TN EDUCATION RESEARCH ALLIANCE

Principal Sorting and the Distribution of Principal Quality

Jason A. Grissom, Brendan Bartanen, & Hajime Mitani

Abstract

Numerous studies document the inequitable distribution of teacher quality across schools. We focus instead on the distribution of principal quality, examining how multiple proxies for quality, including experience, teachers' survey assessments of leaders, and rubric-based practice ratings assigned by principals' supervisors, vary by measures of school advantage, using administrative data from Tennessee. By virtually every quality measure, we find that schools serving larger fractions of low-income students, students of color, and low-achieving students are led by less qualified, less effective principals. These patterns persist across urban, suburban, and rural settings. Both differential hiring/placement and differential turnover patterns by principal quality across school characteristics contribute to these patterns. Simulation evidence suggests that hiring and turnover vary in relative importance to principal sorting patterns according to the measure of quality examined, and that differential principal improvement across contexts may matter as well. Complementary analyses of national survey data corroborate our main results.

WORKING PAPER
2019-02

This is a working paper. Working papers are preliminary versions meant for discussion purposes only in order to contribute to ongoing conversations about research and practice. Working papers have not undergone external peer review

Abstract

Numerous studies document the inequitable distribution of teacher quality across schools. We focus instead on the distribution of principal quality, examining how multiple proxies for quality, including experience, teachers' survey assessments of leaders, and rubric-based practice ratings assigned by principals' supervisors, vary by measures of school advantage, using administrative data from Tennessee. By virtually every quality measure, we find that schools serving larger fractions of low-income students, students of color, and low-achieving students are led by less qualified, less effective principals. These patterns persist across urban, suburban, and rural settings. Both differential hiring/placement and differential turnover patterns by principal quality across school characteristics contribute to these patterns. Simulation evidence suggests that hiring and turnover vary in relative importance to principal sorting patterns according to the measure of quality examined, and that differential principal improvement across contexts may matter as well. Complementary analyses of national survey data corroborate our main results.

Principal Sorting and the Distribution of Principal Quality

The quality of a school's leadership is a key determinant of its performance and the opportunities it provides. Aside from higher test scores (Branch, Hanushek, & Rivkin, 2012; Coelli & Green, 2012; Grissom, Kalogrides, & Loeb, 2015), effective principals have been linked to stronger teacher instructional practices, greater teacher morale and satisfaction, reduced teacher turnover, more positive learning climates, higher quality of professional development and coherence of programs, and more positive parental assessments of functioning (e.g., Boyd et al., 2011; Brewer, 1993; Grissom, 2011; Grissom & Loeb, 2011; Ladd, 2011; Sebastian & Allensworth, 2012). Recognition of the link between how effective the school's principal is and school outcomes has led state policymakers to target principal performance in a spate of recent reform efforts, including changes to preparation, licensure requirements, and evaluation (Cheney & Davis, 2011; Clifford & Ross, 2011; Cosner, Tozer, Zavitkovsky, & Whalen, 2015). It has also increased federal attention to principal quality, evidenced most recently in the Every Student Succeeds Act's highlighting of leadership as a school improvement target towards which states and districts could direct federal funds (Herman et al., 2017).

With the observation that high-quality principals matter for school success comes concern that principal quality is not allocated equitably across schools. This concern stems from the robust conclusion from teacher labor market studies that less qualified, lower-performing teachers are systematically found in the schools with the largest numbers of historically marginalized students (e.g., Clotfelter, Ladd, & Vigdor, 2005; Hanushek, Kain, & Rivkin, 2004; Lankford, Loeb, & Wyckoff, 2002). Studies generally assert that because working conditions are poorer, teaching positions in those schools are less desirable, resulting in greater difficulties recruiting qualified teachers, higher turnover, and systematic migration of high-quality teachers towards higher achieving schools (Boyd, Lankford, Loeb, & Wyckoff, 2005; Guarino, Santibañez, & Daley, 2006), though it is possible that race and class biases affect mobility as well. To the degree that principal

sorting mirrors teacher sorting, students from marginalized backgrounds will have less access to the kind of high-quality leadership that can sustain school improvement, likely reinforcing disparities in students' opportunities to learn among schools at opposite ends of the socioeconomic spectrum.

Although it makes sense that principal sorting would be similar to teacher sorting—nearly all principals are former teachers, after all, and likely have many similar work preferences—institutional differences between teacher and principal labor markets may disrupt this alignment. For example, in many districts, teacher hiring decisions are decentralized to the school level, potentially creating competition for good teachers among schools even in the same district. Uniform salary schedules, which prevent districts from differentiating teacher pay across schools, mean that schools with higher non-pecuniary benefits (i.e., working conditions) are in a better position to compete for high-quality teachers. Moreover, collective bargaining agreements typically give districts minimal power to dismiss or transfer teachers (Cohen-Vogel & Osborne-Lampkin, 2007). These factors combine to limit districts from acting strategically with regard to the allocation of teachers across schools. In contrast, principals are more akin to "middle managers" in the larger district bureaucracy (Morris, Crowson, Hurwitz, & Porter, 1982), and districts have greater control over where they are placed. Principal hiring decisions typically are centralized, so principals seeking alternative employment as an administrator will have fewer local options. Few districts collectively bargain principal contracts, meaning that districts have greater freedom to move principals to different schools.[1] Also, because salary schedules are much less common for principals than for teachers, districts may have greater power to compensate principals for moving them to more difficult schools.[2] In short, patterns of principal sorting need not be as inequitable as those for teachers, particularly within school

[1] Only 19% of school districts report collective bargaining or meet-and-confer discussions with principal associations (source: authors' calculations from the 2011–12 Schools and Staffing Survey).

[2] Approximately 47% of school districts report using a salary schedule for principals, whereas 89% use one for teachers (source: author's calculations from the 2011–12 Schools and Staffing Survey).

districts, where district administrators theoretically have greater capacity to place high-quality principals into needier schools.

The question of whether principal quality, like teacher quality, is distributed inequitably is an empirical one. Unfortunately, research on principal sorting (or the principal labor market more generally) is sparse (Clotfelter, Ladd, Vigdor, & Wheeler, 2006; Loeb, Kalogrides, & Horng, 2010). We contribute to this small literature by examining the hiring and retention of public school principals using two data sources. The first is a decade-long longitudinal administrative data set from Tennessee that allows us to observe the distribution and sorting of principal quality measures across schools with different characteristics over time. The second is nationally representative data on principals from the Schools and Staffing Survey (SASS), with which we assess the degree to which the patterns of sorting observed in the Tennessee data hold more generally. These data allow us to extend analysis beyond the urban districts that often have been the focus of prior work to make comparisons to principal labor market patterns in suburban and rural districts.

Using these two rich data sources, we ask the following main research questions. First, how are measures of principal quality distributed across schools with different concentrations of traditionally marginalized students, including low-income students, students of color, and low-achieving students? Second, to what degree do differences in the kinds of principals hired into schools educating different groups of students explain principal sorting patterns? Finally, to what degree does differential principal turnover contribute to gaps in principal quality among different types of schools?

Sorting among Educators

A substantial literature documents the inequitable distribution of educator quality across schools. Most of this work focuses on the distribution of teachers, showing a robust pattern that schools with larger proportions of students of color, students from low-income

backgrounds, and low-achieving students tend to be staffed with teachers with lower qualifications, including less teaching experience and lower likelihoods of full certification (Clotfelter et al., 2005; Goldhaber, Lavery, & Theobald, 2015; Lankford et al., 2002). More recently, the growth of student-level data systems has allowed researchers to move beyond qualifications measures to examine the distribution of outcome-based measures of teacher performance, such as value-added, as well (Glazerman & Max, 2011; Goldhaber et al., 2015; Goldhaber, Walch, & Gabele, 2014; Isenberg et al., 2013; Sass, Hannaway, Xu, Figlio, & Feng, 2012). This line of work reports similar overall patterns: students from marginalized backgrounds have less access to high-performing teachers.

In contrast, surprisingly few studies have systematically explored the distribution of principal quality, though some evidence suggests similar distributional patterns with respect to principal qualifications. For example, in North Carolina in the 1990s and early 2000s, the average licensure-related test scores among principals serving the highest-poverty schools were in the range of 0.3 to 0.5 standard deviations lower than those of principals serving the lowest-poverty schools (Clotfelter et al., 2006). Other principal qualifications, such as length of tenure, were also unequally distributed. Similarly, principals of Texas schools with larger low-income populations and lower math achievement had less experience in that school (Branch et al., 2012). In Miami, schools with a large number of marginalized students tended to have principals with less experience, less education, and an undergraduate degree from a less competitive institution (Loeb et al., 2010).

Less evidence exists on the distribution of job performance-related measures.[3] Conceptual and empirical challenges make accurate measures of principals' value-added to student test scores, a common outcomes-based measure for teachers, difficult to estimate (Grissom et al., 2015). A few studies look instead at the distribution of low-stakes assessments of principal leadership. For example, studies that use teacher surveys to

[3] The distinction between qualifications and job performance is important because the two may not be strongly correlated.

quantify the quality of leadership in a school tend to find negative correlations with the fractions of Black students, Hispanic students, and/or students receiving free or reduced-price lunch (Boyd et al., 2011; Clotfelter et al., 2006; Grissom, 2011; Ladd, 2011).

The mechanisms driving these patterns have been left underexplored. That is, it is unclear to what extent the inequitable distribution of principal quality arises from patterns across schools in principal placement, turnover, or sorting across schools over the principal career. Many principals express preferences for working in schools that are high-achieving and have fewer low-income students, perhaps because they also express preferences for other school factors correlated with student characteristics, such as parent participation, resource availability, and school safety (Loeb et al., 2010). These preferences can affect hiring by limiting the pool of qualified principal candidates for vacancies in such schools and affect turnover by making it more likely that a qualified principal seeks employment in schools more aligned with his or her preferred working conditions. At the same time, district leaders may mitigate these impacts by, for example, prioritizing hiring principals for lower-income or lower-achieving schools who are both highly qualified and committed to working in a more challenging school environment, or targeting compensation or other resources towards retaining principals in those schools. Because principals are in middle management positions under more direct control of central leadership and less likely to be protected by collectively bargained contracts, principals may be more constrained than teachers in moving to schools with their preferred characteristics.

Prior work finds that the characteristics of principals hired differ by the concentration of student poverty in the school. For example, in their study of North Carolina, Clotfelter et al. (2006) find that brand-new principals account for 67% of principals hired into schools with a majority of low-income students, compared to just 60% for other schools. They also conclude that the concentration of less qualified principals in high-poverty schools is driven largely by patterns of entry, with novice principals in those schools having systematically lower qualifications than their colleagues in low-poverty schools. The authors hypothesize

that these differences may reflect the generally lower qualifications of the teaching pool in those schools from which administrators systematically are drawn. Loeb et al. (2010) similarly find that high-poverty schools in Miami are much less likely than low-poverty schools to fill a principal vacancy with an experienced principal. They find, however, much less stark patterns in principal qualifications among novice principals than the North Carolina study. First-time principals in their data in high- and low-poverty schools had similar levels of total experience in the district and similar educational qualifications.

Higher rates of turnover can also contribute to principal quality disparities. Several studies suggest that principals are more likely to leave schools with large numbers of low-income, low-achieving, or racial/ethnic minority students (Clotfelter et al., 2006; Fuller & Young, 2009; Gates et al., 2006; Grissom & Bartanen, 2018). Higher turnover rates mean lower principal tenure in the school and more frequent vacancies, which can further exacerbate quality differences resulting from differential principal hiring in schools with larger traditionally marginalized populations. That is, if principal job performance improves with experience (Clark, Martorell, & Rockoff, 2009), then frequent turnover means systematic replacement of outgoing principals with new principals who are less likely to be effective.

Building on prior research, we begin by documenting patterns in the distribution of principal quality in Tennessee. We then use the 2011–12 Schools and Staffing Survey to explore (to the extent possible) the degree to which these patterns hold nationally. The Tennessee data offer two main advantages in examining the distribution of principal quality over what has been possible in prior work. First, we can use the Tennessee data to construct more comprehensive measures of principal qualifications from principals' prior job histories. This is particularly important for examining hiring disparities, as many new-to-school principals do not have prior principal experience. Second, the Tennessee data permit examination of the distribution of quality using not only qualifications measures and the kinds of survey-based teacher reports of leadership used in prior work

but with measures of effectiveness from the statewide principal evaluation system, such as supervisors' rubric-based ratings of principal practice. Standardized performance ratings can provide a more explicit accounting of gaps across schools based on actual leadership practice, which may not correlate strongly with qualifications. While more limited in the measures it makes available, SASS allows for a nationally representative look at principal quality measures.

Data and Measures

Our main analyses draw on longitudinal administrative personnel data provided by the Tennessee Department of Education. Tennessee is a relatively large and diverse state, operating approximately 1,800 schools in 146 districts that serve 996,000 students, 31% of whom are Black or Hispanic and 58% of whom are eligible for subsidized lunches.[4] The personnel files provide principal job history data from 2002 to 2017. These files include work location, highest degree attained, age, and number of years of experience in Tennessee schools. We match these data to school files that contain annual information on schools' racial/ethnic composition, free/reduced lunch rates, and average performance on the state's standardized math and reading achievement tests[5], as well as information from the Common Core of Data (CCD) about school locale type (urban, suburban, town/rural[6]). This latter information is only available beginning in 2007, so we focus our analysis on the

[4] https://www.tn.gov/education/topic/report-card

[5] From 2006–07 to 2014–15, achievement scores come from the Tennessee Comprehensive Assessment Program, or TCAP, includes math, reading, science, and social studies exams for students in grades 3–8, as well as end of course exams in various high school subjects. In 2015–16, the state switched to a new testing program, called TNReady. To construct a measure of average school achievement, we first standardize each grade-test score by year, then aggregate the student-level scores into a school average. For K–8 schools, the achievement index only includes math and reading scores. For high schools, we include end of course scores for Algebra I, Algebra II, English I, English II, and English III. Due to logistical challenges in implementing TNReady, the state cancelled testing for grades 3–8 in 2015–16. To avoid dropping principals in this year, we impute the achievement index by averaging the scores from 2014–15 and 2016–17. However, all of our findings are robust to simply excluding these principals.

[6] These locale distinctions admittedly can mask important variation within categories.

period from 2007 to 2017.[7] In each year, we identify approximately 1,700 principals. Descriptive statistics for principals and schools are shown in Table 1.

