CALDER

National
Center for Analysis of Longitudinal Data in Education Research

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

Urban Institute

A program of research by the Urban Institute with Duke University, Stanford University, University of Florida, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

Estimating Principal Effectiveness

Gregory Branch,
Eric Hanushek,
and Steven Rivkin

WORKING PAPER 32 • DECEMBER 2009

Estimating Principal Effectiveness

Gregory F. Branch
University of Texas at Dallas

Eric A. Hanushek
Stanford University

Steven G. Rivkin
Amherst College

Preliminary Conference Version
Please do not cite without permission from authors

The research presented here is supported by the Packard Humanities Institute and the Smith Richardson Foundation, the Spencer Foundation, and the Hewlett Foundation. The authors are grateful for support from the National Center for Analysis of Longitudinal Data in Education Research (CALDER), supported by Grant R305A060018 to the Urban Institute from the Institute of Education Sciences, U.S. Department of Education.

CALDER, The Urban Institute
2100 M Street N.W., Washington, D.C. 20037
202-261-5739 ♦ www.caldercenter.org

Contents

Abstract 1

Introduction 2

The Texas Database 5

Distribution of Principals 6

Estimation of Principal Effectiveness 8

Empirical Model 8

Test Measurement Issues 9

Differences in Principal Effectiveness 11

Baseline Results 13

Sensitivity Analysis 14

Differences by Transition 15

Teacher Quality and Principal Effectiveness 16

Empirical Model 16

Cumulative Model of Learning 16

Conclusion 18

References 20

Tables 22

Estimating Principal Effectiveness
Gregory F. Branch, Eric A. Hanushek, and Steven G. Rivkin
CALDER Working Paper No. 32
December 2009

Abstract

Much has been written about the importance of school leadership, but there is surprisingly little systematic evidence on this topic. This paper presents preliminary estimates of key elements of the market for school principals, employing rich panel data on principals from Texas State. The consideration of teacher movements across schools suggests that principals follow patterns quite similar to those of teachers – preferring schools that have less demands as indicated by higher income students, higher achieving students, and fewer minority students. Looking at the impact of principals on student achievement, the authors find some small but significant effects of the tenure of a principal in a school. More significant, however, are the estimates of variations in principal effectiveness. The variation in principal effectiveness tends to be largest in high poverty schools, consistent with hypothesis that principal ability is most important in schools serving the most disadvantaged students. Finally, considering principal mobility, the authors find that principals who stay in a school tend to be more effective than those who move to other schools.

Estimating Principal Effectiveness

I. Introduction

School leadership is frequently described as the key element of a high-quality school, and stories of the inspirational and effective principal are plentiful and oft-repeated.[1] However, it is by no means clear that market forces related to the choice of neighborhood and school provide strong incentives for principals to act in ways that foster highly effective schools. Rothstein (2006) discusses a number of potential impediments to such market forces and reports evidence consistent with absence of strong demand for effective schools. Unfortunately, little systematic evidence exists about the importance quantitative importance of principals, making it difficult to sort through alternative policy proposals.

School accountability provides an alternative and more direct incentive structure for schools and administrators that could potentially remedy information failures and strengthen market forces pushing effective schools. Federal accountability statutes in No Child Left Behind insist that all states take testing, accountability, and sanctions to the individual school level, thus elevating the importance of leadership in the individual building. Publicly available reporting on student achievement, particularly in comparison with schools serving similar populations, places additional pressure on principals to raise achievement.

Cullen and Mazzeo (2007) investigate the link between principal salary growth and employment transitions on the one hand and state accountability rating, achievement, and productivity on the other using Texas administrative data. They find a positive relationship between salary on the one hand and accountability rating and student achievement on the other. In addition, they find that principals of more highly rated and higher achieving schools are more likely to persist in their current positions.

These results suggest that a higher accountability rating and higher achievement raises salary and job security, but the possible influences of confounding factors

[1] A large qualitative literature focuses on "effective schools" and in that generally places special emphasis on principals and leadership issues. See, for example, Edmonds (1979), Purkey and Smith (1983), or the case studies in Carter (2000).

including peer composition suggest caution in interpretation of the results. Specifically, the limited set of student, school, and district controls leaves open the possibility that unobserved student or school characteristics contribute to the higher achievement or accountability rating. Consequently, the outcomes may not provide meaningful measures of actual principal effectiveness. It may also be the case that unobserved factors such as family commitment to education raise both school performance and principal pay or job persistence, but the fact that school switchers realize the largest salary increases does suggest that the accountability rating and student achievement do improve labor market outcomes.

We focus on principal value-added and take a somewhat different approach in an effort to identify the variation in principal effectiveness and factors that contribute to that variation. Specifically, we estimate a series of different measures of principal effectiveness using value added models.[2] These include principal fixed effects from models with and without school fixed effects and principal specific gradients linking length of time at a school with student outcomes.

In order to avoid complications introduced by differences in tenure, the samples are limited to observations from the first two or three years of a principal's tenure at a school. Shrinkage methods are also used to mitigate the influences of test error and other sources of measurement error.

In addition to describing the aggregate distribution of principal effectiveness, we also examine differences by student demographic variables in the distribution of principal effectiveness that may derive either from systematic differences in the underlying distributions of principal quality or differences in the magnitude of the effect of principal quality on student outcomes. From case studies and anecdotal accounts, the importance of principals seems most apparent when considering schools serving disadvantaged populations. The demands of schools with poor-performing students including their more difficult working conditions and added difficulty attracting and retaining teachers may

[2] Brewer (1993) and Eberts and Stone (1988) also use panel data to control for student differences by adopting value-added specifications. These studies find evidence that principal quality positively affects achievement, but the possibility that unobserved student or school level characteristics introduce bias remains.

inflate the importance of having an effective leader as compared to the situation in a higher achieving school.[3]

Following the description of the distribution of principal value added we investigate the contribution of teacher quality to differences in principal effectiveness. The composition of the teaching force provides a potentially important channel through which principals affect the quality of instruction. By not rehiring less effective teachers or encouraging them to leave, a successful principal can raise the effectiveness of the teaching force if the replacements outperform the departing teachers. In contrast, a principal that alienates the more effective teachers or hirers poorly may lower the quality of instruction. Because high teacher turnover is associated with both improvement and decline in the quality of instruction, the level of turnover provides little information on the wisdom of principal personnel decisions.

