roughly 0.04 standard deviation increase in grade 8 reading scores in Texas and a 0.07 increase in CPS. For grade 8 mathematics, the effects are larger (0.07 standard deviations in Texas and 0.10 standard deviations in CPS). In general, the estimated standard deviations tend to be larger in Chicago, consistent with the additional grades in CPS K-8 elementary schools and potentially larger effects at younger ages.

Principal effects on longer-term academic and behavioral outcomes also tend to range between 0.1 and 0.15 of a standard deviation of the raw outcome for CPS but vary somewhat more in Texas. Specifically, differences in effects on 9th grade achievement and absences tend to be smaller in Texas, while differences in effects on post-secondary school attendance exceed 0.2 in Texas.

In Table 2 we present correlations between estimated principal effects on multiple student outcomes for our Chicago sample; the estimates come from the full specification with both elementary school and high school by year fixed effects. The top panel reports correlations between principal effects on 8th grade math and reading achievement and 9th grade reading and math achievement, absences and disciplinary infractions. We estimate strong and significant correlations between principal effects on 9th grade mathematics achievement and the effects on 8th grade achievement in both subjects; the correlations between the effect on 9th grade reading achievement and the effects on the two 8th grade achievement measures are smaller and less significant in the case of reading. Moreover, none of the correlations between the effects on 8th grade mathematics or reading scores on the one hand and effects on 9th grade absences or disciplinary infractions had meaningful or significant correlations. This is consistent with the findings in Jackson (forthcoming) of a weak correlation between effects on cognitive and non-cognitive skills.

The second panel reports the correlations between elementary-school principal effects on 9th grade mathematics and reading GPA on the one hand and effects on 8th grade achievement and 9th grade behaviors on the other. Although effects on GPA are positively related to effects on 8th grade mathematics and reading achievement are all positive, they are small and insignificant at conventional levels. In contrast, effects on both mathematics and reading GPA are strongly related to effects on absences and the probability of receiving a disciplinary infraction. Remarkably, the correlations with the effect on absences exceed 0.6 for both mathematics and reading GPA, and the correlations with the effects on disciplinary infractions exceed 0.25. The primacy of non-cognitive skills in the determination of GPA mirrors the findings in Jackson (forthcoming).

The second panel shows correlations between effects on 9th grade mathematics and reading GPA on the one hand and effects on cognitive and non-cognitive skills. Although correlations between the effects on cognitive skills and GPA are all positive, they are small and insignificant. In contrast, the correlations with effects on absences and disciplinary infractions are large and highly significant, again highlighting the importance of behavioral skills. The absence of controls for prior absences suggests some caution in the interpretation of these correlations, and unfortunately GPA data are not available for Texas. However, even if these correlations were to decline in half as suggested by the results for Texas reported below, they would remain large.

The third panel shows correlations between post-secondary schooling on the one hand and the elementary and secondary school outcomes on the other. The correlations between effects on 8th grade achievement and effects on post-secondary schooling are stronger for math and roughly twice as large for the effects on college persistence than on college attendance. In contrast, effects on the behavioral outcomes and GPA are more strongly correlated with the effect on college attendance than the effect on college persistence. This is consistent with the notion that fostering non-cognitive skills can increase high-school attainment and elevate expectations while the production of cognitive skills is relatively more important to success in college.

Table 3 reports the corresponding correlations for Texas based on specifications that alternately include or omit the control for prior absences. As in the case for the standard deviations, only the correlations that involve principal effects on absences are sensitive to the inclusion of prior absences. In contrast to the pattern for CPS, the correlations between effects on cognitive skills as measured by 8^{th} grade mathematics and reading achievement and effects on college persistence are not significant, and only the effect on reading achievement is significantly correlated with college attendance. Moreover, the correlations between the effect on absences and effects on post-secondary schooling are much smaller and only marginally significant in specifications that control for prior absences. Effect on absences is also not significantly correlated with effect on the probability of being employed or in school, controlling for prior absences.

The next part of the analysis explores the possibility that the patterns of principal effects may vary by prior academic preparation. Students below CPS and Texas median test scores and those above the medians are placed into separate groups, and principal effects on low and high achievement students are estimated. Estimates are positively correlated, meaning that more productive principals tend to be more productive with students across the achievement distribution. Note that although the CPS and Texas test scores are not comparable, average achievement in CPS falls well below the average for Illinois. The more diverse circumstances and larger variation among schools and districts in the state of Texas as opposed to the single urban district in Illinois suggests that differences by initial achievement might be expected to be larger in Texast.

