Technical appendix.

How Much Does Leadership Matter?
Effects of Elementary School Principals in Chicago on Students' High School Outcomes

By Steven G. Rivkin, Jeffrey C. Schiman, and Lauren Sartain

Updated on April 19, 2023

This study examines the extent to which elementary school principals have effects on student achievement and long-term outcomes, creating value-add metrics on principals' effects on students' eighth grade achievement (while in the school with the principal), as well as principals' effects on their students' ninth grade outcomes (achievement, absences, and disciplinary infractions), and post-secondary outcomes (college attendance and college persistence). We show the degree to which these principal effects are correlated—that is, whether principals who have stronger effects on some outcomes also have stronger effects on other outcomes. Finally, we examine the relationships between principal effectiveness and teacher reports about principal leadership and school climate on annual 5Essentials surveys.

CPS Data Description

The CPS panel data sets span twenty years, from 1993-94 to 2013-14, and follow children as they progress through elementary, secondary, and post-secondary schooling. Specifically, we use administrative and survey data from the University of Chicago Consortium on School Research.

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 1993-94 to 2010-11, and the ACT Explore test in the spring of their 9th grade year from 2011-12 to 2013-14. There is also National Clearinghouse data on college attendance for CPS high school graduates matched with the administrative records.

Data on CPS personnel contain information about the current position, allowing us to observe the principal of record at a given school 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. The availability of rich teacher, principal, and student surveys along with the longitudinal quantitative outcome data also enables us to investigate other channels through which effective principals raise achievement. This descriptive analysis controls for unobserved student heterogeneity with prior student achievement and for unobserved school differences with school fixed effects, thus enabling us to extend substantially the evidence presented in Bloom, Lemos, Sadun, and Van Reenen (2015) on the relationship between student achievement and principal practices and skills. In the study years, CPS administered the UChicago Consortium's 5Essentials survey to all CPS teachers and CPS students in grades 6-12 every other year (now annually).

We use the following items from the 2013 survey and construct a single measure of leadership quality:

- Makes teaching expectations clear
- Communicates clear vision for school
- Understands how students learn
- Sets high standards for student learning
- Presses teachers to implement professional development
- Tracks student academic progress

We take the resulting measure and standardize it within year, so the average principal in a given school year has a value of 0 with a standard deviation of 1. Principals who have higher (lower)-than-average reported leadership quality have values above (below) 0 on the measure. In our analysis, we then correlate that measure of leadership quality with our estimated principal effects on student test score growth (i.e., principal value-added measures).

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.[1] As highlighted in Branch et al. (2020), 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 mitigated when studying principals since performance is measured at the

[1] These include including (Clark and Martorell 2009; Branch, Hanushek, and Rivkin 2012; Chiang, Lipscomb, and Gill 2016; Coelli and Green 2012; Hochbein and Cunningham 2013; Dhuey and Smith 2014; and Grissom, Kalogrides, and Loeb 2015).

school level and principals oversee many more students than do individual teachers.[2] 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.

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 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

[2]Kane et al. (2013), Chetty, Friedman, and Rockoff (2014), Rothstein (2010), and Guarino et al. (2015) investigate the presence and magnitude of biases introduced by nonrandom assignment to classrooms.

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 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 variation restricts comparisons to those that are meaningful.

Estimates of principal effectiveness are generated from the following specification:

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

Equation (1) models outcome (A) for student i in school s in year t with principal p and who attends high school has 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), henceforth referred to as elementary school; and high school (η) fixed effects, a principal by school fixed effect (θ) that serves as the measure of principal effectiveness, a year fixed effect (π) and a random error (ε).[3] The high school fixed effect absorbs all time-invariant differences in high schools. By subsequently demeaning the estimates by elementary school, we account for fixed differences among elementary schools.

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

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

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 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 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. Finally, random shocks to schools may inflate the variances, and we address this issue by applying Bayesian shrinkage methods.