Fortunately, administrative data that link students with teachers provide the opportunity to estimate teacher value-added and therefore the impact of personnel changes on the quality of instruction. These data are available for only a single large district, and the second portion of the empirical analysis examines the correlation between estimates of principal quality on the one hand and the quality of departing and entering teachers during the initial three years of a principal's tenure. If estimates of principal quality actually capture differences in principal effectiveness as opposed to other confounding influences one would expect to find that the quality of instruction tends to improve during a principal's first two or three years at a school.

The limited evidence to date suggests that principal turnover is higher in high poverty schools, implying that students in these schools are much more likely to have a principal with little or no experience (Gates et al. (2006)). This pattern mirrors the findings for teachers and provokes concern that unstable leadership provides an additional obstacle to academic success for many high poverty communities.[4] However, as is the case for teachers, the consequence of turnover depends in large part on the difference in quality between the departing and entering administrators.

[3] Hanushek and Rivkin (2007)

[4] Hanushek, Kain, and Rivkin (2004) and Boyd, Lankford, Loeb, and Wyckoff (2005) provide information on teacher turnover.

The next section of the paper describes the data, and Section 3 presents a description of the distribution of principals by experience, tenure, and student demographic composition. Section 4 provides a conceptual framework for the consideration of principal effectiveness. Section 5 discusses the empirical framework used to measure principal effectiveness and reports our findings on the distribution of principal quality and differences by student demographic composition. The analysis pays particular attention to problems introduced by test measurement error and differences in student characteristics among schools.

II. The Texas Database

The administrative data used in this project were constructed as part of the UTD Texas Schools Project. Working with the Texas Education Agency (TEA), this project has combined different data sources to create matched panel data sets of students and teachers. The panels include all Texas public school teachers, administrators, staff, and students in each year, permitting accurate descriptions of the schools for each principal.

The Public Education Information Management System (PEIMS), 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 can be merged with information on student achievement by campus, grade, and year. Beginning in 1993, the Texas Assessment of Academic Skills (TAAS) was administered each spring to eligible students enrolled in grades three through eight.[5] These criterion 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

[5] 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.

variance equal to one for each grade and year. Thus, our achievement measures describe students in terms of their relative position in the overall state performance distribution.

Because the years of experience in the Texas public schools variable combines both time as a teacher and as an administrator, it is not possible to measure tenure as a principal accurately for those who begin their principal career prior to 1990/91 school year, the initial year of our personnel data. Therefore, for both the descriptive analysis and the achievement modeling we concentrate on the period 1995-2001, and we allocate principals to precise experience and tenure categories in the early career while aggregating experience for six or more years.

The linkage of principal and teacher effectiveness relies on a special subset of data for a single large district in Texas (referred to as the Lone Star District). These data, described and used in Hanushek, Kain, O'Brien, and Rivkin (2005), provide the student-teacher linkages that can be put together with the principal data.

III. Distribution of Principals

This section describes the distribution of principals by demographic characteristics, tenure, and student income, race, ethnicity, and achievement using data for school years 1995 to 2002. This descriptive information forms the backdrop from which to examine principal effectiveness and differences by student characteristics.

Table 1 describes trends over time in the share of all principals and first year principals who are female, black, and Hispanic respectively. Although the shares of each of these groups increase between 1995 and 2002, women show the largest percentage point gains. In 1995 roughly 53 percent of all elementary and middle school principals were women, and that increased to 62 percent by 2002. While slightly more than half of the new principals were women in 1995, almost two thirds of the entering principals were women in 2002. The shares of black and Hispanic principals also increased by more than 10 percent during this period. Between 1995 and 2002 the black principal share increased from 9.8 to 11.2 percent, while the Hispanic principal share rose from 18 percent to 20.5 percent. It appears that growth in enrollment and the number of schools and principals contributed to increases in the shares of principals who are female, black and Hispanic.

Tables 2 describes differences in tenure at the current school by student income, race-ethnicity, and quartile of the mathematics test score distribution. Schools are divided into quartiles on the basis of enrollment shares for each of these characteristics, and the proportions of principals with one, two, three, four, five and six or more years of tenure are reported. Note that the relatively short time frame of the sample prevents additional detail beyond six years, as even the experience variable reported in the administrative data combines years as a principal with years spent in other roles including teacher.

Although high proportion low income schools are more likely than low proportion low income schools to have first year principals and less likely to have principals who have been at the school at least six years, the division of schools by initial achievement produces much larger differences. The proportion of principals in their first year in schools with the lowest average initial achievement is roughly 40 percent higher than in schools with the highest average initial achievement, while the proportion of principals that have been at their current school at least six years is roughly 50 percent higher in the highest achievement schools. Similar differences are smaller when schools are ordered by income and far smaller when schools are ordered by black or Hispanic enrollment shares. The extent to which this captures the duel impacts of factors on achievement and principal turnover or the higher turnover of ineffective principals is not clear.

Table 3 reports principal transitions categorized by destination, new role, and tenure. Roughly 70 percent of principals remain principals in the same school for the subsequent year regardless of tenure. The probability of changing schools and remaining a principal rises from 5.9 percent following the first year at a school to 8.3 percent following the third though fifth years prior to falling back to 5.7 percent for those with tenures of at least six years. By comparison, the probability of transitioning to a non-principal role at a campus declines from 4.5 percent to 2.8 percent as tenure increases from one to at least six years. For those with at least two years of tenure approximately 2 percent transition to work as district administrators. Finally, between 1 in 5 and 1 in 6 principals exit the Texas public schools entirely regardless of tenure category.

IV. Estimation of Principal Effectiveness

Separation of the impact of principals on student achievement from the contributions of various student, school, and district characteristics is complicated by the fact that those factors contributing to achievement likely also make the school more desirable for the typical applicant for a principal position. Because of the impediments to the identification of principal quality, we present a number of different quality measures and examine their relationship with estimated changes in teacher quality following a principal's arrival at a school.

The impact of a principal on school quality likely increases with tenure, and comparisons among principals should account for differences in length of service at a school. Importantly, the impact may become more positive or more negative over time, so simple linear or polynomial controls for tenure are not appropriate. Therefore we estimate separate tenure gradients for each principal in some specifications and restrict the sample to the first k years in a school in others.