Table 4 reports standard deviations of principal effects from the full model with both elementary school and high school by year fixed effects, and the standard deviations tend not to vary much by prior achievement. The primary exceptions are effects on 9th grade reading and absences in CPS, where the standard deviations are much larger for below median achievement students.

Table 5 presents correlations of the effects for CPS by prior achievement, and pronounced differences emerge in the correlations between effects on college attendance and persistence on the one hand and effects on both achievement and non-cognitive skills on the other. First, associations with value added to 8th grade math score are stronger for higher achievers; the correlations are less than half as large for lower achievers. Second, the associations with effects on absences are also much stronger for high achievers, though they are all highly significant. Third the correlation between effect on disciplinary infractions and college attendance is large and significant for low achievers but essentially zero for high achievers.

Table 6 presents results for Texas by prior achievement for specifications with and without the control for prior absences, and a sharply different pattern emerges. Specifically, the correlations between effects on college attendance and persistence on the one hand and effects on achievement and absences tend to be much larger in magnitude for below median achievement students. In the case of college persistence, the effects on achievement and absence are positively and significantly correlated only for below median achievement students. Importantly, the inclusion of prior absences as a control has little effect on these correlations for below median achievement students but a much larger effect for above median achievement students. The CPS data do not include information on employment, but the correlations in the bottom panel suggest that raising both cognitive and non-cognitive skills raises the probability of engaging in productive activity following high school for below median achievement students.

The lack of sensitivity of the principal effect estimates based on low-achievement students in Texas supports the belief that the lack of a control for prior absences does not substantially inflate effects for CPS because of the differences in demographic composition. While roughly 45 percent of Texas middle school students are eligible for a subsidized lunch, over 80 percent of CPS students are so eligible. This suggests that the pattern of results for the below median achievement Texas effects likely provides a better benchmark for CPS than the sample of all students.

V. Patterns of effects by principal turnover

Controls for prior achievement and absences in combination with the inclusion of elementary (or middle) and high school by year fixed effects enable the identification of elementary (or middle) school by period value added to outcomes. Because different principals lead a school during the different periods, we attribute the effects to principals. However, there may be other factors not under the control of the principal that differ between periods, and therefore it is possible that these account for some, most or even the entire variation between periods. Examples of such factors include the composition of teachers or changes in curriculum.

Our approach to providing additional information on the contributions of principals involves the estimation of the variance in ‘principal’ effects for a sample in which two cohorts of students have the same principal. Consider a sample of schools with long-serving principals whose tenure equals or exceeds ten years in Chicago or six years in Texas. Division of these spells in half creates two ‘principals’ who in actuality are the same person. Under the assumption of constant principal effects on value-added, the principal would contribute nothing to within-school differences in outcomes for these cohorts; both would have had the same principal. Therefore, variance estimates from this sample would provide counterfactual estimates for what the variation would have been in the schools that experienced a principal transition in the absence of the change in principal. The difference between the variances based on the sample of schools with multiple principals and the sample in which the tenure of a single, long-serving principal is divided in half would provide an estimate of the true variances in principal value-added to the various outcomes.

A key assumption underlying this falsification comparison is constant principal effects throughout a spell at a school, and this contradicts the notion that school quality evolves under a principal through changes in the stock of teachers, support for teacher development, and introduction of practices related to learning and school climate. Therefore, the comparison between single- and multi-principal schools provides a lower bound estimate of the variation attributed to principals which may not be particularly informative if longer-serving principals tend to be able to enact more profound changes to school operations.

Differences between single and multi-principal schools could dampen the value of the variance comparisons between the two sets of schools, and we take steps to make the samples more comparable. First, we use propensity score matching on the probability of having a single, long-serving principal to generate a comparison group of schools with multiple principals. The predictors include number of test-takers, a primary determinant of the variance of the principal fixed effect, and demographic variables. Second, we restrict the sample to one cohort per principal spell. Third, we continue to control for prior achievement and demographic characteristics, though we exclude campus average characteristics due to the limitation of the sample to one student cohort per spell; the basic results are not sensitive to the exclusion of these variables. The final sample includes 56 schools in each category.

Tables 7 present fixed effect variance estimates based on the divided spells of long-serving principals and two samples of schools with multiple principals, one with all such schools and the other with a comparable sample based on propensity score matching. Each school includes only two cohorts: one per principal in the multiple principal schools and two for the principal in the single principal schools.

