Strategic Retention: Principal Effectiveness and Teacher Turnover in Multiple-Measure Teacher Evaluation Systems

Jason A. Grissom
Brendan Bartanen
Vanderbilt University

***

Pre-print version. Later published as:

Grissom, J. A., & Bartanen, B. (2019). Strategic retention: Principal effectiveness and teacher turnover in multiple-measure teacher evaluation systems. *American Educational Research Journal*, *56*(2), 514-555.

***

Studies link principal effectiveness to lower average rates of teacher turnover. However, principals need not target retention efforts equally to all teachers. Instead, strong principals may seek to strategically influence the composition of their school's teaching force by retaining high performers and *not* retaining lower performers. We investigate such strategic retention behaviors with longitudinal data from Tennessee. Using multiple measures of teacher and principal effectiveness, we document that indeed more effective principals see lower rates of teacher turnover, on average. Moreover, this lower turnover is concentrated among high-performing teachers. In contrast, turnover rates of the lowest-performing teachers, as measured by classroom observation scores, increase substantially under higher-rated principals. This pattern is more apparent in advantaged schools and schools with stable leadership.

***

Research suggests that teacher turnover has negative effects on school performance, particularly for schools that are low-achieving and serve larger high-needs populations (Ronfeldt, Loeb, & Wyckoff, 2012; Hanushek, Rivkin, & Schiman, 2016). Motivated by this inverse relationship between turnover and school outcomes and an assumption that teacher turnover rates are higher than optimal (e.g., Alliance for Excellent Education, 2014), numerous studies have investigated the factors that lead teachers to leave their schools or the profession altogether (see Borman & Dowling, 2008; Grissom, Viano, & Selin, 2016; Guarnino, Santibañez, & Daley, 2006). Studies find that the quality of the school's leadership is among the most important predictors of whether a teacher continues in the school from year-to-year (e.g., Boyd et al., 2011; Grissom, 2011; Ladd, 2011). This research suggests that more effective principals are better able to retain teachers because they create more positive school climates, supply teachers with greater

support, provide more beneficial opportunity for professional growth, and otherwise positively shape teachers' working conditions in ways that lead to greater job satisfaction and attachment (Ingersoll, 2001; Johnson, Kraft, & Papay, 2012; Kraft, Marinell, & Yee, 2016).

An important characteristic of prior studies linking leadership to teacher retention is that they find that more effective leadership is associated with lower turnover for the *average* teacher. From a school improvement standpoint, however, all teacher turnover is not created equal. Although schools no doubt benefit from retaining effective teachers, they likely also benefit from failing to retain teachers who are ineffective (Adnot et al., 2016; Hanushek, 2009; Ronfeldt et al., 2012; Staiger & Rockoff, 2010). We refer to this phenomenon—the retention of effective teachers coupled with the non-retention of ineffective teachers—as *strategic retention.* Past work has found that high-growth schools indeed have higher retention rates among high-performing teachers and higher turnover rates among low performers (Loeb, Béteille, & Kalogrides, 2012).

As the chief human capital managers in their schools, principals are uniquely positioned to pursue differential retention strategies. Personnel decisions are an important channel through which principals can influence student achievement (Branch, Hanushek, & Rivkin, 2012; Jacob, 2011), and several studies provide evidence that principals can act strategically in the area of personnel management to try to improve school outcomes, particularly when they have access to good information about teacher performance (e.g., Cohen-Vogel, 2011; Goldring et al., 2014; Grissom, Loeb & Nakashima, 2014; Rockoff et al., 2010). Effective principals may pursue strategies to reduce the overall teacher turnover rate in their schools, but we expect that they may work particularly to lower turnover among their high performers. In contrast, effective principals may find ways to encourage turnover among low performers, either through administrative

means, such as contract nonrenewal, or through less formal means, such as "counseling out" or finding ways to make the job less palatable (Drake et al., 2016).