Test measurement issues complicate the estimation of principal quality, and we consider both measurement error and the possibility that a focus on basic skills disadvantages principals in schools with larger shares of high achieving students. Below we describe steps taken to address these concerns in the empirical analysis.

A. Empirical Model

Our approach is to embed the analysis of principal effects within a more general achievement model, where panel data techniques relying on multiple observations of individuals and schools are employed to purge the models of systematic but poorly identified influences on student learning. Our basic models relate achievement (A) for student i in school s with principal p in year y as a function of prior achievement, observed student characteristics (X), time varying school and peer characteristics (C), and a vector of principal by school fixed effects. Because of concerns about unmeasured school factors, we also include a school fixed effect (ω) in some specifications, meaning that only within school differences over time identify principal effects. Thus, adding a random error (ε), the empirical model is:

$$(1) \qquad A_{ispy} = \lambda A_{i,y-1} + \beta X_{isy} + \delta C_{sy} + \theta_p + \omega_s + \varepsilon_{ispy}$$

The vector X includes a full set of race/ethnicity indicators and indicators for subsidized lunch eligibility, special education participation, female and English as a second language classification, a switch to the earliest grade offered in a different school (including structural transitions from elementary to middle school), and a switch to other than the earliest grade offered in a new school; and the vector C includes average demographic characteristics for students in school s in year y including proportion low income, proportion classified as special needs, proportion that are recent immigrants and proportion female. All regressions also include a full set of year-by-grade indicators to account for test changes and other statewide policy changes.[6]

The administrative data enable us to account for unobserved school differences through the inclusion of school fixed effects. The school fixed effects absorb time invariant differences in school factors including facilities, attractiveness to teachers, district policies, and peer group composition. Importantly, the fixed effect for a principal in a school where she is the only principal in the sample cannot be identified in a model that includes school fixed effects, as the principal effects in such a model compare principals with others in the same school.

Because the school fixed effects do not account for time varying school factors, even comparisons of school average value added for multiple principals in the same school may fail to produce consistent estimates of within school differences in principal effectiveness. Therefore we consider an additional specification that adds an interaction between tenure at a school and the principal by school fixed effect. In this model the principal by school fixed effects capture all differences across principal spells including fixed principal differences and school factors specific to the spell, and the tenure gradients provide information on changes over time in school quality that we attribute to the principals.

B. Test Measurement Issues

Given the substantial variation in both enrollment and student demographic characteristics among schools, test error and the structure of tests potentially complicate

[6] Prior to running the regressions, the data are aggregated to the campus by grade-by-year level to reduce the computational burden. All tables report absolute values of t-statistics based on robust standard errors clustered by campus.

the measurement of principal quality. Specifically, measurement error in the estimation of principal fixed effects is likely to vary inversely with enrollment, and differences in the underlying distribution of student achievement may affect the translation of principal quality into student achievement. For example, in schools where many students would score near the top at the beginning of the school year, principal quality could have very little effect on standardized test scores even if it is having substantial impact on the overall level of intellectual engagement and quality of instruction.

Measurement error in the principal fixed effect estimates clearly must be addressed. As Kane and Staiger (2002) and Jacob and Lefgren (2005, (2006) point out, even in the absence of confounding influences quality estimates capture both random error and true effects. Consequently variance estimates overstate the actual variation in principal effectiveness, and the magnitude of any upward bias is likely to increase as school size decreases. Following Morrison (1983) we utilize a shrinkage estimator to mitigate the impact of the test error. Normalizing average principal quality to zero, the adjusted quality estimate $\hat{\eta}_s^a$ for principal s in year y equals

$$\hat{\eta}_{sy}^a = \frac{V_{sy}}{V_{sy} + A}\hat{\eta}_{sy}$$

where $\hat{\eta}_{sy}$ is the coefficient on the principal s dummy variable in year y, V_{sy} is the estimated variance of that estimate, and A is the estimated variance of the principal by spell fixed effect distribution. Essentially, the larger the fixed effect error variance for a principal spell the more the adjusted fixed effect is shrunk toward the grand mean.

A related but clearly separate issue concerns the possibility that differences in the underlying distribution of student achievement alter the translation of principal quality into student achievement. For example, principal efforts may have little effect on the TAAS scores of high achieving students who could do very well on this test focused on lower level skills even without attending school. Consequently our test score based estimates of principal quality may produce a more compressed distribution for groups of schools with higher shares of initially high achieving students. Because we are particularly interested in the possibility that principals may have larger effects in schools serving predominantly disadvantaged students who tend to have lower initial scores, this

concern must be addressed in order to produce valid comparisons across schools grouped by poverty rate.

We use two alternative methods to investigate the sensitivity of the estimates to student composition. The first includes a more flexible specification of prior achievement in order to capture differences in expected test score growth by initial score. Although this likely mitigates the problem, even the inclusion of quadratic or cubic terms may not fully address the problem given the skewness of the test score distribution. Therefore our preferred method is to weight observations in all schools and years with fixed weights in order to produce estimates of principal quality not influenced by differences in the test score distribution among schools.

Equation 2 shows the calculation used to produce weighted school by grade by year mean test score:

$$\overline{A}_{sgy} = \frac{1}{N_{sgy}} \sum_{i=1}^{10} \sum_{c=1}^{N_{sgyi}} \frac{F_i}{share_{sgyi}} A_{csgyi}$$

Where F_i is the fixed weight share for decile i, $share_{sgyi}$ is the actual share of students in school s in grade g in year y in decile i, N is enrollment, A is test score, and c indexes child. The weights come from the distribution of 3rd grade mathematics achievement scores in 1994 for students in the top quartile of schools in terms of proportion of students eligible for a subsidized lunch. F_1 is the share in the bottom test score decile, F_2 in the second decile, and on up to F_{10} in the top decile. The scores of students at the low end of the distribution receive disproportionate weight in schools with small shares of such students relative to the weighting sample of higher poverty schools, while the scores of such students receive less than proportional weight in schools with a high concentration of initially low achieving students.