Results

We estimate a series of principal effects on academic and behavioral 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 in college following high school graduation. We define college persistence as being enrolled in a college for three consecutive terms. Any evidence that effects on the probability of

attending college are strongly related to principal effects on cognitive and noncognitive skill acquisition would support the belief that they capture meaningful differences in the likelihood of engaging in no productive activities. Finally, we estimate four specifications for each outcome that differ by whether elementary and high school fixed effects are included.

Variation of principal effects

In Table 1, we present the standard deviations of principal effects on each outcome. The estimates reveal substantial variation across principals over each dimension of performance. The magnitudes remain largely unchanged by the addition of high-school fixed effects. Focusing on specifications with high-school fixed effects that demean by elementary school averages (Column 4), the results suggest that a one standard deviation improvement in elementary-school principal quality is associated with a roughly 0.08 standard deviation increase in grade 8 reading scores and a slightly higher 0.10 standard deviation in grade 8 math scores. Principal effects on high school and longer-term academic and behavioral outcomes also tend to hover around 0.1 of a standard deviation of the raw outcome.

Correlations among principal effects on different outcomes

Table 2 report correlations among the estimated principal effects on multiple student outcomes. The estimates come from the full specification with high school fixed effects and subsequent demeaning of the estimates by the elementary school averages. 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. The results show strong and significant correlations between principal effects on immediate effects on 8th-grade achievement and longer-term effects on those students' 9th-grade mathematics, though the relationships weaker 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 on the other had are sizeable or significant. This is consistent with the findings in Jackson (2018) of a weak correlation between teacher 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, 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 (2018).

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 50 percent larger 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, though effects on absences and both math and reading GPA are significantly related to effects on college attendance and persistence. This is consistent with the notion that the acquisition of mathematics and reading skills becomes more important relative to the acquisition of noncognitive skills as a student progresses through post-secondary schooling.

Correlations between principal practices and student achievement growth

Measurement of the variance in principal effectiveness highlights the importance of school leadership, and the previous section provides causal evidence that better personnel management is one channel through which more effective principals raise the quality of instruction. Principals have complex jobs that range from providing instructional leadership to making personnel decisions to purely administrative tasks. Using survey responses of school managers across eight countries, Bloom, Lemos, Sadun, and Van Reenen (2015) describe the relationships between student outcomes and particular management practices in a range of public and private schools. Their rigorous survey methodology provides information about differences in school management within and across countries, and the empirical analysis reveals a positive relationship between achievement on the one hand and manager skills and use of practices considered to be effective on the other.[4] However, the absence of controls for prior achievement and limited information on students and schools in their empirical analysis likely amplify the influences of student and school heterogeneity on the estimated effects of management skills and practices. By combining the survey information with longitudinal achievement data on Chicago students, we are able to extend this line of research with much richer controls for underlying differences in students and schools.

CPS principal surveys contain information on each principal's practices, though questions vary considerably from year to year.[5] Teacher and student surveys contain questions related to leadership along with questions on various aspects of school climate. We focus on survey responses that appear particularly relevant to principal effectiveness, re-appear in

[4] These include practices elucidated in Fryer (2014) that were found to have a causal effect on achievement.
[5] Measuring principal management practices through surveys of the principals themselves was an innovation of Bloom, Lemos, Sadun, and Van Reenen (2015) and their earlier management studies in other areas.

virtually all years of the data, and are reasonably objective in nature to lessen the contribution of the principal's popularity to the variation in responses.[6] Teachers are asked the extent to which their principal sets clear expectations, communicates a vision, understands student learning, sets high learning standards, implements professional development, and tracks student achievement. Students are asked the extent to which they feel safe in the classroom and the hallways, the extent to which they find the curriculum interesting and challenging, and whether they work hard to do their best.