Investigating strategic teacher retention is especially timely in the context of the widespread adoption of multiple-measure teacher evaluation systems across the United States in recent years (Steinberg & Donaldson, 2016). These systems typically incorporate both school leaders' observations of classroom practice using a detailed instructional rubric and achievement-based metrics (such as value-added), sometimes supplemented with other measures (Grissom & Youngs, 2016). Facilitation of strategic retention is one motivation for these systems, which are predicated on the idea that differentiating teacher performance—which prior, less rigorous evaluation systems failed to do (Weisberg et al., 2009)—can provide schools opportunities to tailor approaches to different teachers, including merit pay, career ladders, improvement plans, and dismissal (Steinberg & Donaldson, 2016).[1] Often ignored in policy discourse around teacher evaluation, principals are both important targets and important facilitators of these systems. They are targets in the sense that requirements for frequent, detailed teacher observations push principals to collect and process their own evidence about their teachers' effectiveness. They are facilitators in that the implementation capabilities and investment choices of the principal directly influence both the quality of the collected data (in the form of observation scores) and the local response to evaluation information (see Grissom & Loeb, 2017; Kraft & Gilmour, 2016; Reinhorn, Johnson, & Simon, 2017). These roles suggest that principal effectiveness may be an important prerequisite for teacher evaluation systems to realize their potential to shape the effectiveness of a school's teacher workforce.

[1] Recent research questions the degree to which newer systems of teacher evaluation are appreciably better at differentiating teacher performance than the systems they replaced in many states (e.g., Kraft & Gilmour, 2016).

In this article, we investigate whether schools led by more effective principals show evidence of strategic retention. We leverage longitudinal administrative data from Tennessee spanning the years 2011–12 to 2016–17, which correspond to the first six years of the implementation of TEAM, the state's principal and teacher evaluation system. TEAM collects multiple measures of performance for both principals and teachers, including both rubric-based observation measures and measures based on student achievement. The evaluation system thus provides us with rich data for operationalizing both principal and teacher effectiveness, which we further augment—for principals—with additional measures from low-stakes surveys that ask Tennessee teachers to rate leadership quality in their schools. Using these measures of principal and teacher effectiveness, we address several research questions. First, replicating prior work, we ask whether principal effectiveness is associated with lower teacher turnover on average. Second, we ask to what extent the association between principal effectiveness and teacher turnover differs for high-performing and low-performing teachers, and, furthermore, whether these associations vary by different measures of teacher performance. Third, we ask whether and how these patterns differ by other characteristics of teachers (e.g., novice vs. veteran teachers) and their schools (e.g., high-achieving vs. low-achieving schools).

We answer these questions using a school fixed effects approach to exploit within-school variation in principal effectiveness over time, which helps eliminate some (but not all) threats to interpreting our estimates as causal. In contrast to most prior studies, which have relied on teacher perceptions of principals from surveys to operationalize principal effectiveness, our main analyses make use of ratings assigned to principals using a rubric scored by supervisors trained to rate principal effectiveness against state standards, which may provide more accurate measures of their performance. Our results contribute to the research base not only on teacher

mobility and retention but to the literatures on the talent management role of principals, which is increasingly recognized as a key component of principal work (Cannata et al., 2017; Goldring et al., 2014), and on the role of teacher evaluation systems in shaping the teacher workforce.

STRATEGIC TEACHER RETENTION

Numerous quantitative studies going back nearly two decades identify the effectiveness of school leadership as an important predictor of whether teachers stay or leave their schools (e.g., Griffith, 2004; Ingersoll, 2001; Stockard & Lehman, 2004). Many of these studies use data from teacher surveys to operationalize leader effectiveness, then use those perceptual measures to predict the likelihood of turnover. For example, Boyd et al. (2011) find that beginning teachers in New York City are less likely to leave their positions when they have more positive perceptions of school administrators, and when their colleagues do as well. Moreover, in a follow-up survey, nearly half of the teachers who left their position identified "support from administrators" as the most important reason for their decision. Likewise, Ladd's (2011) study using teacher working conditions survey data from North Carolina finds that teachers working in schools with better leadership—as measured by aggregate responses to leadership perceptions questions among teachers in the school—express lower intent to leave and, in fact, turn over less often, controlling for other factors. Grissom's (2011) analysis of data from the nationally representative Schools and Staffing Survey similarly finds that principal effectiveness, measured via school-level averages of teachers' ratings of leadership, is associated with increased teacher satisfaction and a decreased likelihood of teacher turnover. These associations were larger in schools with larger numbers of historically marginalized students. In another analysis of New York City data—this time longitudinally—Kraft, Marinell, and Yee (2016) again find teachers'

perceptions of leadership quality to predict lower rates of teacher turnover, more so than other school climate factors, such as quality of teacher relationships and collaboration.