V. Differences in Principal Effectiveness

This section examines the variation in principal effectiveness as measured by value-added to mathematics achievement. We estimate both average principal value added over the first two (or three) years at a school and also the trend in value added over the first two (or three) years at a school. Separation of the impact of principals from the

contributions of unobserved student, school, and district characteristics is complicated by the fact that factors contributing to achievement likely also make the school more desirable for the typical applicant for a principal position. Because of the impediments to the identification of principal quality we present a number of different quality measures and examine their relationships with principal transitions and teacher turnover.

The impact of a principal on school quality likely increases with tenure, and that is why we limit comparisons among principals to those with the same length of service at a school. Importantly, the impact of tenure may become more positive or more negative over time, so simple linear or polynomial controls for tenure are not appropriate. On the one hand, over time a principal would be expected to learn about school operations, the effectiveness of various teachers, and other school specific factors, and such learning would presumably improve job performance. On the other hand, however, principal personnel decisions alter the stock of teachers and the school environment, and the impact of a principal increases over time as a principal accounts for more and more of the hiring and retention of the existing stock of teachers. Thus whether or not the effect is positive depends crucially upon whether the curricular and personnel decisions of the current principal are superior to those of the prior principal.

Following the presentation of the fixed effect estimates we examine the sensitivity of the observed patterns to efforts to mitigate problems introduced by inadequacies of the achievement tests. We begin by illustrating the impact of shrinking the estimates to account for test measurement error and then consider the effects of re-weighting to account for underlying differences in the achievement distribution

The discussion focuses on differences in the distribution of principal effectiveness by school demographic composition but also considers variation by principal transition status. Such differences receive considerable attention, and we describe the variation in effectiveness by quartile of the share of students eligible for a subsidized lunch. Because high poverty, high student turnover schools may confront more difficulties in attracting and retaining teachers and in maintaining discipline, it would not be surprising if principal quality were to have a larger effect in on outcomes in such schools. In addition, underlying differences in the distribution of principal quality may also contribute to any observed variation in principal added. We take advantage of a small sample of principals

who move among schools in different poverty quartiles in an effort to learn more about the contributions of these respective factors to differences by quartile in the dispersion of principal value added.

IVa. Baseline Results

Table 4 reports the mean, variance and 10th, 25th, 50th, 75th, and 90th percentiles of the principal by spell fixed effect distribution by the school poverty rate based on a sample of the first three years at a school. Going down the rows reveals monotonic relationships between the poverty rate on the one hand, and the mean and variance on the other. Specifically, the mean principal by spell fixed effect declines but the variance increases with the poverty share. An examination of the various quantiles shows that the increase in dispersion as the poverty level rises is most pronounced at the lower end of the distribution: The difference between the top and bottom poverty rate quantiles equals -0.20 standard deviations at the 10th percentile, -0.17 standard deviations at the 25th percentile but only 0.07 standard deviations at the 90th percentile.

Because confounding factors may influence the principal by spell fixed effects we also estimate separate returns to tenure for a principal spell at a school in specifications that control for principal by spell fixed effects. The gradient estimate captures the rate of improvement or deterioration of school performance and provides an alternate measure of principal effectiveness for a specification in which the principal by school fixed effect accounts for all other factors affecting achievement.

Table 5 reports the mean, variance and 10th, 25th, 50th, 75th, and 90th percentiles of principal value added as measured by a principal by spell specific return to tenure. In contrast to the principal by spell fixed effects, there is little difference in mean effectiveness, suggesting that other disadvantages at high poverty schools introduced downward bias into the estimate of the principal by spell fixed effects. There is, however, a similar pattern of the variance rising with the poverty rate. Interestingly, the differential at the 90th percentile of 0.10 standard deviations (0.9 to 0.19) between the 90th percentile estimates in the bottom and top poverty rate quantiles is similar to that for the principal by spell fixed effects, while the absence of a sharp difference at the 10th and 25th percentiles contrasts the sizeable gap observed in the principal by spell fixed effects. This pattern suggests that other factors account for what appears to be extremely poor

principal performance in many cases. Henceforth we focus on these gradient estimates as our preferred measures of principal quality.

IVb. Sensitivity Analysis

The aforementioned measurement issues may not only inflate the variance estimates but may also affect the pattern of estimated principal fixed effects by school poverty share. The next table presents estimates that address the two measurement issues, and the subsequent table further divides schools on the basis of size to investigate the pervasiveness of any patterns by school poverty rate.

Table 6 presents three sets of estimates that address these two issues separately and then together. The top panel contains information using the same gradient estimates underlying the statistics reported in Table 5 but shrunk to the grand mean on the basis of the standard error estimates; the second panel reports the mean, variance, and quantiles for estimates of principal specific tenure returns based on reweighted data that eliminates differences in the distribution of initial achievement among principals; finally, the bottom panel reports results based on reweighted data that have been shrunk to remove the effects of test error.

The very close similarity between the results in the top panel and those in Table 5 and between the results in the bottom two panels of Table 6 show that shrinkage has virtually no effect on the estimated distribution of principal quality. Unlike estimates of teacher value added that often rely on fewer than 50 observations for many teachers, estimated value added for principals of even quite small schools typically come from at least several hundred test scores. Consequently the variance of the error is likely to be quite small, and it is not surprising that shrinkage has little effect on the results.

In contrast, there are marked differences among schools in the initial achievement distribution, and this raises the possibility that the greater concentration of test scores at the lower end of the achievement distribution causes the larger variance in principal quality observed for high poverty schools. Re-weighting the scores such that the estimates for all principals are based on the same underlying distribution among test score deciles mitigates this source of bias. Both the middle and bottom panels report results based on the reweighted data, and there is little or no evidence in support of the belief that the larger dispersion in quality observed for the poorest schools comes from

the disproportionate concentration of scores at the lower end of the distribution. To the contrary, a comparison of the statistics in the bottom two panels with those in the top panel reveal that re-weighting increases the variance in all quartiles of the school poverty distribution but does not narrow the differential between the estimated variance in principal quality in the highest poverty schools versus the remaining schools.

Table 7 presents information on the distribution of principal quality based on reweighted and shrunk data by school size, where schools are divided at the median enrollment level. Regardless of school enrollment, the dispersion in principal quality is highest in the high poverty schools, though the differences by poverty rate are larger in the larger schools. In both school size categories the spread rises at both the top and bottom of the distribution in the high poverty schools, consistent with the notion that the variation in principal is larger in high poverty schools.