The simplest way to see how management and leadership (as shown through these surveys) relates to principal effectiveness is to array average estimates of principal value-added by survey responses to each question. Table 3 shows that average principal value-added increases monotonically for all questions as student and teacher responses become more positive about principal leadership and school climate. The patterns are remarkably similar for student responses regarding safety and their level of engagement with the material and for teacher responses regarding instructional leadership and focus on achievement. The gaps between the top and bottom responses are around 0.02 standard deviations for the student responses and somewhat larger for the teacher responses, particularly for questions related to the focus on achievement.

Table 4 reveals strong, highly significant relationships between achievement and all three student and teacher indices in specifications that include prior achievement and school fixed effects regardless of whether the specifications also include indicators for first or last year of a spell (Columns 2 and 3). Removal of the school fixed effects leads to little change in the coefficient on the teacher leadership impression index but substantial changes to the coefficients

[6] We thank Principals Susan Kick and Linda Shay for their guidance on item selection.

on the student interest and safety indexes. The decrease in the magnitude and significance of the coefficient on the interest index and increase in magnitude of the coefficient on school safety index moving from Column 2 to Column 1 suggests the presence of unobserved school differences that affect perceptions about academic engagement and school safety and are related to achievement growth.

Not surprisingly, the magnitudes of the coefficients in the specification with neither prior achievement nor school fixed effects are much larger, though the coefficient is surprisingly negative in the case of the student interest index. Moreover, the substantial declines in magnitude following the inclusion of school fixed effects in these specifications suggests that the inclusion of prior achievement controls for unobserved student and school heterogeneity.

On balance, the findings in the preferred specifications that account for prior achievement and school fixed effects are consistent with instructional leadership and the establishment of a safe and engaging climate as channels through which principals raise achievement. Though not causal, these estimates strongly suggest that more effective principals tend to be perceived as better instructional leaders and tend to produce safer and more engaging school environments.

References

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.

Branch, Gregory F., Eric A. Hanushek, and Steven G. Rivkin. 2012. "Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals." NBER Working Paper W17803. Cambridge, MA: National Bureau of Economic Research (January).

Branch, Gregory F., Eric A. Hanushek, and Steven G. Rivkin. 2020. "Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals." NBER Working Paper W17803. Cambridge, MA: National Bureau of Economic Research (January).

Cannon, Sarah, David Figlio, and Tim Sass. "Principal quality and the persistence of school policies." *Unpublished manuscript* (2012).

Chetty, Raj, John N. Friedman, and Jonah Rockoff. 2014. "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates." *American Economic Review* 104, no. 9 (September): 2593-2632.

Chiang, Hanley, Stephen Lipscomb, and Brian Gill. 2016. "Is School Value Added Indicative of Principal Quality?" *Education Finance and Policy* 11, no. 3 (Summer): 283–309.

Clark, Damon, Paco Martorell, and Jonah Rockoff. 2009. "School Principals and School Performance." CALDER Working Paper 38. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research (CALDER) (December).

Coelli, Michael, and David A. Green. 2012. "Leadership effects: school principals and student outcomes." *Economics of Education Review* 31, no. 1 (February): 92-109.

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

Grissom, Jason A., Demetra Kalogrides, and Susanna Loeb. 2015. "Using Student Test Scores to Measure Principal Performance." *Educational Evaluation and Policy Analysis* 37, no. 1 (March): 3-28.

Guarino, Cassandra M., Mark D. Reckase, and Jeffrey M. Wooldridge. "Can value-added measures of teacher performance be trusted?." *Education Finance and Policy* 10, no. 1 (2015): 117-156.

Hochbein, Craig, and Brittany C. Cunningham. 2013. "An Exploratory Analysis of the Longitudinal Impact of Principal Change on Elementary School Achievement." *Journal of School Leadership* 23, no. 1: 64-90.

Jackson, C. Kirabo. "What do test scores miss? The importance of teacher effects on non–test score outcomes." *Journal of Political Economy* 126, no. 5 (2018): 2072-2107.