Measuring Principal Quality in Tennessee

The key empirical challenge to investigating the distribution of principal quality across schools is that direct measures are difficult to come by. Fortunately, Tennessee has invested in creating multiple measures of principal effectiveness, two of which we use in this analysis. Additionally, we construct a variety of plausible proxies for principal quality that have been used in prior work.

The first measure comes from the Tennessee Educator Acceleration Model (TEAM) for the 2011–12 through 2016–17 school years. TEAM is the state's educator evaluation system. Fifty percent of the TEAM evaluation for principals comes from ratings of principal performance on a rubric pegged to the Tennessee Instructional Leadership Standards.[8] These ratings are based on formal observations conducted by the principal's supervisor. In this analysis, we use principals' average yearly observation scores—the exact measure used by the state to calculate summative evaluation ratings.[9] We refer to this measure as "supervisor ratings".

A potential concern with rubric-based observation scores is that they conflate differences in effectiveness with differences in school context or differences in how principal performance is judged by the district. One approach to mitigate this bias is to

[7] We use data going back to 2002 to calculate measures of principal experience and length of tenure in current school.

[8] For more information about TEAM, see http://team-tn.org/evaluation/administrator-evaluation/

[9] Prior work shows that principals' ratings across indicators are highly inter-related and can be reduced to a single underlying performance score using factor analysis (Grissom, Blissett, & Mitani, 2018b). Using the average observation score instead of the factor score described in Grissom et al. (2018b) allows us to include principals in districts that used alternative observation rubrics (approximately one-quarter of principals in the state), as these districts do not report domain-specific scores for principals. However, for principals for whom we can calculate factor scores, the average observation score and the factor score are correlated at 0.95 or higher each year. Beginning in 2011–12, we can access these ratings for 90% of principals.

"residualize" the scores—i.e., regress them on school characteristics and compute the residuals. This procedure mechanically removes any correlation between ratings and the contextual characteristics included in the model. However, to the extent that there are *true* differences in principal quality by school context (e.g., between high- and low-poverty schools), this type of residualization will over-correct for bias from contextual differences.[10] Instead, we residualize supervisor ratings on district fixed effects, which forces the average score in each district to be zero. We also estimate a handful of multivariate models that regress the unadjusted scores on school contextual variables and district fixed effects, which limits the identification of principal quality gaps to schools within the same district.[11]

Additionally, we use low-stakes survey responses of teachers that assess their school's leadership. The responses are from the Tennessee Educator Survey, a yearly statewide survey of teachers jointly administered by the Tennessee Education Research Alliance and the Tennessee Department of Education.[12] In the first three years of the survey, teachers were randomly assigned to respond to different modules, one of which contained a set of questions evaluating their principal's leadership. Items ask, for example, whether the school's principal consistently monitors student academic progress, communicates a clear vision for the school, or sets high standards for teaching. Beginning in 2014–15, the survey was redesigned to administer these leadership items to all teachers. We found that responses measured a single underlying latent construct of principal performance, so we compute the standardized factor score, which we refer to as "teacher ratings."[13]

---

[10] One potential solution is to include principal fixed effects in the residualization step, similar to models that have been used to estimate teacher value-added (e.g. Chetty, Friedman, & Rockoff, 2014). However, successful identification in these models requires sufficient within-person variation in school characteristics. Due to our short panel (supervisor ratings start in 2011–12) and the fact that principals tend to move among similar groups of schools, these models likely rely on variation that is idiosyncratic.

[11] We also explored a two-step residualization process that takes the district-adjusted scores and residualizes them on school characteristics and principal fixed effects. Scores from this approach are highly correlated ($r = 0.96$) with the district-adjusted scores.

[12] For information about the survey, see https://www.tn.gov/education/data/educator-survey.html

[13] Appendix B shows, for each year, the questions used to construct the score and descriptive statistics for the factor analysis. Note that the some of the survey items used to generate the factor score change from

In addition to these direct measures of principal quality, we examine several principal qualifications measures that are plausible proxies for principal quality. The first of these measures is principal experience. Research suggests that school performance is lower under novice principals and that principal effectiveness increases as they gain experience (Béteille, Kalogrides, & Loeb, 2012; Clark et al., 2009; Dhuey & Smith, 2014). We thus create two measures of novice principal: an indicator for the principal being in his or her first year as a principal and an indicator for being in the first three years. Across years, 11% of Tennessee principals are in their first year, on average, and 33% are in their first three years (see Table 1).[14] Prior work also suggests that school performance is lower following administrative turnover (Béteille et al., 2012; Miller, 2013), so we create a variable for first year in school and first three years in school for principals beginning in a new school regardless of whether they have prior experience as a principal elsewhere. Eighteen percent of principals are in their first year as principal in the school, and 47% are in their first three years, on average.

The next set of measures is based on principals' educational experiences. We create an indicator for holding an education specialist degree or doctorate as one's highest degree. Evidence on the link between degree attainment and principal performance is minimal, though at least one study found that principals with specialist and doctoral degrees were more likely to engage in management behaviors associated with greater student learning gains (Grissom & Loeb, 2011). Forty percent of Tennessee principals hold one of these degrees. Also, for principals seeking initial certification since the 2003–04 school year

year to year. Given that we found strong evidence of a single underlying factor in each year, we chose to retain items that did not appear on previous surveys but were relevant to evaluating principal performance and school culture. Response rates among teachers for each of the six years of the survey are (starting from 2012): 24.8%, 38.7%, 41.9%, 55.3%, 48.1%, 56.2%. The percentages of principals for whom we can construct these ratings are (in order of year) 68, 72, 80, 97, 96, and 97. Unlike with supervisor ratings, missingness is not correlated with whether the principal turned over at the end of the year. However, missingness is somewhat greater among high-poverty and low-achievement schools.

[14] The administrative data file does not have a variable that indicates which year the employee became a principal for the first time. We coded *first year* as the first year the employee was observed as a principal in the longitudinal personnel file, beginning in 2002. Since our analysis begins in 2007, we should be able to accurately identify novice principals.

(when the state first required the test), we obtained School Leaders Licensure Assessment (SLLA) scores from the Educational Testing Service (ETS) and matched them to the personnel file.[15] The SLLA is a test of knowledge believed necessary for competent professional practice, and is aligned to the Interstate School Leaders Licensure Consortium leadership standards. In total, we have SLLA scores for 25% of the sample, though we found that rates of missingness for SLLA scores were very similar across each of our categories of school context.[16]

The Unequal Distribution of Principal Quality across Schools

We begin by examining the distribution of principal qualifications and effectiveness across different categories of Tennessee schools. Specifically, Table 2 categorizes schools according to four proxy measures of societal (dis)advantage: average test score performance (levels), student poverty, percentage of students of color in the school, and locale type (i.e., urban, suburban, town, rural). The achievement index, which is the average standardized student-weighted score for math and reading in the school, is shown by quintile: lowest quintile, middle 60%, and highest quintile. Student poverty is split into three categories by the fraction of the school's student body that is eligible for free or reduced-price lunch: less than 20%, 20–80%, and more than 80%.[17] We refer to these groups as "low-poverty", "medium-poverty", and "high-poverty," respectively. The percentage of students of color is similarly broken into three groups (0–20%, 20–80%, 80–100%). For each quality measure (columns), we conduct significance tests for the difference between the first and second/third school group. Specifically, the p-values refer to the coefficients from a regression model where the relatively "advantaged" school group (highest quintile of achievement, 0–20% FRPL, 0–20% students of color, suburban locale) is the omitted

[15] This matching is described in Grissom, Mitani, and Blissett (2017).

[16] More information about the SLLA can be found at https://www.ets.org/sls/

[17] Approximately 10% of school-by-year observations fall into the first group, 70% in the second group, and 20% in the third group. Using quintiles of student poverty leads to qualitatively similar conclusions.

category.[18]

Looking across measures of principal quality, a clear, consistent pattern emerges. As in prior studies of the distribution of teacher qualifications (e.g., Lankford et al., 2002), schools with large proportions of low-achieving students, low-income students, and students of color are more likely to employ inexperienced principals, principals who are new to that school, principals who obtained lower scores on the SLLA, and principals who were rated lower by their supervisors and teachers. For example, 13% of schools in the lowest achievement quintile employ a first-year principal, compared to only 9% of schools in the highest quintile. Low-achievement schools are similarly more likely than the highest-achievement schools to employ a principal who is new to the school (23% to 15%). Principals in the lowest-achieving schools also score 3.7 points lower on the SLLA than principals working in low-poverty schools, which equates to 44% of a standard deviation in the SLLA distribution, a large difference.

Figure 1 further illustrates the disparities in principal experience and tenure. For each individual panel, the breakdown by school characteristics (school achievement in Panel A, student poverty in Panel B, and nonwhite students in Panel C) is shown overall and within locale groups (urban, suburban, town/rural). Importantly, patterns tend to hold across locale types, despite their differences in labor markets and distributions of student characteristics. Schools attended by larger numbers of marginalized students in all geographic contexts have more inexperienced principals.

Table 2 shows that principals in low-achievement schools are, on average, more than one-third of a standard deviation below average in terms of effectiveness as measured by supervisor ratings, and the gap between principals in low-achievement versus high-achievement schools is an astounding 0.88 SD ($p < 0.01$), or about half a rating point. Similar disparities exist between schools across poverty, race/ethnicity, and locale.

[18] Although we have population-level data for many of the quality measures (experience, tenure, education), we conduct hypothesis testing for all measures to maintain consistency throughout the text.

Comparing high- and low-poverty schools, the disparity in supervisor ratings is more than a full standard deviation. As an additional check, we examine "adjusted" scores that rely on within-district variation in school characteristics. Here, we still see that there are substantial differences between low-achievement and high-achievement schools, high-poverty and low-poverty schools, and schools serving low and high percentages of students of color, though the magnitude of the disparities is smaller. Additionally, there are no differences in adjusted scores by locale type, which makes sense given that there is very little variation in school locale within districts. That we still find disparities in ratings even after adjusting for between-district differences further suggests true differences in average principal quality between schools serving higher and lower numbers of marginalized students.[19]

When effectiveness is measured using teacher perceptions of leadership performance (teacher ratings), principals in low-poverty and high-achievement schools are more effective than principals in high-poverty and low-achievement schools ($p < 0.01$), though the disparities are smaller in magnitude. Similarly, schools in urban areas and with large numbers of students of color tend to have lower-rated principals.

Figure 2 shows meaningful disparities in terms of the full distribution of supervisor (panel A) and teacher ratings (panel B). The vertical lines on each plot show the mean differences from Table 2. One possibility is that the mean differences in ratings are driven by the propensity to have more or fewer very low or very high scoring principals. Figure 2 demonstrates that this is largely not the case—the distribution of ratings has a similar bell shape across school contextual categories. The one clear exception to this pattern is supervisor ratings in low-poverty schools, where there are a large number of principals who have nearly perfect supervisor ratings and almost no principals who score more than one

[19] If we run a random effects model that partitions variance in supervisor ratings into between- and within-district, 32% of the variance in scores is between districts. We can also add an additional random effect for principal, which further partitions variance. Here, 29% is between districts, 32% is between principals in the same district, and 39% is within principal. These patterns suggest to us that disparities may indeed be amenable to mitigation by district actors.

standard deviation below the mean.[20]

Studies demonstrate that sorting patterns for teachers systematically disadvantage urban schools (e.g., Lankford et al., 2002). Table 2 shows that town/rural schools in Tennessee face principal quality deficits that are as large or larger. For instance, town/rural principals have the lowest scores on the SLLA, and are more likely to employ inexperienced and new-to-school principals. As in studies of teachers, suburban schools show a consistent pattern of advantage in most measures of leadership quality.

Finally, one question raised by Table 2 is the extent to which the descriptive disparities in principal quality by different measures of school context are driven by one contextual measure that happens to be correlated with the others. To explore this possibility, we estimate multivariate models for a selected set of principal quality measures. By including all of the school contextual categories in a single model, we can see whether one is relatively more important in explaining disparities. As shown in Appendix Table A1, the disparities in principal quality are most clearly tied to differences in the average achievement level of the school, though for some measures poverty maintains a correlation even after achievement is accounted for.

Disentangling Sorting: Hiring and Turnover

Inequitable sorting of principals by school characteristics can occur in two main ways: schools attended by higher concentrations of marginalized students can tend to hire less qualified and less effective principals, or they can be less likely to retain high-quality principals when they hire them. This section examines these two mechanisms. Differentiating between hiring and turnover helps illuminate what policy approaches might be useful for reducing the quality gaps.

[20] Interestingly, we find no such pattern for ratings from teachers, though low-poverty schools still have higher-rated principals, on average.

Quality of New Hires

First, we examine the characteristics of newly hired principals in Tennessee. That is, we summarize the characteristics of principals in their first year in a given school, regardless of whether they have prior principal experience. Means for all new hires and for new hires broken down by school characteristics are shown in Table 3. The first columns show different types of prior administrative experience. The last two columns show the proportion of new hires with an education specialist or doctoral degree and the average SLLA score.

The average new hire in a high-achievement school has 5.3 years of prior experience as a principal or AP, compared to 4.6 years for new hires in low-achievement schools. This gap reflects longer preparation as an AP (3.3 years vs. 2.7 years) rather than more prior experience as a principal. There also are large differences between high-/medium-achievement (28%) and low-achievement (12%) schools in whether the new hire served as an AP in the same school. We also find that low-achievement schools are less likely to hire (or be able to hire) principals with an Ed.S./doctorate degree and that their new hires have lower average SLLA scores. Patterns are similar by student poverty, with differences between high- and low-poverty schools in AP experience even more pronounced.

Disparities in hiring among schools serving different proportions of students of color are less clear, due to different hiring patterns in urban/suburban versus town/rural schools. Across almost all measures of administrator experience, schools in town/rural settings hire less qualified principals. The average newly hired principal in a town/rural school has roughly 1.3 years of prior principal experience, compared to two years in urban/suburban schools. The same disparity exists for AP experience; roughly 80% of newly hired principals in urban/suburban schools have AP experience, compared to roughly 66% in town/rural schools. One reason for the large disparity in AP experience across locales is that town/rural schools tend to have smaller schools with fewer administrators, providing aspiring principals fewer opportunities to serve as APs. Additionally, while town/rural

schools have new hires with lower average SLLA scores, they are slightly more likely to have a terminal degree, mirroring a pattern in Table 2.