IVc. Differences by Transition

Many bemoan high rates of teacher and administrator turnover in high poverty schools, but the magnitude of the problem depends in large part on which principals are leaving. In order to gain a better understanding of this issue, we describe the distribution of principal quality by status in the fourth year of the school and share of students eligible for a subsidized lunch. Principals may remain in the same school as principal, remain in the same school in another capacity, take a principal position in another school, take a different position in another school, move to a central administrative position, or exit the public schools entirely. The principal quality measures are shrunk estimates of the gradient in the first three years based on reweighted data.

Tables 8 reports the transition distribution by principal effectiveness quartile based on shrunk principal-spell tenure gradients from reweighted data. With the exception of the lowest poverty schools, principals in the top category in terms of estimated quality are much more likely to remain in the current school. The gap largest gap between the top category and the rest emerges for the high poverty schools, the schools with the largest variation in achievement. In these schools almost three out of four principals in the highest quality category return for a fourth year, while fewer than two thirds return in each of the other three lower categories in terms of estimated value added. Those in the top category are more than five percentage points less likely to exit

the public schools entirely and also much less likely to take a principal position at another school.

One possibility is that Table 8 confuses exits related to school switching with retirement by principals. In order to get at the actual school mobility, we restrict the sample to principals with less than 25 years of experience (Table 9). Again, while there is a little more noise in the pattern, more effective principals tend to stay in their school.

V. Teacher Quality and Principal Effectiveness [NOTE: INCOMPLETE SECTION]

This section examines the relationship between principal quality and changes in the quality of teachers during the principal's first three years in a school. The analysis is limited to one large urban district for which we are able to link students to teachers; the state administrative data do not match students with teachers. We begin with a description of the estimation of teacher value added and then describe differences within schools in the quality of teachers who stay, enter, and leave by the quartile of principal effectiveness. We use two measures of teacher quality, one comparing a teacher to all other teachers in the district and one comparing a teacher to all other teachers in her school.

Va. Estimation of Teacher Value added

I. Empirical Model

The primary analytical task for the estimation of teacher value added is the separation of teacher contributions to achievement from other student, family, school, and community factors. This analysis builds on a cumulative model of learning, and highlights the specific issues relevant to the estimation of teacher fixed effects.

A. Cumulative Model of Learning

We focus on the growth in learning that occurs during a specific grade and relate this to the flows of educational inputs from schools and elsewhere. Equation (1) models achievement of student *i* in grade G and year y (suppressed in the equation since year is unique to grade G for student i) as a function of initial student skill at entry to grade G (α_{iG}), of family background and other influences outside of schools (X), of peer

composition (P), of school factors – including resources, principal quality, and school or district determined curriculum – (S), of teacher quality (τ), and of a random error (e).

$$A_{iG} = \alpha_{iG} + \beta X_{iG} + \lambda P_{iG} + \delta S_{iG} + \tau_{jy} + e_{iG} \quad (3)$$

In the absence of random assignment, unobserved peer and school factors for a given class could confound estimates of the quality of the teacher assigned to that class. The problems can come from a variety of behavioral outcomes: principal assignment of better teachers to classrooms with better students (or worse students, if seeking to equalize achievement across classes); better teachers gravitating toward higher resource schools; families with the most educational concerns and most resources to support children moving to the school districts with the best teachers. All complicate the estimation of teacher value-added to achievement, as teacher quality becomes intertwined with characteristics of students or schools.

The desirability of any particular approach to isolating the value-added of teachers depends upon the extent to which it accounts for the potential confounding factors. Teacher quality is identified only if all potentially confounding factors are included and properly specified as explanatory variables in the regression. Either omission or misspecification of factors that determine α or in *e* corrupts the estimates of teacher quality.

A value-added regression of achievement in grade G on achievement in grade G-1 along with contemporaneous family, school, and peer characteristics and a fixed effect for each teacher in each year provides a natural way to account for prior influences while estimating teacher effects on achievement.[7] Yet although the inclusion of prior achievement mitigates bias from omitted family, neighborhood, and school influences, dynamic behavioral choices by families and school authorities introduce bias even to value added models.

For example, notable sources of "across school" unobserved heterogeneity include the quality of the principal, family background, the extent to which the curriculum for grade G comports with the state test, and the level of student disruption. Because available data typically have limited controls for differences in the quality of

[7] See Hanushek (1979, (1986) for a discussion of value-added models.

administration and other subtle aspects of schools, it is quite difficult to separate teacher and school effects in specifications that produce teacher fixed effects relative to all other teachers in the district. Therefore it is appealing to control for school or even school by grade by year fixed effects in order to account for both observed and unobserved persistent differences among schools and districts, though as noted above this approach prohibits comparisons of teacher quality among schools. Moreover, even estimates of teacher value added based solely on within school variation could suffer from omitted variables bias if classroom assignments are not random.[8] Therefore we produce estimates from specifications that include and specifications that exclude school by grade by year fixed effects.

Vb. Results

VI. Conclusion

An important facet of many school policy discussions is the role of strong leadership, particularly of principals. Leadership is viewed as especially important in revitalizing failing schools. This discussion is, however, largely uninformed by systematic analysis of principals and their impact on student outcomes.

Understanding the impact of principals on learning is a particularly difficult analytical problem. The non-random sorting of principals among schools and consequent difficulty separating the contributions of principals from the influences of peers and other school factors raise questions about the degree to which principals are responsible for differential outcomes.

Panel data on student performance that are linked to principals and schools permit circumventing the most serious difficulties in identifying principal effectiveness. Embedded within a value-added that controls for initial student achievement, we investigate models of principal fixed effects, both with and without school fixed effects, and models of returns to principal tenure at a school. These provide alternative measures of principal effectiveness that deal with different types of potentially confounding influences.

[8] Clotfelter, Ladd, and Vigdor (2006) and Rothstein (2008) document the existence of extensive within-school sorting on the basis of student characteristics and prior performance.

The results suggest the existence of substantial variation in principal effectiveness, particularly in higher poverty and lower achieving schools. In fact the variance estimates for principal effectiveness are roughly twice as large in high as opposed to low poverty schools and in low as opposed to high achieving schools. Allowance for test issues including measurement error and test difficulty does not change these results. These results are consistent with a hypothesis that principal skill is more important in the most challenging schools.