The table does not reveal much larger variation in the multiple principal schools which would be expected under constant principal effects and substantial fixed differences in principal productivity. Rather some of the variances are larger in the single principal schools and some are larger in the multiple principal schools. For example, the variances of the grade 9 test score and absence effects are much larger in the multiple principal schools, but there is little difference in the variances for the effects on grade 8 test scores or college persistence and the single principal schools have a larger variance in the effect on math GPA. Note that matching makes little difference in most cases.

Table 8 provides additional information on the cohort differences in effects by principal turnover and the average changes in schools with a single principal throughout the period. The right panel shows mean absolute differences in later cohort minus earlier cohort effects, and these tend to be more consistently larger in the multiple principal schools than the variances. Only mean absolute differences in effects on disciplinary infractions and 8th grade reading scores are not larger in the multiple principal schools. Disciplinary infractions in high school tend to be concentrated among a small number of students, and middle schools tend not to have large effects on reading scores, so this pattern is more consistent with expectations leadership productivity differences being larger for different principals. Clearly the variation between schools with a single and those with multiple principals requires further exploration.

The right panel of Table 8 reports average differences in effects between the late earlier cohorts, and not surprisingly the differences tend to show relatively higher school quality for later cohorts in single principal but not in multiple principal schools. Comparisons of single principal schools to the matched sample of multiple principal schools reveal more favorable changes in all outcomes except 9th grade math score. In some cases the differences are small, but in the cases of effects on 8th grade test scores, both 9th grade math and reading GPA, and college attendance and persistence the differences are sizeable.

VI. Conclusion

Using a fixed effects framework that accounts for fixed differences in elementary schools and time-varying differences in secondary schools as well as prior achievement, the analysis provides strong evidence that elementary and middle school principals and schools affect test scores, behaviors, and longer-term educational and labor market outcomes. In addition, the analysis highlights the importance of considering longer-term and behavior outcomes, as the correlation between school effects on test scores and school attendance tend to be positive but small and effects on attendance tend to be more highly correlated with longer-term secondary and post-secondary outcomes for initially lower achievement students.

The pattern of variation in school effects by principal turnover is not consistent with a simple model of fixed principal effects and substantial differences among principals. Rather it appears that long-serving principals tend to make substantial changes to schools that produce significant differences in the quality of schooling that are mostly positive. However, this issue clearly merits additional investigation.

Bibliography (preliminary and incomplete)

Bertrand, Marianne, and Antoinette Schoar. 2003. "Managing with Style: The Effect of Managers on Firm Policies." *The Quarterly Journal of Economics* 118, no. 4 (November): 1169-1208.

Bloom, Nicholas, and John Van Reenen. 2010. "Why do management practices differ across firms and countries?" *Journal of Economic Perspectives* 24, no. 1 (Winter): 203-224.

Brewer, Dominic J. 1993. "Principals and student outcomes: Evidence from U.S. high schools." *Economics of Education Review* 12, no. 4 (December): 281-292.

Branch, Gregory, Eric A. Hanushek, Steven G. Rivkin, and Jeffrey C. Schiman. "Variation and Determinants of the Productivity of Public Sector Managers: The Case of School Principals "

Card, D., AND A. B. Krueger. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," Journal of Political Economy, 100, 1–40.

Carneiro, P. and J. J. Heckman (2003). Human capital policy. In J. J. Heckman, A. B. Krueger, and B. M. Friedman (Eds.), *Inequality in America: What Role for Human Capital Policies?*, pp. 77–239. Cambridge, MA: MIT Press.

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

Coelli, Michael and David Green. 2012. "Leadership Effects: School Principals and Student Outcomes." *Economics of Education Review*, 31(1)

Dhuey, Elizabeth, and Justin Smith. 2014. "How important are school principals in the production of student achievement?" *Canadian Journal of Economics/Revue canadienne d'économique* 47, no. 2 (May): 634-663.

Eberts, Randall W., and Joe A. Stone. 1988. "Student achievement in public schools: Do principals make a difference?" *Economics of Education Review* 7, no. 3: 291-299.

Grissom, Jason, and Susanna Loeb. forthcoming. "Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills." *American Educational Research Journal*.

Hanushek, Eric A., and Steven G. Rivkin. 2010a. "Constrained Job Matching: Does Teacher Job Search Harm Disadvantaged Urban Schools?" NBER w15816. Cambridge, MA: National Bureau of Economic Research (March).