While observable characteristics such as prior experience help illustrate hiring disparities among schools with different characteristics, we are ultimately interested in comparing the effectiveness of these new hires. Supervisor ratings follow similar patterns as found in Table 2. There are stark differences across different categories of school advantage. New hires in high-achievement and low-poverty schools have above-average supervisor ratings (0.03 and 0.29 SD), compared to new hires in low-achievement and high-poverty schools, where ratings are well below average (-0.63 and -0.63, respectively). As before, we find an advantage for suburban schools (-0.27) relative to town/rural (-0.38) and urban (-0.49) schools. Teacher survey measures show similar patterns.

One challenge in examining the supervisor and teacher ratings from the first year in a principal's school is that their distributions may not be clean to interpret if they are biased by school characteristics. Again we can compare the adjusted ratings, which account for differences in average ratings between districts. Using these adjusted ratings narrows the gaps compared to the first column; however, our substantive findings remain the same.

As as an additional measure, we examine the ratings of new-to-school principals from their prior job, if they were working as an administrator. These ratings may be affected by the characteristics of the prior school, but should nonetheless reflect the district's assessment of the leader's effectiveness (and, presumably, the best available evidence about the leader's effectiveness at the time he or she was hired into the new school). Table 4 reports the results.[21] The average new hire with prior principal experience has a below-average supervisor rating (-0.23). However, more effective principals sort to more advantaged schools. For example, the average of prior-year ratings among newly hired principals in high-achievement schools is 0.13, compared to -0.47 in low-achievement

[21] There is no prior teacher rating for former APs, since questionnaire items do not include AP-specific questions.

schools. Comparing the adjusted prior ratings, the gap shrinks to roughly 0.25 SD but is still statistically significant at a 95% confidence level. Similar differences exist when we classify schools by student poverty, race/ethnicity, and locale. Disparities in prior ratings of new hires are similar in magnitude to the disparities in first-year scores, supporting the contention that disparities are not completely driven by a tendency for principals in advantaged schools to receive higher evaluation scores.[22] Finally, among new hires with supervisor ratings as an AP, the difference between low-achievement and high-achievement schools is more than half a standard deviation, with an even larger disparity between high-poverty and low-poverty schools (0.79 SD).

Principal Turnover

We turn to principal turnover in Tennessee to investigate the second possible reason for principal sorting. Table 5 summarizes the proportions of principals who leave their positions (binary turnover) in addition to five differentiated turnover categories: (1) transfer to a different school in the same district; (2) transfer to a school in a different district; (3) move to a central office position; (4) move to a non-principal, school-level position; and (5) leave the education system. On average, 18% of Tennessee principals leave their positions each year. Four percent move to another school in the district, less than 1% move to another school in a different district, 3% are promoted to a central office position (e.g., instructional supervisor, superintendent), 3% are demoted to a school-based position (e.g., assistant principal, teacher), and 7% are no longer working in the Tennessee public education system.

Schools with larger marginalized populations systematically face higher turnover rates. The largest gap is between low-achievement and high-achievement schools (23% vs. 14% turnover rate). Additionally, 22% of principals at high-poverty schools do not return

[22] While traditionally advantaged schools also tend to hire principals with higher teacher ratings in their prior schools, these differences are smaller in magnitude and only statistically significant when comparing schools by the percentage of students of color.

as principals the following year, compared to 17% at low-poverty schools. Urban schools (21%) have substantially higher turnover rates than suburban (16%) and town/rural (17%) schools. Examining specific types of turnover, we find that principals in low-achievement, high-poverty, more nonwhite, and urban schools are the most likely to transfer within the district (6%). Approximately one-third of principal turnover cases involve position changes—to either central office or lower school-level positions. Principals working in low-achievement and high-poverty schools have the highest rates of demotion and are the most likely to exit the education system.

Figure 3 breaks down principal turnover rates by school characteristics within locale groups. Across all three panels, patterns for both urban and town/rural schools tend to mirror the overall pattern: substantially higher turnover in the least advantaged schools. Suburban schools follow this same pattern with respect to achievement, but their rates of turnover between traditionally advantaged and disadvantaged schools are more similar for the other two measures.[23]

Simulating the Principal Quality Gap

While the previous section documents substantial hiring and turnover differences across schools according to measures of marginalization, we have not yet identified the degree to which each of these mechanisms drives the principal quality gap in Tennessee. Next we conduct a simulation to uncover the relative importance of these components. We draw on a framework described in Goldhaber, Quince, and Theobald (2018), which examines teacher quality gaps in North Carolina and Washington. Here we provide a basic description and results; Appendix C contains the details of the simulation (including all of the parameters used). The intuition of our analysis is to use the observed rates of hiring

[23] In other work, we explore the relationship between principal turnover and measures of principal quality (Grissom & Bartanen, 2018). This analysis finds that less effective principals are much more likely to turn over, even conditioning on school and other individual characteristics. Less effective principals are especially more likely to be demoted or to exit the education system, and also somewhat more likely to move to other districts.

and turnover to simulate the quality gap over time, beginning from an arbitrary equal distribution of principal quality. Starting from an assumed equal distribution allows us to examine the extent to which differential rates of hiring and turnover contribute to the unequal distribution of principal quality.

To be specific, we conduct separate simulations that examine gaps among high-achievement, middle-achievement, and low-achievement schools for two measures of principal quality: principal experience and supervisor ratings. For principal experience, we split principals into three groups: 0–2 years, 3–5 years, and 6 or more years of prior principal experience. For supervisor ratings, placement is determined by the quartile ranking of average *adjusted* observation score from the current year and all prior years. We use the average of prior and current scores to minimize instability in ratings while still allowing principals to vary in their effectiveness over time. We operationalize the principal quality gap as the difference in the percentage of "low-quality" (i.e., 0–2 years of experience or bottom quartile of ratings) principals between low-achievement (or middle-achievement) and high-achievement schools.

The components of the principal quality gap in our simulation are: exits, promotions, demotions, transfers, new hires, and reclassifications (i.e., moves from one quality category to another).[24] Differences in these rates across categories of school advantage will affect the distribution of principal quality. Reclassifications capture changes in principal experience and supervisor ratings over time. Instead of assuming that a principal's quality is fixed, our simulation allows for principals to gain experience and change effectiveness (that is, receiving higher or lower supervisor ratings). For principal experience, differential reclassification cannot contribute to principal quality gaps, as all principals gain experience at the same rate. For supervisor ratings, however, we allow for the possibility that

[24] As with any simulation approach, we make some important simplifications and assumptions. First, we model the principal quality gap as a Markov process (Goldhaber et al., 2018), meaning that the gap in year t is completely determined by the baseline gap (i.e., year $t-1$) and the time-invariant simulation parameters. In reality, rates of hiring and turnover are constantly changing in response to many factors (e.g., labor market dynamics, educator evaluation reforms) that we do not capture in our simulation.

principals in high-achievement and low-achievement schools reclassify (i.e., move between quartiles in the distribution of scores) at different rates.

To parse out the individual contribution of each component of the principal quality gap, we run a separate simulation for each component in which all other components are equal across school groups. We repeat this procedure for each component and sum the individual gaps to obtain the total gap in principal quality (Goldhaber et al., 2018). For both principal experience and supervisor ratings, the simulated gap is very close to the actual gap, which suggests our simulation is a reasonable approximation for sorting dynamics in Tennessee.

Table 6 contains the simulation results for principal experience. Panel A shows the gap between low-achievement and high-achievement schools in terms of the proportion of principals with fewer than three years of prior principal experience. Panel B shows the gap between middle-achievement and high-achievement schools. The simulation begins (year 0) with no principal quality gaps and runs for ten years. The rightmost column shows the contribution of each component as a percentage of the total gap. We focus our discussion of the results on the gaps between low-achievement and high-achievement schools.

The gap between low-achievement and high-achievement schools is 0.11, which means a low-achievement school is 11 percentage points more likely to be led by an inexperienced principal than is a high-achievement school. This gap closely mirrors the empirical gap shown in Table 2. The largest contributor to this experience gap is differential exit rates. For example, if the only difference between principals in high- and low-achievement schools was their exit rates, the principal experience gap would be 5.3 percentage points. Similarly, higher demotion rates in low-achievement schools increase the relative proportion of inexperienced principals. These components operate similarly in terms of increasing the principal quality gap; principals lost to exit or demotion are filled by new hires, who tend to have less experience.

Principal transfers also contribute to quality gaps, though to a lesser extent. This

contribution is driven by two factors. Low-achievement schools have higher overall transfer rates (see Table 5) and the probability of moving from a low-achievement to high-achievement school is greater than the probability of moving from a high-achievement to low-achievement school.[25]

In contrast, promotion and hiring actually decrease gaps, albeit only slightly. While overall promotion rates are roughly equal between principals in high- and low-achievement schools, promoted principals in high-achievement schools tend to be more experienced than their counterparts in low-achievement schools. Thus, the loss in experience due to promotions is greater in high-achievement schools, which shrinks the quality gap. For hiring, low-achievement schools are slightly less likely than high-achievement schools to hire an inexperienced principal, which also shrinks the quality gap. [26]

Table 7 shows the simulation results for supervisor ratings. Panel A shows that the quality gap is 0.14, meaning that the difference between low-achievement and high-achievement in the proportion of principals in the bottom quartile of supervisor ratings is 14 percentage points, which is almost exactly equal to the empirical gap. The gap between middle-achievement and high-achievement schools in Panel B is smaller (0.09), but the contribution of the individual components to the total gap are similar.

In contrast to the experience simulation, exits and demotions contribute little to the supervisor ratings gap. Instead, two-thirds of the total gap is explained by differential reclassification rates of principals in high-achievement versus low-achievement schools. Table C15 shows that principals in high-achievement schools are more likely than

[25] The construction of our simulation implies that transfers can only affect the gap if principals are moving to a different school group (e.g., moving from a low-achievement to high-achievement school). Furthermore, if principals who transfer out of a given school group are replaced by principals transferring in from other school groups, the experience gap will not change.

[26] In Panel B, which compares middle-achievement and high-achievement schools, we again find that exits, transfers, and demotions contribute most greatly to the total experience gap. However, relative to low-achievement schools, differential exit and demotion rates contribute less to the gap, while transfers are slightly more important (in percentage terms). Again we find that promotions and hiring are the least substantial components, though in this case they increase slightly the gap between middle-achievement and high-achievement schools, rather than decrease it.

principals in low-achievement schools to improve their ratings over time, particularly among those in the bottom quartile. For example, among those with ratings in the bottom 25% of the distribution in a given year, the probability of moving into the middle 50% in the following year is 34% for principals in high-achievement schools, compared to 26% for principals in low-achievement schools. Relatedly, principals in low-achievement schools are also more likely to move down in the distribution. Among principals working in low-achievement schools who score in the top quartile of ratings in a given year, 17% will move out of the top quartile in the following year, compared to 12% of principals in high-achievement schools. These differential reclassification rates help produce the disparities in adjusted supervisor ratings shown in Table 2.

The remainder of the quality gap is explained by transfers and hiring. Movement from low-achievement to middle- or high-achievement schools is infrequent, but it is more likely among principals with average to high supervisor ratings.[27] Low-achievement schools are also more likely to hire a principal that scores in the bottom quartile in their first year (48%) than are high-achievement schools (41%). Despite their higher frequency in low-achievement schools, exits and demotions explain very little of the ratings gap because they are concentrated among principals with the lowest ratings. Finally, promotion rates actually serve to *decrease* the quality gap between high-achievement and low- or medium-achievement schools, as the greatest rates of promotion to central office are among highly rated principals in high-achievement schools.

What do we learn from these simulation results? First, the drivers of principal sorting vary by the measure of principal quality. The disparity in principal experience between low-achievement and high-achievement schools shown in Table 2 is largely a function of higher turnover rates, which we documented in Table 5. Because replacement principals tend to be relatively inexperienced regardless of the achievement level of the school, these

[27] In fact, we observe no cases where a principal in the bottom quartile of ratings moved from a low achievement school to a high achievement school. Further, transfers from middle- or high-achievement schools to low-achievement schools are almost non-existent, regardless of the principal's average rating.

higher turnover rates translate to low levels of principal experience in low-achievement schools. However, when we consider how those principals are rated by their supervisors, we come to different conclusions. Principals who leave their positions tend to have lower ratings, on average (Grissom & Bartanen, 2018), and partly as a result, higher turnover rates are not the main driver of disparities in supervisor ratings. Instead, disparities result from the tendency of low-rated principals in high-achievement schools to move out of the bottom quartile of ratings at higher rates than principals in low-achievement schools. Principals in low-achievement schools do not improve their ratings over time to the same degree as their colleagues in other schools.[28]

Do Tennessee's Patterns Hold Nationally?

To assess the generalizability of the Tennessee findings, we draw on data from the 2011–12 Schools and Staffing Survey (SASS), administered by the National Center for Educational Statistics (NCES). Although cross-sectional and more limited in measures of principal quality, SASS data are nationally representative, collected from a stratified random sample of public schools. The analysis reported below utilizes data on approximately 7,500 public schools. In addition to the main survey, NCES implemented the Principal Follow-up Survey (PFS) in the year following to collect information on the responding principal's whereabouts. We utilize PFS data to calculate measures of principal turnover.

Table 8 summarizes the distribution of principal quality nationally. Schools are categorized by three of the same four categories, with the only difference being that we do not have a measure of school achievement. Distributional patterns observed among schools in Tennessee appear to hold nationally. Schools with the highest concentrations of marginalized students are led by less qualified principals. High-poverty schools, for

[28] We also investigated whether the lower reclassification rates from the first quartile to the middle quartiles was due to higher transfer rates. When we allow for reclassification to vary by principals who remain in the same school versus those who transfer, the simulation results are identical.

example, employ principals with 1.3 fewer years of experience, on average, than low-poverty schools. They are also more likely to employ inexperienced principals (i.e., principals in their first-to-third year) than those in the lowest category (28% to 20%). Their principals have spent 0.5 fewer years in that school building. They are also less likely to hold an education specialist degree or a doctoral degree (34%, compared to 40% in the low-poverty group).