Contrary to commonly held views, more effective principals are less likely to switch districts and are more likely to remain in the same school. This skill-biased moving is particularly prevalent in schools with lower initial achievement.

The initial results suggest that principal movements parallel teacher movements. Specifically, principals are affected by the racial and achievement distribution of students in schools, and this enters into mobility patterns. Yet the common view that the best leave the most needy schools is not supported.

An important element of the role of principals is how they interact with teachers. Our on-going analysis links principals to measures of teacher effectiveness to understand how principals affect teacher outcomes.

References

Boyd, Donald J., Hamilton Lankford, Susanna Loeb, and James H. Wyckoff. 2005. "The draw of home: How teachers' preferences for proximity disadvantage urban schools." *Journal of Policy Analysis and Management* 24 (1): 113–32.

Brewer, Dominic J. 1993. "Principals and student outcomes: Evidence from U.S. high schools." *Economics of Education Review* 12(4): 281–92.

Carter, Samuel Casey. 2000. *No Excuses: Lessons from 21 High-Performing, High-Poverty Schools*. Washington, D.C.: The Heritage Foundation.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2006. "Teacher-student matching and the assessment of teacher effectiveness." *Journal of Human Resources* 41(4): 778–820.

Eberts, Randall W., and Joe A. Stone. 1988. "Student achievement in public schools: Do principals make a difference?" *Economics of Education Review* 7(3): 291–99.

Edmonds, Ronald R. 1979. "Effective schools for the urban poor." *Educational Leadership* 37(1): 15–24.

Gates, Susan M., Jeanne S. Ringel, Lucrecia Santibañez, Cassandra Guarino, Bonnie Ghosh-Dastidar, and Abigail Brown. 2006. "Mobility and turnover among school principals." *Economics of Education Review* 25(3): 289–302.

Hanushek, Eric A. 1979. "Conceptual and empirical issues in the estimation of educational production functions." *Journal of Human Resources* 14(3): 351–88.

———. 1986. "The economics of schooling: Production and efficiency in public schools." *Journal of Economic Literature* 24(3): 1141–77.

Hanushek, Eric A., John F. Kain, Daniel M. O'Brien, and Steve G. Rivkin. 2005. "The market for teacher quality." Working Paper No. 11154. Cambridge, MA: National Bureau of Economic Research.

Hanushek, Eric A., John F. Kain, and Steve G. Rivkin. 2004. "Why public schools lose teachers." *Journal of Human Resources* 39(2): 326–54.

Hanushek, Eric A., and Steve G. Rivkin. 2007. "Pay, working conditions, and teacher quality." *Future of Children* 17(1): 69–86.

Jacob, Brian A., and Lars Lefgren. 2005. "Principals as agents: Subjective performance measurement in education." mimeo John F. Kennedy School of Government.

———. 2006. "When principals rate teachers." *Education Next* 6(2):59–69.

Kane, Thomas J., and Douglas O. Staiger. 2002. "Volatility in school test scores: Implications for test-based accountability systems." In *Brookings Papers on Education Policy 2002*, edited by Diane Ravitch (235–69). Washington, DC: Brookings.

Purkey, Stewart C., and Marshall S. Smith. 1983. "Effective Schools: A Review." *The Elementary School Journal* 83(4): 427–452.

Rothstein, Jesse. 2008. "Teacher quality in educational production: Tracking, decay, and student achievement." Working Paper 14442. Cambridge, MA: National Bureau of Economic Research.

Table 1. The Distribution of Principals by Gender, Race, Ethnicity, and Experience, 1995-2002

	All Principals				First Year Principals		
	Female	Black	Hispanic	N	Female	Black	Hispanic
1995	52.6%	9.8%	18.0%	3,793	57.5%	11.3%	20.8%
1996	54.2%	10.0%	18.6%	3,864	62.4%	14.2%	21.6%
1997	55.7%	10.0%	18.7%	3,965	63.8%	9.6%	19.8%
1998	57.6%	10.3%	19.2%	4,026	63.5%	13.1%	23.3%
1999	59.1%	10.0%	19.6%	4,083	65.0%	8.6%	22.6%
2000	59.6%	10.2%	20.2%	4,158	61.7%	14.3%	20.9%
2001	60.8%	10.9%	20.4%	4,258	65.0%	12.9%	20.9%
2002	61.5%	11.2%	20.5%	4,331	65.2%	13.0%	23.1%

Table 2. Distribution of Principals by Tenure at Current School and Student Demographic Characteristics

		Principal Tenure						
	quartile	1	2	3	4	5	6 or more	
proportion eligible for subsidized lunch								
	bottom	17.8%	15.3%	12.6%	10.2%	8.2%	36.1%	100.0%
	2nd	19.8%	15.9%	12.3%	9.7%	7.9%	34.5%	100.0%
	3rd	20.5%	17.1%	13.7%	9.7%	7.9%	31.1%	100.0%
	top	19.5%	17.1%	13.2%	10.5%	8.2%	31.6%	100.0%
proportion black								
	bottom	20.7%	17.1%	13.5%	10.3%	7.9%	30.6%	100.0%
	2nd	18.5%	15.5%	12.8%	10.2%	7.9%	35.0%	100.0%
	3rd	18.3%	15.9%	12.6%	9.7%	8.6%	35.0%	100.0%
	top	20.1%	16.8%	12.8%	9.9%	7.8%	32.7%	100.0%
proportion Hispanic								
	bottom	19.1%	15.7%	12.2%	9.9%	7.9%	35.2%	100.0%
	2nd	19.2%	15.9%	12.8%	9.5%	7.9%	34.8%	100.0%
	3rd	19.3%	16.3%	12.9%	10.1%	8.4%	33.0%	100.0%
	top	19.9%	17.5%	13.8%	10.6%	8.0%	30.2%	100.0%
Initial Math Achievement								
	bottom	22.7%	19.4%	14.3%	9.8%	7.4%	26.3%	100.0%
	2nd	20.4%	16.7%	12.7%	10.0%	8.4%	31.9%	100.0%
	3rd	18.1%	15.3%	12.5%	10.1%	7.9%	36.3%	100.0%
	top	16.4%	14.0%	12.2%	10.2%	8.5%	38.8%	100.0%