Numerous mechanisms may link more effective leadership to lower turnover, but explanations in most studies center on effective leaders' positive impacts on school climate (Burkhauser, 2017), which in turn improves teachers' satisfaction and commitment to the school and makes it less likely that they leave. In this vein, early-career teachers interviewed by Johnson and Birkeland (2003) who chose to stay in their schools reported having principals who fostered supportive cultures and found ways to encourage teachers and recognize their efforts, while non-stayers instead tended to describe their principals as absent, controlling, overly critical, and not attuned to teacher support. Similarly, Allensworth, Ponisciak, and Mazzeo (2009) found that teachers who reported a high level of trust of their principal and viewed their principal as a strong instructional leader were substantially more likely to remain in their schools.

Although we do not discount the importance of the connections among more effective leadership, more positive school climates, and lower average teacher turnover rates, we propose that effective school leaders may pursue additional strategies to improve the quality of their teaching staffs, presumably in pursuit of more positive student outcomes. In particular, we suggest that more effective leaders may be more likely to target their retention efforts at more effective teachers while simultaneously working to *not* retain ineffective teachers. Such *strategic retention* strategies likely are more likely to lead to gains in student outcomes than strategies aimed at retaining all teachers. To this point, Loeb et al.'s (2012) study of Miami-Dade County Public Schools found that high-growth schools were more successful both at retaining teachers with high value-added to student achievement and at turning over teachers with low value-added.

The hypothesis that more effective principals are more likely to selectively retain teachers rests on a number of assumptions about what leaders know and can do with respect to teacher retention. In particular, strategic retention presumes that principals can identify their high and low performers, that they can influence retention decisions among teachers they wish to retain, and that they have the means to release or influence turnover among teachers they wish to not continue in their schools. We discuss each of these in turn.

Identifying High and Low Performers

Differentiated retention strategies targeted at high- and low-performing teachers rely on principals' capacities to tell one type of teacher from the other. Even prior to the recent widespread implementation of multiple measures-based teacher evaluation systems, studies suggested that principals could differentiate teachers by performance (Harris & Sass, 2014; Jacob & Lefgren, 2008; Rockoff et al., 2012). In the post-Race to the Top era, when more rigorous teacher evaluation systems that make multiple measures of teacher performance accessible to principals have become commonplace, principals have even more information at their disposal for differentiating between high- and low-performing teachers (see Grissom & Youngs, 2016). More frequent principal observations of teachers using a standardized rubric are a hallmark of these new systems, providing principals with more opportunities to observe teacher practice and, in many cases, facilitating their engagement with other measures of teacher effectiveness (Goldring et al., 2014). As with prior studies in low-stakes settings, studies of principals' formal evaluation ratings of teacher practice show that they can accurately differentiate teacher performance (Grissom & Loeb, 2017; Sartain, Stoelinga, & Brown, 2011).

Principals may be particularly responsive to teacher effectiveness measures derived from their own observations in making talent management decisions. As Goldring et al. (2015)

document in a study of principals' use of these measures in six urban districts, principals trust observation data because they collect the data themselves. In contrast, they discount other measures, such as value-added, because of lack of transparency and because those measures often are returned to them too late to be useful in talent management decisions. Moreover, principals value dimensions of teacher practice that are not captured by student achievement-based measures but that often are reflected in observational rubrics (Grissom, Loeb, & Doss, 2016; Master, 2014). Thus, we expect that teacher effectiveness reflected in the observation score is more likely to drive principals' teacher retention decisions than other components of an overall evaluation score.

Targeting Retention Efforts at High-Performing Teachers

Knowing who their high-performing teachers are does not guarantee that principals have tools available to influence those teachers to stay. In few districts do principals play any role in determining teacher compensation, meaning that principals' work to retain their effective teachers must come through affecting their working conditions.[2] In several studies linking principal effectiveness to teacher turnover, effectiveness has been operationalized by teacher responses to survey questions focused on principals' impacts on working conditions, such as the degree to which principals provide support and encouragement and recognize teachers for a job well done (e.g., Grissom, 2011; Ladd, 2011). Principals' ability to retain their effective teachers likely depends on the degree to which they can differentiate their efforts to create more positive working conditions for their high performers.

[2] One exception might be in districts in which teacher incentive pay models are employed. Some schools in Tennessee, for example, have experimented with incentive pay through awards from the Teacher Incentive Fund (TIF) from the U.S. Department of Education. Including TIF school status as a covariate in the models we present later, however, did not affect our results.