Moreover, on average, they score 0.24 standard deviations lower on the subjective performance measure constructed from teacher survey responses.[29] Each of these differences is statistically significant. The patterns are similar or even more pronounced when looking across percent students of color. However, somewhat in contrast to Tennessee, the national data show less of a clear suburban advantage in terms of principal quality measures. While rural/town schools are more likely to employ novice principals than suburban schools (8% and 6%, $p < 0.05$) and their principals are less likely to hold an education specialist degree or a doctoral degree (33% and 40%, $p < 0.01$), they have similar experience levels and performance ratings from teachers. Urban schools are led by principals with lower performance ratings than suburban schools ($p < 0.01$). In addition, they tend to hire principals with fewer years of experience than those in suburban schools ($p < 0.01$).

Table 9 replicates, to the extent possible, the hiring and turnover patterns using the SASS data. In comparison to the Tennessee data, for new hires we have fewer measures of prior experience and performance. We find that while new hires in schools serving low proportions of FRPL-eligible students or students of color tend to have more years of prior principal experience, the differences are small in magnitude and not statistically significant. Similar to our findings in Tennessee, new hires in traditionally disadvantaged schools have

[29] SASS data allow us to create a teacher ratings measure similar to that constructed in Tennessee. The SASS teacher survey features five questions about principal leadership which we use to conduct a factor analysis similar to the one performed for the teacher ratings in Tennessee. These items, shown in Appendix Table B7, identify one underlying subjective leadership performance factor; factor scores from a similar measure have been used to capture principal effectiveness in other studies (e.g., Grissom, 2011). Standardized factor scores are averaged at the school level from all teacher responses in the school.

lower average performance (as rated by teachers) in their first year than new hires in more advantaged schools. Also, while principal turnover rates nationally (22%) are higher than in Tennessee (18%), turnover disparities by school categories mirror our prior findings. For example, 28% of principals in high-poverty schools turned over after 2011–12, compared to 21% in low-poverty schools ($p < 0.01$).[30] Nationally, transfers are the largest driver of turnover disparities.

Discussion and Conclusions

Our analysis of principal quality measures by student characteristics finds strong evidence of inequitable leadership sorting. By virtually every measure we examine, less advantaged schools face leadership quality deficits; schools with large proportions of low-income students, students of color, and low-achieving students are led by principals with weaker qualifications and lower performance ratings. Also, although mostly overlooked in prior work, these gaps are often just as apparent in rural schools as in their urban counterparts.

The lamentable punch line of such sorting is that the kind of effective leadership required for school success is scarcer in precisely the kinds of school that would benefit from it most. For instance, in Tennessee, 38% of principals in schools in the bottom quintile of achievement are in their first three years on the job. Research demonstrates that novice principals generally have less developed instructional leadership, school management, and problem-solving skills, which likely are necessary for school improvement (e.g., Daresh, 1986; Grissom & Loeb, 2011; Leithwood & Steinbach, 1995). Less experienced principals in these schools may be even less prepared, having served less time—though only about half a year less—in AP positions, often the training ground for future principals (Bastian & Henry, 2015). More important than the experience differences, practice ratings for principals in these same schools fall well below average. Prior work has shown that lower

[30] Principal turnover rates were also higher in urban schools (26%) than suburban or town/rural schools (22% and 21%), though this difference is not statistically significant.

practice ratings predict lower student achievement growth, higher turnover rates among effective teachers, and other key school outcomes (Grissom & Bartanen, 2018; Grissom, Blissett, & Mitani, 2018a). In short, principals in the schools with the greatest needs are the least positioned to drive improvement, and the patterns of principal sorting we observe likely contribute to opportunity and performance gaps between schools serving higher and lower concentrations of marginalized student populations.

Inequitable principal sorting is thus a significant policy problem. Addressing it requires understanding its drivers. Our analysis suggest that these inequitable distribution patterns are driven both by higher turnover among principals in challenged schools and disparities in hiring of new leaders to replace departing principals. For instance, the yearly principal turnover rate of schools in bottom quintile of student achievement in Tennessee is 23%, compared to 14% for schools in the highest quintile. High rates of turnover in such schools in rural areas in Tennessee are particularly striking. Similar disparities exist across a broad range of principal characteristics and categories of school advantage. New hires have fewer total years of administrator experience, lower performance ratings from their prior roles as assistant principals, lower licensure examination scores, and lower effectiveness in their first year, though, importantly, data limitations mean that we cannot be sure whether these differences are driven by differences districts' hiring practices or decisions or by differences in application and job-seeking behavior by candidates with higher or lower qualifications. Moreover, our simulation results uncover that principals in more challenged schools may improve more slowly with experience as well, further exacerbating quality gaps.

Importantly, districts likely have more policy options in addressing inequitable principal sorting than they have in the case of teachers. Administrators typically do not collectively bargain contracts, and salaries often are not set by salary schedules. As middle management, principals are more subject to district-level reassignment decisions. Indeed, district reassignment decisions may constitute part of the problem, if district leaders choose

to move principals in challenging leadership environments more frequently; a limitation of our analysis is that we cannot differentiate principal-initiated transitions from district-initiated ones. We suggest that districts direct their efforts towards stemming principal turnover in low-income or low-achieving schools by prioritizing school leadership stability in their own personnel decisions and through targeted retention strategies aimed at reducing voluntary turnover, such as retention bonuses or increased mentoring, coaching, and other supports for principals leading especially challenging schools. When vacancies in such schools arise, districts should concentrate efforts on recruiting high performers with proven track records of leadership effectiveness, potentially varying pay to compensate principals for taking on ambitious leadership assignments. As Clotfelter et al. (2006) argue, less traditional options, such as housing assistance programs, may also be effective in addressing principal sorting, since principals often seek jobs near their homes, which may not be near the kinds of schools that need high-quality leadership most. In addition, proactive programs to build pipelines of effective leadership candidates can help districts compensate for what might otherwise be a scarce supply of high-quality leaders available to fill such leadership positions (Turnbull, Anderson, Riley, MacFarlane, & Aladjem, 2016).

Attention to creating pipelines of well-prepared leaders coupled with coaching and related strategies may also help to address the pattern of lower returns to experience that our simulation results identify as a potentially important contributor to principal quality gaps. However, we caution that few studies have investigated principal development and how it may vary by school characteristics, and measurement in this area is a particular challenge. Additional research on principal improvement is necessary before we draw firm conclusions from this finding.

The analysis we have provided here is descriptive, aimed at providing an initial look at principal quality and some suggestion of the mechanisms that drive it. Future research should consider the strategies school districts employ to recruit, hire, and place principals, and how these strategies affect sorting patterns. It should also delve deeper into the factors

that drive principal turnover by school characteristics (Grissom & Bartanen, 2018), and how districts can be successful in reducing leadership turnover. Research might also consider how accountability and evaluation systems may exacerbate or ameliorate patterns of principal sorting.

References

Bastian, K. C., & Henry, G. T. (2015). The Apprentice. *Educational Administration Quarterly*, *51*(4), 600–639.

Béteille, T., Kalogrides, D., & Loeb, S. (2012). Stepping stones: Principal career paths and school outcomes. *Social Science Research*, *41*(4), 904–919.

Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S., & Wyckoff, J. (2011). The Influence of School Administrators on Teacher Retention Decisions. *American Educational Research Journal*, *48*(2), 303–333.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2005). The Draw of Home: How Teachers' Preferences for Proximity Disadvantage Urban Schools. *Journal of Public Policy and Management*, *24*(1), 113–132.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals.*

Brewer, D. J. (1993). Principals and Student Outcomes: Evidence From U.S. High Schools. *Economics of Education Review*, *12*(4), 281–292.

Cheney, G. R., & Davis, J. (2011). *Gateways to the Principalship:State Power to Improve the Quality of School Leaders* (Tech. Rep.). Washington, D.C.: Center for American Progress.

Chetty, R., Friedman, J., & Rockoff, J. (2014). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, *104*(9), 2633–2679.

Clark, D., Martorell, P., & Rockoff, J. (2009). *School Principals and School Performance.* Washington, D.C.

Clifford, M., & Ross, S. (2011). *Designing Principal Evaluation Systems: Research to Guide Decision-making. An Executive Summary of Current Research* (Tech. Rep.). Washington, D.C.: American Institutes for Research.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2005). Who teaches whom? Race and the

distribution of novice teachers. *Economics of Education Review*, *24*, 377–392.

Clotfelter, C. T., Ladd, H. F., Vigdor, J. L., & Wheeler, J. (2006). High-poverty schools and the distribution of teachers and principals. *North Carolina Law Review*, *85*, 1345–1379.

Coelli, M., & Green, D. A. (2012). Leadership effects: school principals and student outcomes. *Economics of Education Review*, *31*, 92–109.

Cohen-Vogel, L., & Osborne-Lampkin, L. (2007). Allocating Quality: Collective Bargaining Agreements and Administrative Discretion Over Teacher Assignment. *Educational Administration Quarterly*, *43*(4), 433–461.

Cosner, S., Tozer, S., Zavitkovsky, P., & Whalen, S. P. (2015). Cultivating Exemplary School Leadership Preparation at a Research Intensive University. *Journal of Research on Leadership Education*, *10*(1), 11–38.

Daresh, J. C. (1986). Support for Beginning Principals: First Hurdles are Highest. *Theory Into Practice*, *25*(3), 168–173.

Dhuey, E., & Smith, J. (2014). How important are school principals in the production of student achievement? *Canadian Journal of Economics*, *47*(2), 634–663.

Fuller, E., & Young, M. D. (2009). Tenure and retention of newly hired principals in Texas. *Texas High School Project*, 19.

Gates, S. M., Ringel, J. S., Santibanez, L., Guarino, C., Ghosh-Dastidar, B., & Brown, A. (2006). Mobility and turnover among school principals. *Economics of Education Review*, *25*(3), 289–302.

Glazerman, S., & Max, J. (2011). *Do Low-income Students Have Equal Access to the Highest-Performing Teachers?* (Tech. Rep.). Mathematica Policy Research.

Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven Playing Field? Assessing the Teacher Quality Gap Between Advantaged and Disadvantaged Students. *Educational Researcher*, *44*(5), 293–307.

Goldhaber, D., Quince, V., & Theobald, R. (2018). *How Did It Get This Way?*

Disentangling the Sources of Teacher Quality Gaps Across Two States. New York, NY.

Goldhaber, D., Walch, J., & Gabele, B. (2014). Does the Model Matter? Exploring the Relationship Between Different Student Achievement-Based Teacher Assessments. *Statistics and Public Policy*, *1*(1), 28–39.

Grissom, J. A. (2011). Can Good Principals Keep Teachers in Disadvantaged Schools? Linking Principal Effectiveness to Teacher Satisfaction and Turnover in Hard-to-Staff Environments. *Teachers College Record*, *113*(11), 2552–2585.

Grissom, J. A., & Bartanen, B. (2018). Principal Effectiveness and Principal Turnover. *Education Finance and Policy*.

Grissom, J. A., Blissett, R. S. L., & Mitani, H. (2018a). Evaluating School Principals: Supervisor Ratings of Principal Practice and Principal Job Performance. *Educational Evaluation and Policy Analysis*, *40*(3), 446–472.

Grissom, J. A., Blissett, R. S. L., & Mitani, H. (2018b). *Supervisor Ratings as Measures of Principal Performance: Evidence from the TEAM Evaluation System in Tennessee.* Nashville, TN.

Grissom, J. A., Kalogrides, D., & Loeb, S. (2015). Using Student Test Scores to Measure Principal Performance. *Educational Evaluation and Policy Analysis*, *37*(1), 3–28.

Grissom, J. A., & Loeb, S. (2011). Triangulating Principal Effectiveness: How Perspectives of Parents, Teachers, and Assistant Principals Identify the Central Importance of Managerial Skills. *American Educational Research Journal*, *48*(5), 1091–1123.

Grissom, J. A., Mitani, H., & Blissett, R. S. L. (2017). Principal Licensure Exams and Future Job Performance: Evidence From the School Leaders Licensure Assessment. *Educational Evaluation and Policy Analysis*, *39*(2), 248–280.

Guarino, C. M., Santibañez, L., & Daley, G. A. (2006). Teacher Recruitment and Retention: A Review of the Recent Empirical Literature. *Review of Educational Research*, *76*(2), 173–208.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why Public Schools Lose Teachers. *Journal of Human Resources*, *39*(2), 326–354.

Herman, R., Gates, S. M., Arifkhanova, A., Bega, A., Chavez-Herrerias, E., Han, E., ... Wrabel, S. (2017). *School Leadership Interventions Under the Every Student Succeeds Act: Evidence Review* (Tech. Rep.). Santa Monica, CA: RAND Corporation.

Isenberg, E., Max, J., Gleason, P., Potamites, L., Santillano, R., Hock, H., & Hansen, M. (2013). *Access to Effective Teaching for Disadvantaged Students (NCEE 2014-4001)* (Tech. Rep.). Washington, D.C.: National Center for Education Evaluation and Regional Assistance.

Ladd, H. F. (2011). Teachers' Perceptions of Their Working Conditions: How Predictive of Planned and Actual Teacher Movement? *Educational Evaluation and Policy Analysis*, *33*(2), 235–261.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis. *Educational Evaluation and Policy Analysis*, *24*(1), 37–62.

Leithwood, K., & Steinbach, R. (1995). *Expert problem solving: Evidence from school and district leaders.* New York, NY: SUNY Press.

Loeb, S., Kalogrides, D., & Horng, E. L. (2010). Principal Preferences and the Uneven Distribution of Principals Across Schools. *Educational Evaluation and Policy Analysis*, *32*(2), 205–229.

Miller, A. (2013). Principal turnover and student achievement. *Economics of Education Review*, *36*, 60–72.

Morris, V. C., Crowson, R. L., Hurwitz, E., & Porter, C. (1982). The Urban Principal: Middle Manager in the Educational Bureaucracy. *The Phi Delta Kappan*, *63*(10), 689–692.