Table 3. Principal Transitions by Tenure

transition	new role	Years of Tenure as Principal at School			
		1	2	3 to 5	6 or more
same campus	Principal	72.5%	68.0%	69.2%	70.8%
change campus, same district	Principal	3.9%	4.9%	6.0%	4.7%
change district	Principal	2.0%	2.5%	2.3%	1.0%
same campus	Other	0.8%	0.6%	0.3%	0.2%
change campus, same district	Other	2.0%	2.1%	2.2%	2.1%
change district	Other	1.7%	1.7%	1.1%	0.5%
same district central office	administration	0.7%	1.3%	1.6%	1.3%
new district central office	administration	0.5%	0.7%	0.7%	0.4%
exit Texas public schools		16.0%	18.1%	16.7%	19.0%
		100.0%	100.0%	100.0%	100.0%

Table 4. Mean, Variance, and Select Quantiles of the Distribution of Principal by Spell Fixed Effects, by the Share of Students in a School That Are Eligible for Subsidized Lunch

			Percentiles				
	Mean	Variance	10th	25th	50th	75th	90th
Share low income quantile							
Bottom	0.025	0.025	-0.180	-0.063	0.032	0.134	0.215
2nd	-0.030	0.031	-0.243	-0.142	-0.032	0.086	0.190
3rd	-0.043	0.043	-0.301	-0.162	-0.036	0.103	0.207
Top	-0.062	0.069	-0.383	-0.236	-0.068	0.114	0.285
All	-0.028	0.043	-0.286	-0.153	-0.019	0.109	0.222

Table 5. Mean, Variance, and Select Quantiles of the Distribution of Principal by Spell Gradients, by the Share of Students in a School That Are Eligible for Subsidized Lunch

			Percentiles				
	Mean	Variance	10th	25th	50th	75th	90th
Share low income quantile							
Bottom	-0.006	0.007	-0.102	-0.049	-0.004	0.043	0.088
2nd	-0.011	0.008	-0.124	-0.071	-0.010	0.047	0.106
3rd	-0.005	0.012	-0.134	-0.071	-0.009	0.062	0.125
Top	0.020	0.020	-0.137	-0.067	0.013	0.099	0.193
All	-0.006	0.007	-0.102	-0.049	-0.004	0.043	0.088

Table 6. Alternative Estimates of the Mean, Variance, and Select Quantiles of the Distribution of Principal by Spell Gradients, by the Share of Students in a School That Are Eligible for Subsidized Lunch and Steps Taken to Mitigate Problems Related to Test Measurement

			Percentiles				
	Mean	Variance	10th	25th	50th	75th	90th
Estimates Shrunk							
Share low income quantile							
Bottom	-0.006	0.007	-0.101	-0.049	-0.004	0.043	0.088
2nd	-0.011	0.008	-0.124	-0.071	-0.010	0.047	0.106
3rd	-0.005	0.011	-0.134	-0.070	-0.009	0.061	0.125
Top	0.020	0.020	-0.136	-0.066	0.013	0.099	0.192
All	0.000	0.012	-0.125	-0.063	-0.003	0.059	0.129
Re-weighted Data							
Share low income quantile							
Bottom	-0.021	0.044	-0.193	-0.078	-0.011	0.064	0.150
2nd	-0.023	0.027	-0.183	-0.107	-0.024	0.058	0.149
3rd	-0.008	0.042	-0.189	-0.096	-0.015	0.061	0.175
Top	0.008	0.088	-0.231	-0.111	-0.002	0.120	0.282
All	-0.011	0.050	-0.199	-0.099	-0.013	0.073	0.188
Shrunk Estimates Based on Re-weighted Data							
Share low income quantile							
Bottom	-0.020	0.042	-0.192	-0.077	-0.011	0.063	0.150
2nd	-0.023	0.026	-0.182	-0.107	-0.024	0.057	0.149
3rd	-0.008	0.041	-0.187	-0.095	-0.015	0.060	0.175
Top	0.008	0.086	-0.230	-0.111	-0.002	0.119	0.277
All	-0.011	0.049	-0.199	-0.099	-0.013	0.073	0.187

Table 7. Mean, Variance, and Select Quantiles of the Distribution of Principal by Spell Gradients based on shrunk estimates from re-weighted data, by the Share of Students in a School That Are Eligible for Subsidized Lunch and School Size

			Percentiles				
	Mean	Variance	10th	25th	50th	75th	90th
Small Schools							
Share low income quantile							
Bottom	-0.027	0.051	-0.228	-0.100	-0.011	0.071	0.173
2nd	-0.019	0.029	-0.191	-0.112	-0.025	0.068	0.175
3rd	-0.020	0.050	-0.212	-0.114	-0.018	0.057	0.168
Top	0.007	0.075	-0.265	-0.136	-0.007	0.122	0.286
All							
Large Schools							
Share low income quantile							
Bottom	-0.016	0.035	-0.164	-0.067	-0.011	0.058	0.123
2nd	-0.027	0.024	-0.170	-0.100	-0.021	0.051	0.110
3rd	0.005	0.031	-0.158	-0.081	-0.011	0.070	0.179
Top	0.008	0.093	-0.217	-0.103	-0.001	0.118	0.259
All	-0.007	0.048	-0.176	-0.084	-0.011	0.071	0.173

Table 8. Principal Transition Distribution by Quartile of the Tenure Gradient Estimate, by the Share of Students in a School That Are Eligible for Subsidized Lunch