Kane, Thomas J., Daniel F. McCaffrey, Trey Miller, and Douglas O. Staiger. 2013. Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment. MET Project: Bill and Melinda Gates Foundation (January).

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 School Principals on academic and behavioral outcomes

	Standard deviation of estimated principal effects				raw outcome mean	raw outcome standard deviation
	(1)	(2)	(3)	(4)		
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 fixed effects	N	Y	N	Y		
Demean by average elementary school principal effect	N	N	Y	Y		

Table 2. Correlations between principal effects on achievement, behavioral, and longer-term outcomes

Correlations between contemporaneous effects on 8th-grade student achievement and long-run effects on 9th-grade achievement and behavior

	9th grade outcomes			
	math score	reading score	absences	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

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***

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***

Notes. Estimates shown include high school fixed effects and demean by the elementary school mean. Statistical significance indicated by *** p<0.01, ** p<0.05, and * p<0.10.

Table 3. Average Principal Value-Added by Student and Teacher Survey Items

Panel A. Student survey items				
	(1)	(2)	(3)	(4)
How safe do you feel…	*Not safe*	*Somewhat safe*	*Mostly safe*	*Very safe*
In your class	-0.0103	-0.0063	-0.0020	0.0072
In the hallways and bathrooms	-0.0105	-0.0057	-0.0005	0.0109
How much do you agree…	*Strongly disagree*	*Disagree*	*Agree*	*Strongly agree*
The topics we are studying are interesting and challenging	0.0027	0.0028	0.0047	0.0106
I work hard to do my best in this class	-0.0074	-0.0046	0.0020	0.0102
Panel B. Teacher survey items				
The principal at this school…	*Strongly disagree*	*Disagree*	*Agree*	*Strongly agree*
Makes teaching expectations clear	-0.0157	-0.0107	0.0007	0.0149
Communicates clear vision for school	-0.0157	-0.0111	0.0021	0.0152
Understands how students learn	-0.0112	-0.0006	0.0033	0.0131
Sets high standards for student learning	-0.0261	-0.0196	-0.0006	0.0168
Presses teachers to implement professional development	-0.0192	-0.0020	0.0015	0.0146
Tracks student academic progress	-0.0260	-0.0112	0.0012	0.0180

Notes: Student and teacher survey data from Chicago Public Schools covering years in which surveys were administered – 1997, 1999, 2001, 2003, 2005, 2007, 2009, 2011, 2012, 2013, 2014, and 2015.

Table 4. The relationship between achievement and teacher and student rating indexes and principal management indexes

	(1)	(2)	(3)	(4)	(5)	(6)
Student safety index	0.036***	0.019***	0.019***	0.140***	0.056***	0.055***
	(0.004)	(0.003)	(0.003)	(0.012)	(0.006)	(0.006)
Student interest index	0.0022	0.013***	0.013***	-0.060***	0.0011	0.00079
	(0.0028)	(0.003)	(0.003)	(0.010)	(0.0042)	(0.0042)
Teacher leadership impression index	0.015***	0.014***	0.014***	0.038***	0.028***	0.027***
	(0.002)	(0.002)	(0.002)	(0.006)	(0.004)	(0.003)
Prior student achievement	X	X	X			
School fixed effects		X	X		X	X
Indicators for first and last years of spells			X			X
Observations	1,086,081	1,086,081	1,086,081	1,488,497	1,488,497	1,488,497

Notes. Coefficients come from student-level regressions of achievement on indexes base on student and teacher responses to the surveys. The indexes are standardized to have a mean of zero and standard deviation of one. Each regression controls for student-level and school-grade-year averages of student race, sex, special education, enrollment, share new students, and parental SES as well as enrollment. Regressions including prior achievement control for a cubic function of lagged achievement. Standard errors clustered by principal spell are in parentheses.
* p<0.10, ** p<0.05, *** p<0.001.