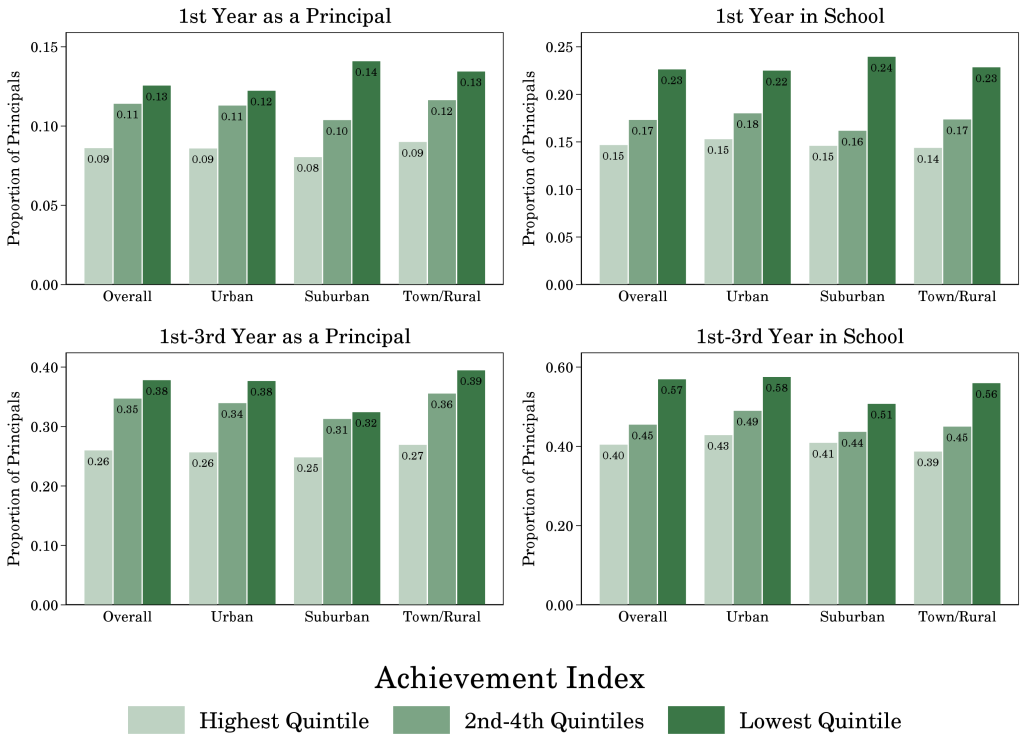
Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban*

Economics, *72*, 104–122.

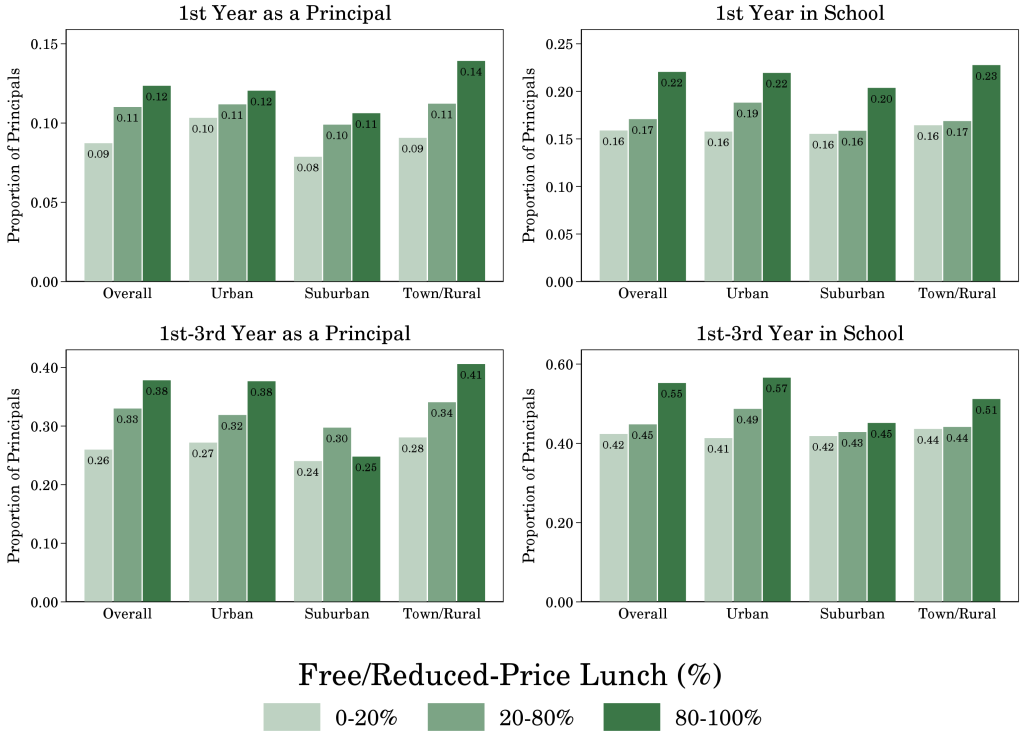
Sebastian, J., & Allensworth, E. (2012). The Influence of Principal Leadership on Classroom Instruction and Student Learning: A Study of Mediated Pathways to Learning. *Educational Administration Quarterly*, *48*(4), 626–663.

Turnbull, B. J., Anderson, L. M., Riley, D. L., MacFarlane, J. R., & Aladjem, D. K. (2016). *The Principal Pipeline Initiative In Action* (Vol. 5; Tech. Rep.). Policy Studies Associates, Inc.

(a) School Achievement Index



(b) Student Poverty



(c) Students of Color

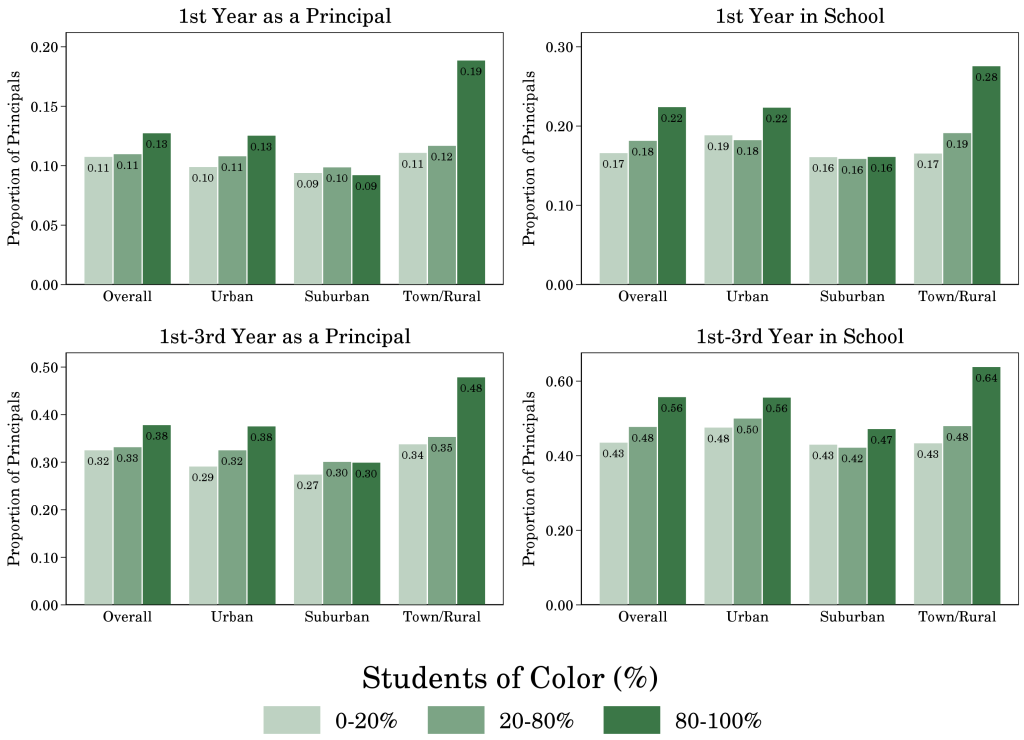
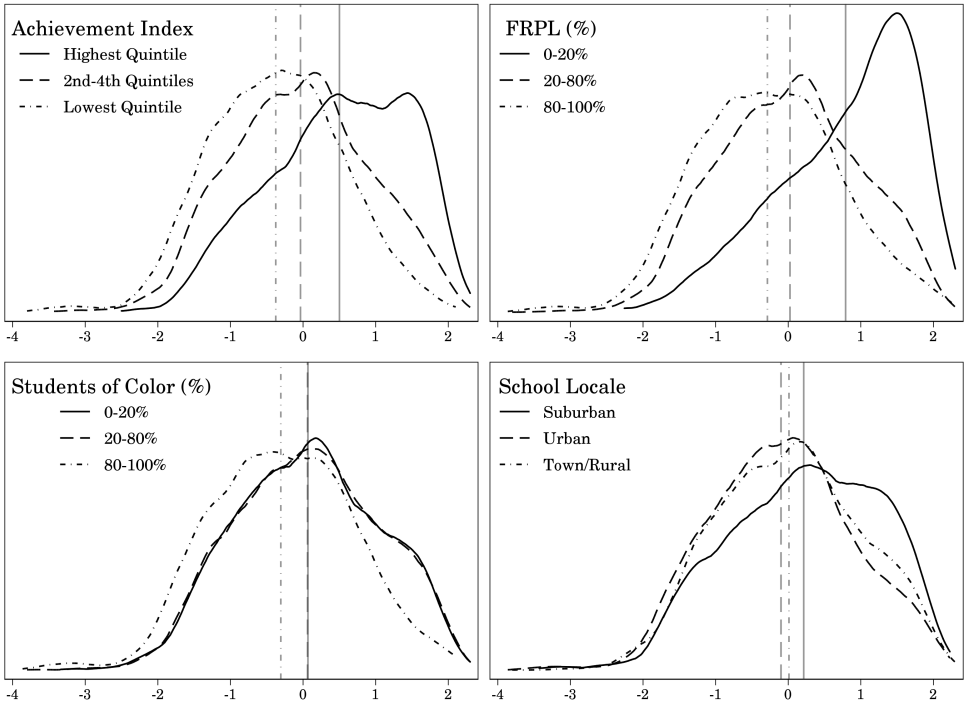


Figure 1. Principal Experience and Tenure by School Characteristics in Tennessee

Notes: Each bar shows the proportion of principals in a given school category (locale by achievement/poverty/nonwhite) who have the given level of experience/tenure listed above the plot.

(a) Supervisor Ratings



(b) Teacher Ratings

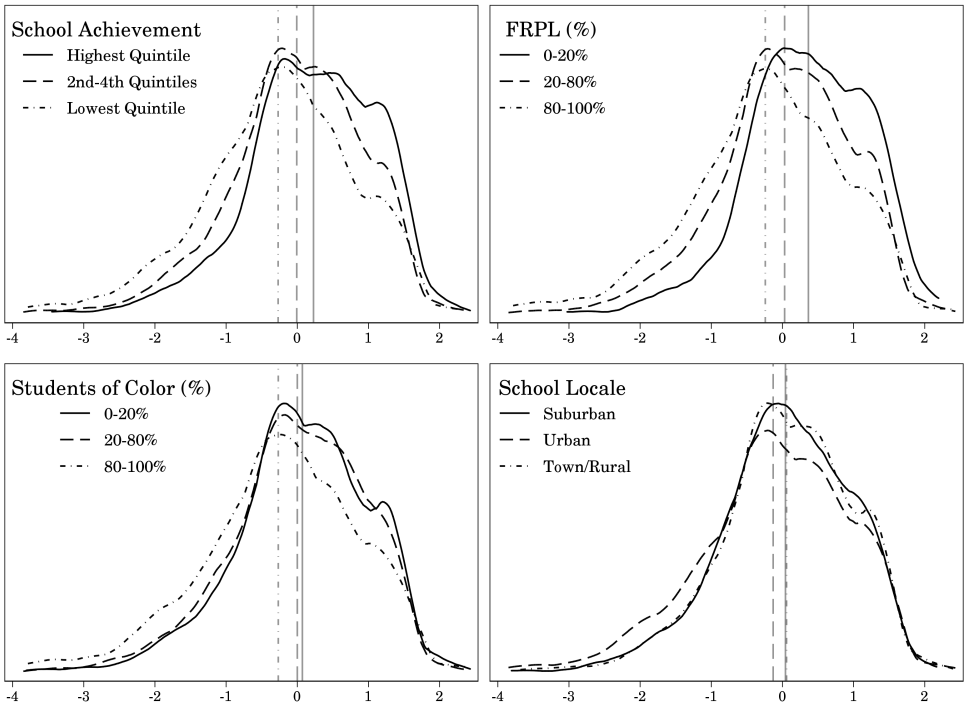
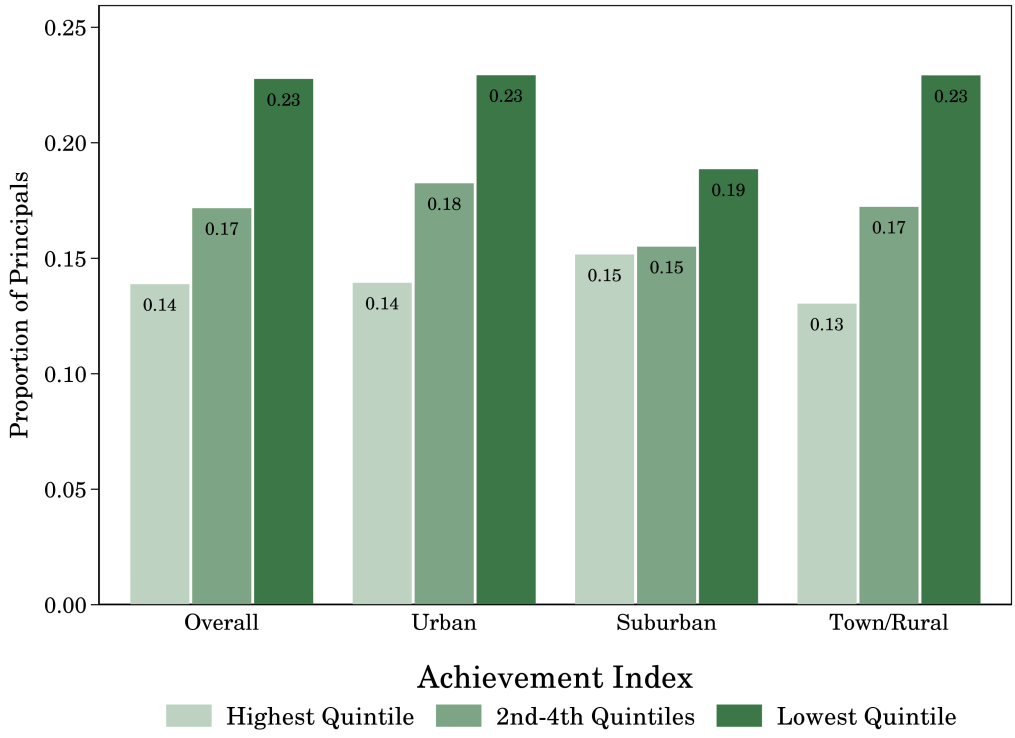