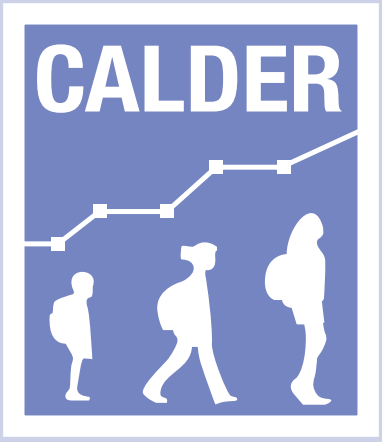
Share low income quartile bottom				
	Quartile of gradient over first three years tenure			
Fourth Year Transition	Q1	Q2	Q3	Q4
Same campus, principal	76.55%	72.61%	60.00%	69.01%
Same campus, other	2.07%	0.00%	0.00%	0.00%
Moves campus, principal	2.76%	4.56%	4.78%	2.11%
Moves campus, other	1.38%	1.24%	2.17%	1.41%
Same district, distr. Admin	0.00%	1.66%	3.48%	2.82%
Moves district, principal	3.45%	0.83%	3.48%	2.82%
Move district, other	2.76%	1.66%	0.87%	0.00%
Move district, district admin	0.69%	0.00%	2.17%	0.70%
Exits	10.34%	17.43%	23.04%	21.13%
	100.00%	100.00%	100.00%	100.00%
Second				
	Quartile of gradient over first three years tenure			
Fourth Year Transition	Q1	Q2	Q3	Q4
Same campus, principal	62.67%	59.02%	64.76%	72.68%
Same campus, other	0.46%	0.55%	0.00%	0.00%
Moves campus, principal	5.07%	7.65%	2.86%	3.61%
Moves campus, other	1.84%	2.19%	1.43%	2.06%
Same district, distr. Admin	2.30%	0.55%	1.90%	1.03%
Moves district, principal	6.45%	0.55%	3.33%	1.03%
Move district, other	2.30%	1.64%	2.86%	0.00%
Move district, district admin	0.46%	1.09%	0.95%	1.03%
Exits	18.43%	26.78%	21.90%	18.56%
	100.00%	100.00%	100.00%	100.00%
Third				
	Quartile of gradient over first three years tenure			
Fourth Year Transition	Q1	Q2	Q3	Q4
Same campus, principal	71.57%	65.48%	62.69%	75.16%
Same campus, other	0.51%	0.00%	0.00%	0.00%
Moves campus, principal	4.06%	5.08%	3.11%	3.18%
Moves campus, other	1.52%	2.03%	0.52%	0.64%
Same district, distr. admin	0.51%	0.51%	2.07%	0.64%
Moves district, principal	4.06%	1.52%	4.66%	1.27%
Move district, other	2.54%	2.03%	2.59%	0.00%
Move district, district admin	1.02%	1.52%	0.00%	0.64%
Exits	14.21%	21.83%	24.35%	18.47%
	100.00%	100.00%	100.00%	100.00%

Top				
	Quartile of gradient over first three years tenure			
Fourth Year Transition	Q1	Q2	Q3	Q4
Same campus, principal	65.31%	60.14%	63.82%	73.76%
Same campus, other	0.00%	0.72%	0.00%	0.00%
Moves campus, principal	8.67%	7.25%	10.53%	6.84%
Moves campus, other	1.53%	2.17%	1.32%	2.28%
Same district, distr. admin	1.02%	0.72%	0.00%	0.00%
Moves district, principal	2.04%	2.90%	2.63%	0.38%
Move district, other	0.00%	0.72%	0.00%	1.14%
Move district, district admin	0.51%	1.45%	0.66%	0.38%
exits	20.92%	23.91%	21.05%	15.21%
	100.00%	100.00%	100.00%	100.00%

Table 9. Principal Transition Distribution by Quartile of the Tenure Gradient Estimate for Principals with Fewer than 25 Years of Total Experience in the Texas Public Schools, by the Share of Students in a School That Are Eligible for Subsidized Lunch

Share low income quartile bottom				
	Quartile of gradient over first three years tenure			
Fourth Year Transition	Q1	Q2	Q3	Q4
Same campus, principal	77.42%	71.03%	65.28%	66.67%
Same campus, other	1.08%	0.93%	0.00%	0.00%
Moves campus, principal	2.15%	3.74%	5.56%	5.56%
Moves campus, other	0.00%	2.80%	1.39%	0.00%
Same district, distr. admin	1.08%	1.87%	1.39%	4.44%
Moves district, principal	3.23%	2.80%	3.47%	2.22%
Move district, other	2.15%	0.93%	1.39%	0.00%
Move district, district admin	0.00%	0.93%	1.39%	2.22%
Exits	12.90%	14.95%	20.14%	18.89%
	100.00%	100.00%	100.00%	100.00%
Second				
	Quartile of gradient over first three years tenure			
Fourth Year Transition	Q1	Q2	Q3	Q4
Same campus, principal	69.29%	71.01%	64.71%	73.00%
Same campus, other	0.79%	0.00%	0.00%	0.00%
Moves campus, principal	3.94%	3.62%	1.96%	3.00%
Moves campus, other	0.00%	2.17%	1.96%	1.00%
Same district, distr. admin	0.79%	0.72%	0.98%	2.00%
Moves district, principal	6.30%	3.62%	3.92%	1.00%
Move district, other	3.94%	0.72%	2.94%	1.00%
Move district, district admin	1.57%	0.00%	0.98%	1.00%
Exits	13.39%	18.12%	22.55%	18.00%
	100.00%	100.00%	100.00%	100.00%
Third				
	Quartile of gradient over first three years tenure			
Fourth Year Transition	Q1	Q2	Q3	Q4
Same campus, principal	51.28%	52.27%	60.34%	55.83%
Same campus, other	0.00%	0.76%	0.00%	0.00%
Moves campus, principal	8.55%	11.36%	4.31%	5.83%
Moves campus, other	2.56%	0.76%	0.86%	3.33%
Same district, distr. admin	4.27%	2.27%	0.86%	0.83%
Moves district, principal	6.84%	4.55%	6.03%	1.67%
Move district, other	2.56%	3.79%	0.00%	0.00%
Move district, district admin	0.85%	0.76%	0.00%	1.67%
Exits	23.08%	23.48%	27.59%	30.83%
	100.00%	100.00%	100.00%	100.00%

Top				
	Quartile of gradient over first three years tenure			
Fourth Year Transition	Q1	Q2	Q3	Q4
Same campus, principal	68.80%	66.67%	61.00%	72.85%
Same campus, other				
Moves campus, principal	7.20%	9.52%	12.00%	5.96%
Moves campus, other	0.00%	2.38%	1.00%	1.32%
Same district, distr. admin	0.80%	0.00%	0.00%	0.00%
Moves district, principal	3.20%	1.19%	3.00%	1.99%
Move district, other	0.00%	2.38%	0.00%	0.66%
Move district, district admin	1.60%	0.00%	2.00%	0.00%
exits	18.40%	17.86%	21.00%	17.22%
	100.00%	100.00%	100.00%	100.00%


CALDER