Hanushek, Eric A., and Steven G. Rivkin. 2010b. "Generalizations about using value-added measures of teacher quality." *American Economic Review* 100, no. 2 (May): 267-271.

Jackson, C. Kirabo. 2018. "What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes. *Journal of Political Economy.*

Morris, Carl N. 1983. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association* 78, no. 381 (March): 47-55.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica* 73, no. 2 (March): 417-458.

Rothstein, Jesse. 2010. "Teacher quality in educational production: Tracking, decay, and student achievement." *Quarterly Journal of Economics* 125, no. 1 (February): 175-214

Table 1. Estimated Effects of CPS Elementary and Texas Middle School Principals on academic, behavioral and labor-market outcomes

	Standard deviation of estimated principal effects				raw outcome mean	raw outcome standard deviation
CPS						
Grade 8 outcomes						
Reading score	0.12	0.12	0.07	0.07	0.015	1
Math score	0.16	0.16	0.10	0.10	0.014	1
Grade 9 outcomes						
Reading score	0.12	0.14	0.08	0.10	0.025	1
Math score	0.15	0.14	0.10	0.09	0.027	0.99
Absences	4.41	4.52	2.88	2.50	21.8	26
Any disciplinary infractions	0.09	0.09	0.06	0.06	0.28	0.45
Math GPA	0.18	0.19	0.12	0.12	1.83	1.21
Reading GPA	0.18	0.20	0.12	0.13	2.02	1.17
post-secondary outcomes						
College attendance	0.09	0.08	0.06	0.04	0.19	0.39
College persistence	0.08	0.06	0.04	0.04	0.12	0.32
High school by year fixed effects	N	Y	N	Y		
Elementary school fixed effects	N	N	Y	Y		

	Standard deviation of estimated principal effects					raw outcome mean	raw outcome standard deviation
Texas							
Grade 8 outcomes							
reading score	0.07	0.07	0.04	0.04	0.04	0.02	0.98
math score	0.11	0.11	0.07	0.07	0.07	0.02	0.98
Grade 9 outcomes							
reading score	0.12	0.12	0.05	0.04	0.04	0.03	0.99
math score	0.14	0.14	0.07	0.07	0.07	0.04	0.98
absences	1.01	1.01	0.42	0.42	0.37	8.7	11.01
any disciplinary infractions	0.06	0.06	0.04	0.04	0.04	0.15	0.39
post-secondary outcomes							
college attendance	0.34	0.34	0.10	0.10	0.10	0.35	0.48
college persistence	0.26	0.26	0.08	0.08	0.08	0.18	0.42
Working or in college after hs	0.40	0.40	0.12	0.12	0.12	0.55	0.5
High school by year fixed effects	N	Y	N	Y	Y		
Elementary school fixed effects	N	N	Y	Y	Y		
Includes prior absences	N	N	N	N	Y		

Table 2. Correlations between principal effects on achievement, behavior and longer-term outcomes for CPS on estimates from specifications that include elementary school and high school by year fixed effects

1. Correlations between effects on 8th grade math and reading achievement and effects on 9th grade achievement and behavior

	9th grade outcomes			
	math score	reading score	absences	any disciplinary infractions
8th grade math achievement	0.38***	0.15***	-0.11	0.02
8th grade reading achievement	0.27***	0.12	-0.07	-0.08

2. Correlations between effects on 9th grade math and reading GPA and effects on 8th grade achievement and 9th grade behavior

	8th grade test scores		9th grade behavior	
	math	reading	absences	disciplinary infractions
9th grade math GPA	0.11	0.11	-0.65***	-0.25***
9th grade reading GPA	0.05	0.1	-0.71***	-0.31***

3. Correlations between effects on college attendance and persistence and effects on 8th and 9th grade academic and behavior outcomes

	8th grade test scores		9th grade behavior		9th grade GPA	
	math	reading	absences	disciplinary infractions	math	reading
college attendance	0.20***	0.09	-0.32***	-0.03	0.29***	0.37***
college persistence	0.29***	0.15**	-0.23***	0.04	0.17**	0.25***

Table 3. Correlations between principal effects on achievement, behavior and longer-term outcomes for Texas on estimates from specifications that include elementary school and high school by year fixed effects

1a. Correlations between effects on 8th grade math and reading achievement and effects on 9th grade achievement and behavior