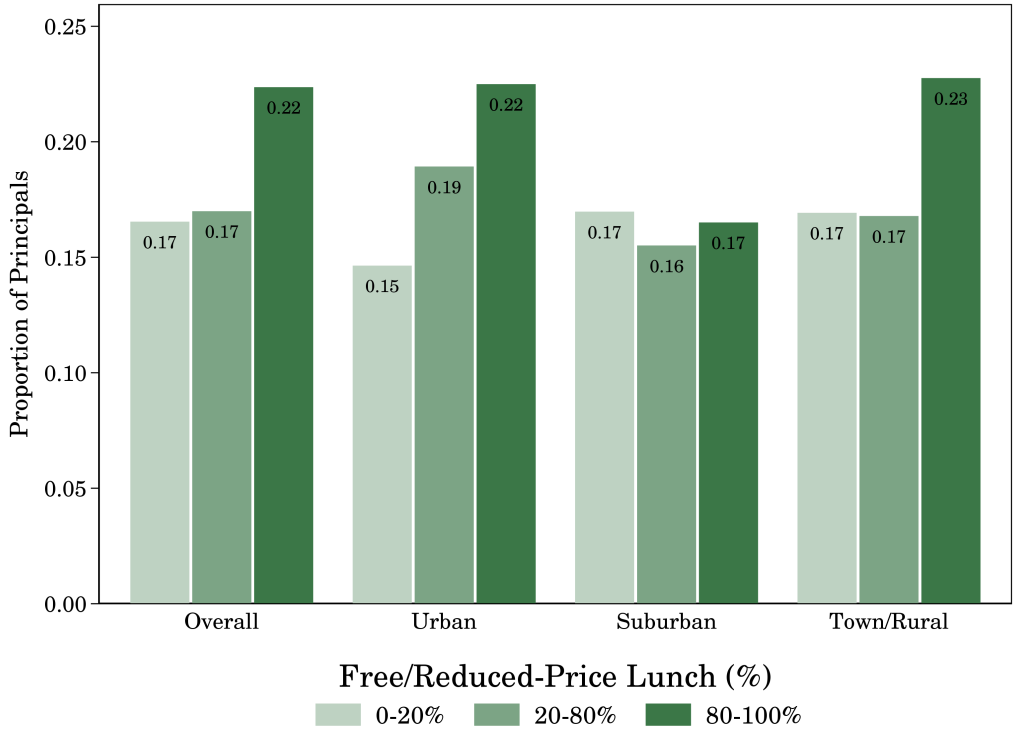
Figure 2. Distribution of Principal Ratings by School Characteristics

Notes: The vertical lines show the mean supervisor/teacher rating by school type. Both measures are standardized. For display purposes, we have excluded observations more than 4 standard deviations below the mean (2 observations for supervisor ratings and 14 observations for teacher ratings).

(a) School Achievement Index



(b) Student Poverty



(c) Students of Color

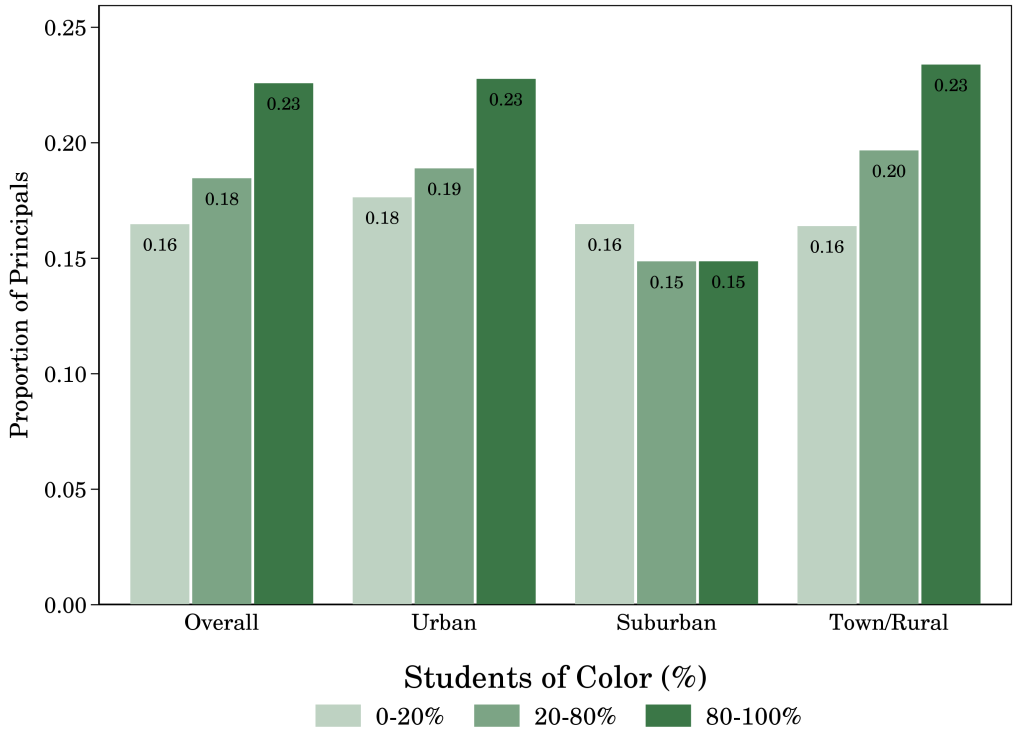


Figure 3. Principal Turnover by School Characteristics in Tennessee

Notes: Each bar shows the proportion of principals in a given school category (locale by achievement/poverty/nonwhite) who leave their positions each year.

Table 1
Descriptive Statistics

	N	Mean	SD	Min	Max
Principal Characteristics					
Female	18305	0.55			
Black	18305	0.19			
Age	18012	50.0	9.1	19	93
Ed.S. or Ph.D.	18230	0.40			
SLLA Score	4524	175.9	8.4	139	198
Prior Principal Experience					
0 Years	18305	0.11			
1–2 Years	18305	0.22			
3–4 Years	18305	0.18			
5+ Years	18305	0.48			
Tenure in School					
0 Years	18305	0.18			
1–2 Years	18305	0.29			
3–4 Years	18305	0.19			
5+ Years	18305	0.34			
Performance Measures					
Supervisor Rating (std)	9120	0.01	0.99	-4.97	2.30
Teacher Rating (std)	8544	0.00	0.98	-4.90	2.43
School Characteristics					
Achievement Index	17248	0.03	0.96	-6.49	6.31
Enrollment (100s)	18305	6.44	3.83	0.26	40.65
Proportion Black	18270	0.25	0.31	0.00	1.00
Proportion Hispanic	18270	0.06	0.09	0.00	0.74
Proportion Gifted	18270	0.02	0.03	0.00	0.56
Proportion w/ Disabilities	18270	0.15	0.08	0.00	1.00
Proportion FRPL	18270	0.57	0.26	0.00	1.00
School Locale					
Urban	18241	0.31			
Suburban	18241	0.15			
Town	18241	0.16			
Rural	18241	0.39			
School Level					
Elementary	18238	0.59			
Middle	18238	0.19			
High	18238	0.18			
Other	18238	0.05			

Notes: Includes principals in Tennessee from 2006–07 to 2016–17. Supervisor and teacher ratings are available beginning in 2011–12.

Table 2
Distribution of Principal Quality by School Characteristics in Tennessee

	Principal Experience		Tenure in School		Other Qualifications		Effectiveness		
	0 years	0–2 years	0 years	0–2 years	Ed.S. or Ph.D.	SLLA Score	Unadjusted Supervisor Ratings	Adjusted Supervisor Ratings	Teacher Ratings
All Schools	0.11	0.34	0.18	0.47	0.40	175.91	0.01	0.00	0.00
Achievement Index									
Highest Quintile	0.09	0.26	0.15	0.40	0.42	178.29	0.50	0.21	0.23
Middle 60%	0.11***	0.35***	0.17***	0.46***	0.41	175.59***	−0.04***	−0.01***	−0.01***
Lowest Quintile	0.13***	0.38***	0.23***	0.57***	0.37***	175.02***	−0.38***	−0.18***	−0.27***
FRPL %									
0–20%	0.09	0.26	0.16	0.43	0.43	178.65	0.79	0.18	0.36
20–80%	0.11**	0.33***	0.17	0.45	0.41	175.90***	0.02***	0.02***	0.03***
80–100%	0.12***	0.38***	0.22***	0.55***	0.38***	175.43***	−0.29***	−0.12***	−0.24***
Students of Color %									
0–20%	0.11	0.33	0.17	0.44	0.44	175.49	0.07	0.01	0.07
20–80%	0.11	0.33	0.18**	0.48***	0.38***	177.08***	0.06	0.05**	0.00***
80–100%	0.13***	0.38***	0.23***	0.56***	0.31***	175.18	−0.31***	−0.16***	−0.27***
Locale									
Suburban	0.10	0.28	0.16	0.43	0.40	178.86	0.21	0.01	0.04
Town/Rural	0.11***	0.34***	0.17	0.45*	0.42**	174.74***	0.01***	−0.01	0.06
Urban	0.12***	0.35***	0.20***	0.52***	0.37**	176.96***	−0.10***	0.01	−0.13***

Notes: Asterisks indicate significant differences from the base categories (Highest quintile of achievement, 0–20% FRPL, 0–20% Nonwhite, Suburban).
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3
Qualifications of New Hires by School Characteristics in Tennessee

	Total Admin Exper.	Total Principal Exper.	Any Principal Exper.	Total AP Exper.	Any AP Exper.	Was AP in Same School	Ed.S. or Ph.D.	SLLA Score
All Schools	4.52	1.71	0.39	2.80	0.72	0.23	0.43	175.96
Achievement Index								
Highest Quintile	5.26	2.00	0.40	3.26	0.79	0.28	0.44	178.65
Middle 60%	4.34***	1.54**	0.36	2.80**	0.71***	0.28	0.44	175.59***
Lowest Quintile	4.57**	1.90	0.45	2.67***	0.74	0.12***	0.40	175.39***
FRPL %								
0–20%	5.70	1.96	0.41	3.74	0.84	0.25	0.45	178.45
20–80%	4.43***	1.60	0.37	2.83***	0.72***	0.28	0.43	175.98**
80–100%	4.51***	2.00	0.45	2.51***	0.69***	0.10***	0.41	175.42**
Students of Color %								
0–20%	4.18	1.52	0.36	2.65	0.68	0.28	0.48	175.78
20–80%	5.07***	1.90**	0.40	3.18***	0.80***	0.24	0.39***	176.74*
80–100%	4.42	1.85*	0.44***	2.56	0.68	0.12***	0.36***	175.16
Locale								
Suburban	5.44	2.08	0.41	3.36	0.84	0.31	0.42	178.35
Town/Rural	3.96***	1.34***	0.34**	2.62***	0.66***	0.27	0.44	175.03***
Urban	4.90**	2.06	0.44	2.84**	0.76***	0.15***	0.41	176.82*

Notes: Asterisks indicate significant differences from the base categories (Highest quintile of achievement, 0–20% FRPL, 0–20% Nonwhite, Suburban).
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4
Effectiveness of New Hires by School Characteristics in Tennessee

	Supervisor Ratings					Teacher Ratings	
	First Year in School	First Year in School (Adjusted)	Prior Years as Prin	Prior Years as Prin (Adjusted)	Prior Years as AP	First Year in School	Prior Years as Prin
All Schools	−0.39	−0.35	−0.23	−0.05	0.08	−0.04	0.04
Achievement Index							
Highest Quintile	0.03	−0.24	0.13	0.08	0.51	0.18	0.07
Middle 60%	−0.41***	−0.34*	−0.19***	−0.00	0.03***	0.02**	0.11
Lowest Quintile	−0.63***	−0.41***	−0.47***	−0.16**	−0.04***	−0.39***	−0.12
FRPL %							
0–20%	0.29	−0.23	0.26	0.12	0.69	0.28	0.15
20–80%	−0.38***	−0.33	−0.17**	−0.01	0.08***	0.02**	0.08
80–100%	−0.63***	−0.44**	−0.46***	−0.18*	−0.09***	−0.33***	−0.11
Students of Color %							
0–20%	−0.35	−0.36	−0.15	−0.01	0.10	0.12	0.11
20–80%	−0.34	−0.27**	−0.20	−0.05	0.07	−0.09***	0.07
80–100%	−0.67***	−0.48**	−0.43**	−0.13	0.04	−0.43***	−0.15**
Locale							
Suburban	−0.26	−0.38	−0.02	0.06	0.30	−0.01	0.13
Town/Rural	−0.38*	−0.36	−0.21*	−0.06	0.04***	0.08	0.06
Urban	−0.49***	−0.32	−0.36***	−0.11	0.03***	−0.27***	−0.05

Notes: Asterisks indicate significant differences from the base categories (Highest quintile of achievement, 0–20% FRPL, 0–20% Nonwhite, Suburban).
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5
Principal Turnover in Tennessee

	All Turnover	Transfer (Within)	Transfer (Across)	Promote	Demote	Exit
All Schools	0.18	0.04	0.00	0.03	0.03	0.07
Achievement Index						
Highest Quintile	0.14	0.03	0.00	0.03	0.01	0.06
Middle 60%	0.17***	0.03	0.00	0.03	0.03***	0.07
Lowest Quintile	0.23***	0.06***	0.01	0.03	0.05***	0.09***
FRPL %						
0–20%	0.17	0.03	0.00	0.03	0.02	0.08
20–80%	0.17	0.03	0.00**	0.03	0.03	0.07
80–100%	0.22***	0.06***	0.00*	0.03	0.04***	0.09
Students of Color %						
0–20%	0.16	0.03	0.00	0.03	0.03	0.07
20–80%	0.18***	0.04***	0.00	0.03	0.03	0.08*
80–100%	0.23***	0.06***	0.01	0.02***	0.04***	0.10***
Locale						
Suburban	0.16	0.04	0.01	0.03	0.02	0.07
Town/Rural	0.17	0.03**	0.00	0.03*	0.03***	0.07
Urban	0.21***	0.06***	0.00	0.02	0.04***	0.09***

Notes: Transfer (Within) are principals who move to another principal position in the same district, while Transfer(Across) are moves to a different district. Promotions are principals who move to a central office position. Demotions are principals who move to a non-principal school-level position (e.g., AP). Exits are principals who are no longer working in the K–12 system (e.g., retirements, moves out of state). Asterisks indicate significant differences from the base categories (Highest quintile of achievement, 0–20% FRPL, 0–20% Nonwhite, Suburban).
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6
Simulated Principal Quality Gap, Inexperienced Principals (0–2 years of experience)

	Simulation Year											
	0	1	2	3	4	5	6	7	8	9	10	%
Low-Achievement												
Total Gap	**0**	**0.051**	**0.080**	**0.095**	**0.103**	**0.107**	**0.109**	**0.110**	**0.111**	**0.111**	**0.111**	**100%**
Exit	0	0.023	0.036	0.044	0.049	0.051	0.052	0.053	0.053	0.053	0.053	48%
Transfer	0	0.014	0.021	0.025	0.026	0.027	0.028	0.028	0.028	0.028	0.028	25%
Promote	0	-0.003	-0.005	-0.007	-0.008	-0.009	-0.009	-0.010	-0.010	-0.010	-0.010	-9%
Demote	0	0.021	0.034	0.041	0.045	0.047	0.048	0.049	0.049	0.049	0.049	45%
Hiring	0	-0.004	-0.006	-0.008	-0.009	-0.009	-0.009	-0.010	-0.010	-0.010	-0.010	-9%
Middle-Achievement												
Total Gap	**0**	**0.032**	**0.049**	**0.059**	**0.064**	**0.067**	**0.068**	**0.069**	**0.069**	**0.069**	**0.069**	**100%**
Exit	0	0.007	0.011	0.013	0.015	0.016	0.016	0.016	0.016	0.016	0.016	24%
Transfer	0	0.010	0.015	0.017	0.018	0.019	0.019	0.019	0.019	0.019	0.019	28%
Promote	0	0.002	0.003	0.003	0.003	0.004	0.004	0.004	0.004	0.004	0.004	5%
Demote	0	0.010	0.016	0.019	0.020	0.021	0.021	0.022	0.022	0.022	0.022	31%
Hiring	0	0.003	0.006	0.007	0.007	0.008	0.008	0.008	0.008	0.008	0.008	12%

Notes: Each gap represents the difference between the given school category and highest quintile achievement schools in the percentage of principals with two or fewer years of prior principal experience. Appendix C describes the simulation details, including tables with the parameters derived from the Tennessee data. The "true gap" between low-achievement (middle-achievement) and high-achievement schools is 0.11 (0.07).

Table 7
Simulated Principal Quality Gap, Low-Rated Principals (bottom 25% supervisor rating)

	Simulation Year											
	0	1	2	3	4	5	6	7	8	9	10	%
Low-Achievement												
Total Gap	**0**	**0.059**	**0.094**	**0.114**	**0.126**	**0.133**	**0.137**	**0.140**	**0.141**	**0.142**	**0.143**	**100%**
Exit	0	0.004	0.007	0.009	0.010	0.011	0.011	0.012	0.012	0.012	0.012	8%
Transfer	0	0.011	0.017	0.021	0.023	0.025	0.025	0.026	0.026	0.026	0.026	19%
Promote	0	-0.006	-0.010	-0.013	-0.014	-0.015	-0.015	-0.016	-0.016	-0.016	-0.016	-11%
Demote	0	0.001	0.002	0.002	0.003	0.003	0.003	0.004	0.004	0.004	0.004	3%
Hiring	0	0.009	0.014	0.017	0.018	0.019	0.020	0.020	0.020	0.021	0.021	14%
Reclassification	0	0.041	0.064	0.077	0.085	0.090	0.092	0.094	0.095	0.096	0.096	67%
Middle-Achievement												
Total Gap	**0**	**0.037**	**0.058**	**0.071**	**0.078**	**0.082**	**0.084**	**0.086**	**0.087**	**0.087**	**0.088**	**100%**
Exit	0	0.002	0.003	0.004	0.005	0.005	0.005	0.005	0.005	0.005	0.005	6%
Transfer	0	0.008	0.013	0.016	0.018	0.019	0.019	0.020	0.020	0.020	0.020	23%
Promote	0	-0.002	-0.003	-0.004	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-6%
Demote	0	0.000	0.001	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002	3%
Hiring	0	0.007	0.010	0.012	0.013	0.014	0.014	0.014	0.015	0.015	0.015	17%
Reclassification	0	0.022	0.034	0.041	0.045	0.047	0.049	0.050	0.050	0.050	0.050	58%

Notes: Each gap represents the difference between the given school category and highest quintile achievement schools in the percentage of principals with supervisor ratings in the bottom quartile. Appendix C describes the simulation details, including tables with the parameters derived from the Tennessee data. The "true gap" between low-achievement (middle-achievement) and high-achievement schools is 0.14 (0.09).

Table 8
Distribution of Principal Quality by School Characteristics in the 2011–12 SASS

	Principal Experience					
	Total Years	0 years	0–2 years	Tenure in School	Ed.S. or Ph.D.	Teacher Ratings
All Schools	7.2	0.08	0.24	4.2	0.36	0.07
FRPL %						
0–20%	7.7	0.09	0.20	4.3	0.40	0.13
20–80%	7.3	0.07	0.23**	4.4	0.35**	0.10
80–100%	6.4***	0.09	0.28***	3.7***	0.34**	−0.11***
Students of Color %						
0–20%	7.8	0.08	0.22	4.7	0.37	0.15
20–80%	7.0***	0.08	0.23	4.1***	0.35	0.12
80–100%	6.3***	0.09	0.30***	3.7***	0.35	−0.20***
Locale						
Suburban	7.3	0.06	0.23	4.2	0.40	0.15
Town/Rural	7.3	0.08**	0.23	4.4	0.33***	0.09
Urban	6.8*	0.09**	0.26	3.9	0.37	−0.04***

Notes: SASS survey weights used. Asterisks indicate significant differences from the base categories (0–20% FRPL, 0–20% Nonwhite, Suburban). Total principal experience does not include current year. Teacher ratings not mean zero due to weighting.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 9
Hiring and Turnover by School Characteristics in the 2011–12 SASS

	Hiring				Turnover			
	Total Principal Exper.	Any Principal Exper.	Ed.S. or Ph.D.	Teacher Ratings (First Year)	All Turnover	Transfer (All)	Position Change	Retire/ Other
All Schools	3.2	0.50	0.33	0.09	0.22	0.07	0.06	0.09
FRPL %								
0–20%	3.6	0.47	0.31	0.18	0.21	0.06	0.06	0.08
20–80%	3.3	0.50	0.33	0.17	0.21	0.06	0.06	0.08
80–100%	3.0	0.53	0.31	−0.19***	0.28***	0.10**	0.06	0.12*
Students of Color %								
0–20%	3.5	0.48	0.30	0.21	0.21	0.06	0.07	0.08
20–80%	3.3	0.50	0.35	0.13	0.21	0.06	0.06	0.09
80–100%	2.8	0.54	0.32	−0.20***	0.29***	0.11***	0.06	0.12***
Locale								
Suburban	3.3	0.55	0.32	0.14	0.22	0.07	0.05	0.10
Town/Rural	3.4	0.48	0.31	0.15	0.21	0.06	0.07**	0.08*
Urban	3.0	0.49	0.36	−0.06	0.26	0.09	0.06	0.11

Notes: SASS survey weights used. Asterisks indicate significant differences from the base categories (0–20% FRPL, 0–20% Nonwhite, Suburban). Total principal experience does not include current year. Teacher ratings not mean zero due to weighting.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix A

Table A1
Multivariate Analysis of Principal Quality Gaps

	0–2 Years Prin Experience		0–2 Years Prin Tenure		Unadjusted Supervisor Ratings		Adjusted Supervisor Ratings		Teacher Ratings	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Achievement Index										
Lowest Quintile	0.097**	0.097**	0.140***	0.102***	-0.780***	-0.543***	-0.417***	-0.543***	-0.337***	-0.201***
	(0.039)	(0.040)	(0.030)	(0.031)	(0.091)	(0.088)	(0.078)	(0.088)	(0.053)	(0.049)
Middle 60%	0.075***	0.078***	0.056***	0.049***	-0.410***	-0.297***	-0.215***	-0.297***	-0.203***	-0.156***
	(0.017)	(0.018)	(0.017)	(0.016)	(0.061)	(0.038)	(0.036)	(0.038)	(0.041)	(0.036)
FRPL %										
20–80%	0.007	0.026	-0.022	-0.031	-0.463**	-0.106*	-0.008	-0.106*	-0.193***	-0.201**
	(0.020)	(0.029)	(0.024)	(0.031)	(0.207)	(0.061)	(0.070)	(0.061)	(0.071)	(0.080)
80–100%	0.025	0.057*	-0.011	-0.014	-0.423*	-0.077	0.025	-0.077	-0.253**	-0.251**
	(0.029)	(0.033)	(0.046)	(0.051)	(0.220)	(0.101)	(0.093)	(0.101)	(0.109)	(0.121)
Students of Color %										
20–80%	0.011	0.005	0.023	0.013	0.052	0.099*	0.034	0.099*	-0.019	-0.148***
	(0.015)	(0.018)	(0.017)	(0.023)	(0.074)	(0.052)	(0.028)	(0.052)	(0.040)	(0.046)
80–100%	0.020	0.042	0.019	0.067**	0.023	-0.124	-0.092	-0.124	-0.105	-0.465***
	(0.030)	(0.034)	(0.035)	(0.030)	(0.141)	(0.090)	(0.087)	(0.090)	(0.083)	(0.077)
Locale %										
Urban	0.010		0.031		-0.033		0.152***		0.018	
	(0.022)		(0.025)		(0.109)		(0.051)		(0.058)	
Town/Rural	0.039**		0.011		-0.022		0.033		0.093*	
	(0.019)		(0.024)		(0.140)		(0.031)		(0.051)	
District FE		✓		✓		✓		✓		✓
N	17171	17171	17171	17171	8679	8679	8678	8678	8174	8174
R^2	0.008	0.037	0.013	0.053	0.092	0.328	0.028	0.040	0.031	0.086

Notes: Standard errors clustered by district shown in parentheses. Omitted categories are highest quintile of achievement index, 0–20% FRPL, 0–20% nonwhite, and suburban locale. When estimating models with district fixed effects, we drop locale as only a handful of districts have within-district variation in school locale.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix B

Factor Analysis for Teacher Ratings Measure

Table B1
Tennessee, 2011–12

Questionnaire Item	Factor Loading	Uniqueness
The principal at my school monitors student academic progress.	0.79	0.37
The principal at my school interacts regularly with students about their learning.	0.75	0.44
The principal at my school presses teachers to implement what they have learned in professional development.	0.75	0.44
The principal at my school communicates a clear vision for this school.	0.87	0.24
The principal at my school sets high standards for student learning.	0.89	0.20
The principal at my school sets high standards for teaching.	0.88	0.23
The principal at my school makes clear to the staff his or her expectations for meeting instructional goals.	0.88	0.23

N = 2,442. Eigenvalue for single factor = 4.9

Table B2
Tennessee, 2012–13

Questionnaire Item	Factor Loading	Uniqueness
The principal at my school monitors student academic progress.	0.77	0.41
The principal at my school interacts regularly with students about their learning.	0.78	0.39
My principal is doing a good job.	0.91	0.17
The principal at my school presses teachers to implement what they have learned in professional development.	0.77	0.40
The principal at my school communicates a clear vision for this school.	0.88	0.23
I am pleased with the way my principal runs this school.	0.90	0.18
The principal at my school sets high standards for student learning.	0.87	0.24
The principal at my school sets high standards for teaching.	0.85	0.28
I would be happy to continue working with my principal in the future.	0.89	0.22
The principal at my school makes clear to the staff his or her expectations for meeting instructional goals.	0.89	0.21
The principal at my school is available to teachers to discuss teacher evaluation results.	0.81	0.34

N = 2,799. Eigenvalue for single factor = 7.9

Table B3
Tennessee, 2013–14

Questionnaire Item	Factor Loading	Uniqueness
The principal at my school monitors student academic progress.	0.78	0.39
The principal at my school interacts regularly with students about their learning.	0.78	0.38
My principal is doing a good job.	0.89	0.19
The principal at my school presses teachers to implement what they have learned in professional development.	0.77	0.40
The principal at my school communicates a clear vision for this school.	0.88	0.22
I am pleased with the way my principal runs this school.	0.90	0.19
The principal at my school sets high standards for student learning.	0.88	0.23
The principal at my school sets high standards for teaching.	0.87	0.25
I would be happy to continue working with my principal in the future.	0.88	0.22
The principal at my school makes clear to the staff his or her expectations for meeting instructional goals.	0.89	0.21
The principal at my school is available to teachers to discuss teacher evaluation results.	0.81	0.34

N = 3,620. Eigenvalue for single factor = 8.0

Table B4
Tennessee, 2014–15

Questionnaire Item	Factor Loading	Uniqueness
The staff feels comfortable raising issues and concerns that are important to them with school leaders.	0.84	0.30
There is an atmosphere of trust and mutual respect within this school.	0.86	0.27
Teachers are held to high professional standards for delivering instruction.	0.56	0.69
The teachers at this school like being here; I would describe us as a satisfied group.	0.78	0.39
I feel appreciated for the job that I am doing.	0.81	0.34
School leadership consistently supports the school staff.	0.90	0.20
School leadership makes a sustained effort to address staff concerns.	0.89	0.21
School leadership provides useful feedback about my instructional practices.	0.78	0.39

N = 31,905. Eigenvalue for single factor = 5.2

Table B5
Tennessee, 2015–16

Questionnaire Item	Factor Loading	Uniqueness
The staff feels comfortable raising issues and concerns that are important to them with school leaders.	0.84	0.30
There is an atmosphere of trust and mutual respect within this school.	0.85	0.28
Administrators hold teachers to high professional standards for delivering instruction.	0.67	0.55
The staff at this school like being here; I would describe us as a satisfied group.	0.78	0.38
I feel appreciated for the job that I am doing.	0.82	0.33
School leadership is adequately visible and available to address staff/student needs.	0.86	0.26
School leadership proactively seeks to understand the needs of teachers and staff.	0.91	0.17
School leadership makes a sustained effort to address staff concerns.	0.91	0.17
School leadership provides useful feedback about my instructional practices.	0.82	0.32
I frequently have the opportunity to receive feedback on my practices from multiple sources.	0.75	0.43