	9th grade outcomes			
	math score	reading score	absences	any disciplinary infractions
8th grade math achievement	0.42***	0.28***	-0.12**	-0.05
8th grade reading achievement	0.18***	0.48***	-0.18***	-0.10**

1b. Correlations between effects on 8th grade math and reading achievement and effects on 9th grade achievement and behavior controlling for prior absences

	9th grade outcomes			
	math score	reading score	absences	any disciplinary infractions
8th grade math achievement	0.42***	0.27***	-0.08	-0.05
8th grade reading achievement	0.17***	0.48***	-0.13***	-0.10**

2a. Correlations between effects on college and work outcomes and effects on 8th and 9th grade academic and behavior outcomes

	8th grade test scores		9th grade behavior	
	math	reading	absences	disciplinary infractions
college attendance	0.07	0.10**	-0.15***	0.07
college persistence	0.01	0.01	-0.14***	0.07
work or college	0.13***	0.22***	-0.14***	0.00

2b. Correlations between effects on college and work outcomes and effects on 8th and 9th grade academic and behavior outcomes controlling for prior absences

	8th grade test scores		9th grade behavior	
	math	reading	absences	disciplinary infractions
college attendance	0.07	0.10**	-0.09*	0.07
college persistence	0.00	0.00	-0.08*	0.07
work or college	0.12**	0.22***	-0.06	0.00

Table 4. Estimated Effects of CPS Elementary and Texas Middle School Principals on academic, behavioral and labor-market outcomes from specifications that include elementary school and high school by year fixed effects, by prior achievement

	Standard deviation of estimated principal effects	
	CPS	Texas

Prior Achievement	below median	above median	below median	above median
Grade 8 outcomes				
reading score	0.08	0.08	0.13	0.13
math score	0.11	0.11	0.20	0.20
Grade 9 outcomes				
reading score	0.17	0.13	0.17	0.19
math score	0.13	0.12	0.21	0.19
absences	3.17	2.56	5.77	4.96
any disciplinary infractions	0.08	0.07	0.13	0.12
math GPA	0.17	0.16	0.24	0.25
reading GPA	0.16	0.16	0.25	0.25
post-secondary outcomes				
college attendance	0.05	0.06	0.11	0.09
college persistence	0.04	0.05	0.09	0.07

Table 5. Correlations between principal effects on achievement, behavior and longer-term outcomes for CPS from specifications that include elementary school and high school by year fixed effects, by prior achievement

1. Correlations between effects on 8th grade math and reading achievement and effects on 9th grade achievement and behavior

	9th grade outcomes			
	math score	reading score	absences	any disciplinary infractions

Below median prior achievement				
8th grade math achievement	0.30***	0.15**	-0.22***	-0.07
8th grade reading achievement	0.27***	0.34***	-0.11	-0.27***
Above median prior achievement				
8th grade math achievement	0.39***	0.24***	-0.20***	-0.03
8th grade reading achievement	0.15**	0.10	-0.11	-0.05

2. Correlations between effects on 9th grade math and reading GPA and effects on 8th grade achievement and 9th grade behavior

	8th grade test scores		9th grade behavior	
	math	reading	absences	disciplinary infractions
Below median prior achievement				
9th grade math GPA	0.24***	0.01	-0.52***	-0.24***
9th grade reading GPA	0.24***	0.08	-0.55***	-0.28***
Above median prior achievement				
9th grade math GPA	0.20***	0.06	-0.63***	-0.23***
9th grade reading GPA	0.05	0.08	-0.56***	-0.20**

3. Correlations between effects on college attendance and persistence and effects on 8th and 9th grade academic and behavior outcomes

	8th grade test scores		9th grade behavior		9th grade GPA	
	math	reading	absences	disciplinary infractions	math	reading

Below median prior achievement						
college attendance	0.08	0.10	-0.24***	-0.20**	0.24***	0.25***
college persistence	0.13*	0.14*	-0.14*	-0.10	0.17**	0.23***
Above median prior achievement						
college attendance	0.27***	0.15*	-0.41***	-0.00	0.36***	0.37***
college persistence	0.29***	0.15*	-0.29***	-0.03	0.30***	0.35***

Table 6. Correlations between principal effects on achievement, behavior and longer-term outcomes for Texas from specifications that include elementary school and high school by year fixed effects, by prior achievement

1a. Correlations between effects on 8th grade math and reading achievement and effects on 9th grade achievement and behavior

	9th grade outcomes			
	math score	reading score	absences	any disciplinary infractions
Below median prior achievement				
8th grade math achievement	0.46***	0.28**	-0.10**	-0.00
8th grade reading achievement	0.14***	0.48***	-0.15***	-0.11**
Above median prior achievement				
8th grade math achievement	0.44***	0.20***	-0.20***	0.25***
8th grade reading achievement	0.33**	0.48***	-0.24***	0.28***

1b. Correlations between effects on 8th grade math and reading achievement and effects on 9th grade achievement and behavior controlling for prior absences

9th grade outcomes

	math score	reading score	absences	any disciplinary infractions
Below median prior achievement				
8th grade math achievement	0.45***	0.28***	-0.05	0.00
8th grade reading achievement	0.14***	0.48***	-0.11**	-0.10**
Above median prior achievement				
8th grade math achievement	0.43***	0.19***	-0.18***	0.25***
8th grade reading achievement	0.31**	0.48***	-0.28***	0.28***

2a. Correlations between effects on college and work outcomes and effects on 8th and 9th grade academic and behavior outcomes

	8th grade test scores		9th grade behavior	
	math	reading	absences	disciplinary infractions
Below median prior achievement				
college attendance	0.18***	0.41***	-0.16***	0.01
college persistence	0.13**	0.24***	-0.16***	0.09*
work or college	0.22***	0.48***	-0.17***	-0.03
Above median prior achievement				
college attendance	0.15***	-0.03	-0.09**	-0.00
college persistence	-0.03	-0.15***	0.05	0.01
work or college	0.14***	-0.05	-0.07	0.06

2b. Correlations between effects on college and work outcomes and effects on 8th and 9th grade academic and behavior outcomes controlling for prior absences

	8th grade test scores		9th grade behavior	
	math	reading	absences	disciplinary infractions
Below median prior achievement				
college attendance	0.18***	0.41***	-0.14***	0.02
college persistence	0.12**	0.23***	-0.14***	0.10**
work or college	0.22*	0.43***	-0.14***	-0.02
Above median prior achievement				
college attendance	0.15***	-0.04	-0.04	0.14***
college persistence	-0.04	-0.16***	0.07	0.02
work or college	0.14*	-0.07	-0.03	0.07

Table 7. Variances of school by cohort estimated effects from specifications that include elementary school and high school by year fixed effects, by principal turnover and comparison sample of multiple principal schools

Same principal for both cohorts	Y	N	N
Matched comparison sample		Y	N
Grade 8 outcomes			
Reading score	0.007	0.006	0.005
Math score	0.012	0.012	0.012
Grade 9 outcomes			
Reading score	0.008	0.012	0.025
Math score	0.007	0.012	0.013
Absences	4.02	5.97	5.87
Any disciplinary infractions	0.0034	0.0040	0.0044
Math GPA	0.042	0.016	0.018
Reading GPA	0.015	0.020	0.022
Post-secondary outcomes			
College attendance	0.0035	0.0008	0.0020
College persistence	0.0011	0.0011	0.0013

Table 8. Mean differences and absolute differences in estimated fixed effects, later cohort minus earlier cohort, from specifications that include elementary school and high school by year fixed effects, by principal turnover and comparison sample of multiple principal schools

	mean differences: later cohort minus earlier cohort			mean absolute differences: absolute value of later cohort minus earlier cohort		
Same principal for both cohorts	Y	N	N	Y	N	N
Matched comparison sample		Y	N		Y	N
Grade 8 outcomes						
Reading score	0.069	0.018	-0.022	0.237	0.232	0.211
Math score	0.073	-0.009	-0.034	0.270	0.304	0.311
Grade 9 outcomes						
Reading score	0.024	0.057	0.016	0.197	0.306	0.270
Math score	-0.045	0.023	-0.029	0.211	0.324	0.318
Absences	-0.17	5.08	4.65	7.84	10.22	9.43
Any disciplinary infractions	0.0792	0.0522	0.0645	0.2011	0.2093	0.2090
Math GPA	0.198	-0.164	-0.130	0.257	0.447	0.407
Reading GPA	0.297	-0.177	-0.199	0.344	0.472	0.409
Post-secondary outcomes						
College attendance	0.019	-0.046	-0.101	0.125	0.178	0.201
College persistence	-0.018	-0.026	-0.065	0.107	0.150	0.158