N = 27,357. Eigenvalue for single factor = 6.8

Table B6
Tennessee, 2016–17

Questionnaire Item	Factor Loading	Uniqueness
The staff feels comfortable raising issues and concerns that are important to them with school leaders.	0.84	0.29
There is an atmosphere of trust and mutual respect within this school.	0.86	0.26
The staff at this school like being here; I would describe us as a satisfied group.	0.81	0.35
I feel appreciated for the job that I am doing.	0.84	0.30
I like the way things are run at this school.	0.89	0.21
I receive sufficient support toward the achievement of my long-term career goals from leaders at my school.	0.84	0.29
School leadership is adequately visible and available to address staff/student needs.	0.84	0.29
School leadership proactively seeks to understand the needs of teachers and staff.	0.91	0.18
School leadership makes a sustained effort to address staff concerns.	0.91	0.18

N = 33,355. Eigenvalue for single factor = 6.7

Table B7
2011–12 SASS Teacher Survey

Questionnaire Item	Factor Loading	Uniqueness
The school administration's behavior toward the staff is supportive and encouraging.	0.87	0.24
My principal enforces school rules for student conduct and backs me up when I need it.	0.87	0.24
The principal knows what kind of school he or she wants and has communicated it to the staff.	0.87	0.25
In this school, staff members are recognized for a job well done.	0.84	0.30
I like the way things are run at this school.	0.86	0.26

N = 37,500 (rounded to nearest 10 per NCES policy). Eigenvalue for single factor = 3.7

Online Appendix C

Simulation Details

The purpose of the simulation is to evaluate the relative importance of the different processes in driving principal quality gaps. Our simulation design draws heavily from Goldhaber et al. (2018), who analyze teacher quality gaps. Besides looking at principals instead of teachers, we make two main adjustments. First, we examine principal quality gaps using three categories of principals and schools, instead of two. Second, rather than assuming principal quality is fixed, we allow principals to change over time. Our school groups correspond to the achievement categories in the main analysis: lowest quintile, 2nd–4th quintiles, and highest quintile of average student achievement level. Additionally, we make this measure time-invariant for each school by averaging over all available years of test score data (2007–2017). We conduct separate simulations based on two measures of principal quality. First, we categorize by prior principal experience: 0–2 years, 3–5 years, and 6 or more years. Second, we categorize by percentile rank based on supervisor ratings in the current year: lowest quartile, 2nd and 3rd quartiles, and highest quartile.

The principal quality gap we focus on is the percentage point difference between the proportion of "low-quality" principals in low-achievement (middle-achievement) versus high-achievement schools. We can explain this total gap through the following components: exits from the education system, transfers (both within and across districts), promotions to central office, demotions to a non-principal school-level position, new hires, and stayers. The parameters used in each simulation are taken from the observed rates in the Tennessee.

Our simulation follows a Markov process derived from the time-invariant probabilities calculated from the Tennessee data. The simulation begins (year 0) with an equal distribution of principal quality across school groups (i.e., the proportion of high-quality, medium-quality, and low-quality principals is the same in each school group). The distribution in year 1 is determined by the distribution in year 0 and the time-invariant probabilities that dictate turnover, hiring, and reclassification. We end the simulation after year 10.

Specifically, the number of principals of each type (henceforth referred to as low-quality, medium-quality, and high-quality) in a given school group (low-achievement, middle-achievement, and high-achievement) is calculated as:

$$\begin{aligned} Total = & Stayers - (TransfersOut + Exits + Promotions + Demotions) \\ & + TransfersIn + NewHires \end{aligned} \tag{1}$$

Stayers are principals who remain in the same school category (low-achievement, middle-achievement, high-achievement) between year t and year $t+1$. $TransfersOut$ are principals who move to a school in a different category (e.g., middle-achievement to high-achievement). *Exits*, *Promotions*, and *Demotions* are no longer principals in year $t+1$. $TransfersIn$ are principals who move into the given school category from a different category. The number of principals in a school group is fixed in the simulation, so the net loss in principals is filled by $NewHires$.

For example, the number of low-quality (i.e., 0–2 years of principal experience or

bottom quartile supervisor rating) principals in low-achievement schools is calculated as follows (the calculations for other combinations of principal quality and school achievement are equivalent):

$$\begin{aligned} Total^{LQ}_{LA,t+1} = & Stayers^{LQ}_{LA,t} \\ & - (Transfers^{LQ}_{LA\rightarrow MA,t} + Transfers^{LQ}_{LA\rightarrow HA,t}) \\ & - (Exits^{LQ}_{LA,t} + Promotions^{LQ}_{LA,t} + Demotions^{LQ}_{LA,t}) \\ & + \left(Transfers^{LQ}_{MA\rightarrow LA,t} + Transfers^{LQ}_{HA\rightarrow LA,t}\right) \\ & + NewHires^{LQ}_{LA,t+1} \end{aligned} \tag{2}$$

We can simplify this to the sum of two groups:

$$Total^{LQ}_{LA,t+1} = Returners^{LQ}_{LA,t+1} + \ NewHires^{LQ}_{LA,t+1} \tag{3}$$

where Returners is the sum of stayers and transfers into the given school group. The total number of new hires in each school group is equal to the net loss of principals in the prior year. The number of low-quality new hires in low achievement schools is the total number of new hires in low-achievement schools multiplied by the probability that a new hire is low-quality:

$$NewHires^{LQ}_{LA,t+1} = NewHires_{LA,t+1} * \Pr\left(LQ|NewHire\right)_{LA} \tag{4}$$

For returners, we must also make an adjustment for changes in classification (i.e., experience gains or changes in supervisor ratings) between year t and year t+1. Thus, the number of returners becomes:

$$\begin{aligned} Returners^{LQ}_{LA,t+1} = & \\ & Returners^{LQ}_{LA,t} \\ & * \left(1 - Pr\left(Reclassification\right)^{LQ\rightarrow MQ}_{LA} \ - Pr\left(Reclassification\right)^{LQ\rightarrow HQ}_{LA} \ \right) \\ & + \left(\ Returners^{MQ}_{LA,t} * Pr\left(Reclassification\right)^{MQ\rightarrow LQ}_{LA}\right) \\ & + \left(Returners^{HQ}_{LA,t} * Pr\left(Reclassification\right)^{HQ\rightarrow LQ}_{LA} \ \right) \end{aligned} \tag{5}$$

which captures reclassification out of a group (i.e., low-quality to medium-quality and low-quality to high-quality) and reclassification in to a group (e.g., medium-quality to low-quality and high-quality to low-quality). When experience is the principal quality measure, this process is simplified, as all principals gain an additional year of experience. Thus, we only examine differential reclassification when examining gaps in supervisor ratings.

To isolate the specific contribution of a process, we hold constant all other processes by using overall average rates, rather than group-specific rates. For instance, if we want to examine the extent to which differences in principal exit rates among high-achievement and

low-achievement schools contributes to the principal quality gap, we adjust the simulation so that the only difference between high-achievement and low-achievement schools is their principal exit rates; rates of transfer, demotion, promotion, reclassification, and hiring are set to the overall rate.

C.1 Exit Rates

Exiters are principals who leave the K–12 public education system after year t. In our simulation, exiting principals are replaced by a new hire in year $t + 1$. The exit rates used (shown below) are the average rates in Tennessee (2007 to 2016 for experience and 2012 to 2016 for principal ratings).

Table C1
Exit Rates by Principal Experience and School Achievement

	0–2 years	3–5 years	6+ years
Overall	0.045	0.068	0.108
Low Achievement	0.064	0.088	0.133
Middle Achievement	0.039	0.065	0.105
High Achievement	0.039	0.056	0.093

Table C2
Exit Rates by Principal Ratings and School Achievement

	Q1 Rating	Q2/Q3 Rating	Q4 Rating
Overall	0.084	0.062	0.083
Low Achievement	0.119	0.078	0.098
Middle Achievement	0.067	0.061	0.079
High Achievement	0.082	0.051	0.086

C.2 Promotion and Demotion Rates

Promotions and demotions are principals who move out of the principalship after year t but remain in the K–12 education system in Tennessee. Like with exits, promoted and demoted principals are replaced by a new hire in year $t + 1$. The promotion and demotion rates used (shown below) are the average rates in Tennessee (2007 to 2016 for experience and 2012 to 2016 for principal ratings).

Table C3
Promotion Rates by Principal Experience and School Achievement

	0–2 years	3–5 years	6+ years
Overall	0.022	0.035	0.035
Low Achievement	0.018	0.034	0.024
Middle Achievement	0.021	0.035	0.038
High Achievement	0.022	0.031	0.036

Table C4
Promotion Rates by Principal Ratings and School Achievement

	Q1 Rating	Q2/Q3 Rating	Q4 Rating
Overall	0.024	0.029	0.051
Low Achievement	0.021	0.017	0.034
Middle Achievement	0.028	0.030	0.053
High Achievement	0.011	0.034	0.056

Table C5
Demotion Rates by Principal Experience and School Achievement

	0–2 years	3–5 years	6+ years
Overall	0.045	0.032	0.021
Low Achievement	0.052	0.046	0.044
Middle Achievement	0.047	0.033	0.019
High Achievement	0.022	0.015	0.008

Table C6
Demotion Rates by Principal Ratings and School Achievement

	Q1 Rating	Q2/Q3 Rating	Q4 Rating
Overall	0.063	0.021	0.015
Low Achievement	0.102	0.029	0.030
Middle Achievement	0.059	0.022	0.016
High Achievement	0.018	0.010	0.009

C.3 Transfer Rates

Transfers are principals who move between principal positions in year t and year $t + 1$. For our simulation, we distinguish between transfers to the same category of schools (e.g., a principal who moves from a low-achievement school to a different low-achievement school) and transfers to a different category (e.g., a principal who moves from a low-achievement school to a high-achievement school). By definition, only across-group transfers affect principal quality gaps, though we acknowledge that high rates of within-group transfer could have negative effects on schools that we will not capture in our simulation. The transfer rates used (shown below) are the average rates in Tennessee (2007 to 2016 for experience and 2012 to 2016 for principal ratings).

Table C7
Transfer Rates to Low Achievement Schools by Principal Experience and School Achievement

	0–2 years	3–5 years	6+ years
Overall	0.011	0.012	0.008
Low Achievement	0.033	0.041	0.036
Middle Achievement	0.005	0.005	0.004
High Achievement	0.002	0.005	0.001

Table C8
Transfer Rates to Middle Achievement Schools by Principal Experience and School Achievement

	0–2 years	3–5 years	6+ years
Overall	0.022	0.022	0.017
Low Achievement	0.016	0.021	0.017
Middle Achievement	0.025	0.024	0.021
High Achievement	0.012	0.015	0.008

Table C9
Transfer Rates to High Achievement Schools by Principal Experience and School Achievement

	0–2 years	3–5 years	6+ years
Overall	0.009	0.012	0.007
Low Achievement	0.004	0.008	0.002
Middle Achievement	0.009	0.011	0.005
High Achievement	0.012	0.020	0.013

Table C10
Transfer Rates to Low Achievement Schools by Principal Ratings and School Achievement

	Q1 Rating	Q2/Q3 Rating	Q4 Rating
Overall	0.010	0.010	0.008
Low Achievement	0.028	0.043	0.038
Middle Achievement	0.005	0.005	0.004
High Achievement	0.004	0.001	0.004

Table C11
Transfer Rates to Middle Achievement Schools by Principal Ratings and School Achievement

	Q1 Rating	Q2/Q3 Rating	Q4 Rating
Overall	0.021	0.021	0.023
Low Achievement	0.014	0.022	0.015
Middle Achievement	0.027	0.025	0.026
High Achievement	0.011	0.008	0.020

Table C12
Transfer Rates to High Achievement Schools by Principal Ratings and School Achievement

	Q1 Rating	Q2/Q3 Rating	Q4 Rating
Overall	0.003	0.011	0.009
Low Achievement	0.000	0.007	0.008
Middle Achievement	0.003	0.009	0.011
High Achievement	0.011	0.016	0.007

C.4 New Hires

When a principal exits, changes positions (promotion or demotion), or transfers to a school in a different category, a newly hired principal must replace them. The probability of hiring an experienced or highly rated principal varies by school type. The probabilities used (shown below) are the average proportions of new hires in each experience/rating category in Tennessee (2007 to 2017 for experience and 2012 to 2017 for principal ratings). To be specific, the rates reflect the proportion of new hires that have a given amount of prior principal experience, or the proportion of new hires whose ratings in the first year at the school fall into the given categories (bottom 25%, middle 50%, or top 25%).

Table C13
Hiring Rates by Principal Experience and School Achievement

	0–2 years	3–5 years	6+ years
Overall	0.884	0.075	0.041
Low Achievement	0.851	0.105	0.044
Middle Achievement	0.900	0.063	0.037
High Achievement	0.874	0.084	0.042

Table C14
Hiring Rates by Principal Ratings and School Achievement

	Q1 Rating	Q2/Q3 Rating	Q4 Rating
Overall	0.452	0.414	0.134
Low Achievement	0.476	0.404	0.120
Middle Achievement	0.458	0.408	0.134
High Achievement	0.408	0.457	0.136

C.5 Reclassification Rates

The majority of principals remain in the same school between year t and year $t+1$. However, we do not assume that their quality is fixed over time. Principals gain experience and their ratings fluctuate from year to year. Thus, we must incorporate into our simulation changes in quality among principals who remain in their schools. Transferring principals can also change in quality, though in this simulation we do not differentiate between changes in quality between "stayer" and "movers". Instead, our applies a "reclassification rate" to all principals who remain in a principal position in year t+1. For our experience simulation, reclassification is straightforward (all stayers gain an additional year of experience) and does not vary by school type. For supervisor ratings, however, we do observe differential rates of reclassification (e.g., moving from the bottom quartile to the middle quartiles) by school type. Reclassification rates, then, are an additional component of the principal quality gap. The reclassification rates used (shown below) are the average rates in Tennessee (2012 to 2017).

Table C15
Reclassification Rates by Principal Ratings and School Achievement

	Q1 to Q2/Q3	Q1 to Q4	Q2/Q3 to Q1	Q2/Q3 to Q4	Q4 to Q1	Q4 to Q2/Q3
Overall	0.285	0.002	0.058	0.093	0.004	0.170
Low Achievement	0.258	0.003	0.088	0.088	0.009	0.171
Middle Achievement	0.281	0.002	0.059	0.091	0.003	0.196
High Achievement	0.337	0.000	0.035	0.103	0.002	0.115