A 2012 report from TNTP suggested that principals indeed could differentiate such efforts and recommended several straightforward strategies for retaining "irreplaceable" teachers (i.e., those in the top tier of performance), including recognizing their contributions, providing them direct encouragement to stay, giving them helpful feedback, and creating opportunities for them to show leadership. Yet in the four urban districts studied for the report, they found use of such strategies to be relatively infrequent. Other research suggests that principals may give more desirable work assignments—by, for example, matching teachers to higher achieving students—to more effective teachers or move those teachers into instructional coach or other leadership roles, suggesting some other strategies through which principals might seek to encourage retention among their high performers (Chingos & West, 2011; Grissom, Kalogrides, & Loeb, 2014; Kalogrides, Loeb, & Béteille, 2013; Player, 2010). Other efforts to retain effective teachers may be indirect. For instance, creating systems and processes to remove *low-performing* teachers may indirectly encourage high performers to stay put by emphasizing the value the school places on high-quality teaching (TNTP, 2012).

Targeting Release Efforts at Low-Performing Teachers

Strategic retention also requires that principals have the capacity to move ineffective teachers out of their schools when they identify them. Principals can induce low-performing teachers to turn over through formal or informal means.

Formal means are administrative processes by which a teacher can be removed from a school or district. Typically, schools have more discretion to dismiss low-performing early-career teachers who are untenured or on probationary status, while the administrative processes surrounding dismissal of a tenured or non-probationary teacher often make (or are perceived to make) forced removal of a more veteran teacher too difficult or costly except in extreme cases

(Weisberg et al., 2009; Gordon, Kane, & Staiger, 2006; Painter, 2000). Recent reforms in several states, including Tennessee and Florida, have sought to make formal teacher dismissal procedures more accessible to school and district leaders by lengthening probationary periods, conditioning tenure on performance, and eliminating some due process rights associated with tenure, among other changes (Wesson, 2012; Harrison & Cohen-Vogel, 2012). Even in the wake of these changes, however, utilization of formal dismissal procedures appears limited, with few teachers in reform states removed for poor performance (Kraft & Gilmour, 2016).

Other evidence suggests, however, that, when given the tools, principals can utilize formal processes to move teachers out of their schools. For example, Jacob (2011) found that principals in Chicago Public Schools were less likely to renew teachers with higher absence rates, lower performance ratings, and lower value-added. In Master's (2014) study of an urban charter district, principals were substantially more likely to dismiss teachers to whom they had given low mid-year ratings. Grissom, Loeb, and Nakashima (2014) found that principals in low-achieving schools in Miami who were given the option to nominate teachers to be involuntarily transferred out of their schools used the policy to move out veteran teachers with low value-added and higher absence rates and replace them with higher performers. In another study, principals reported using data from their teacher evaluation system to place struggling teachers on formal intervention or assistance plans that would lead to dismissal if teachers did not respond to the plans and improve (Drake et al., 2016). The stigma associated with being on a professional improvement plan can be enough to lead ineffective teachers to leave (Cohen-Vogel, 2011).

Principals may make use of informal means to remove ineffective teachers as well. "Counseling out" by, for example, discussing with the teacher that they may be more successful in another environment, is one such strategy (TNTP, 2012), though principals in Grissom and

Loeb's (2011) study of principal management skills rated their ability to counsel out teachers among the lowest of fifty skill areas they assessed. Principals also may implement a strategy of changing teachers' work assignments to make their jobs less desirable (TNTP, 2012), making it more likely that the teacher voluntarily exits. In an experiment in which principals were provided with teacher performance information, turnover among low-performing teachers in the treatment schools increased (Rockoff et al., 2012), evidence that principals indeed utilized the information and available mechanisms to push teachers out. Likewise, other studies have found that teachers are more likely to resign when assigned low evaluation ratings (Master, 2014).

Principal Effectiveness and Strategic Retention

Effective teachers are more likely to stay in their schools, and less effective teachers are more likely to move or exit (e.g., Boyd et al., 2008; Goldhaber, Gross, & Player, 2011). Researchers have not yet examined whether these patterns are moderated by the quality of the school's principal, however. We argue that there are good reasons to suspect that they might. Effective principals produce higher student outcomes (Coelli & Green, 2012; Grissom, Kalogrides & Loeb, 2015), and strategies to improve the effectiveness of teachers in the school are likely important to their approach (Cohen-Vogel, 2011; Loeb et al., 2012). To the degree that more effective principals are more engaged in and serious about teacher observation and evaluation (Goldring et al., 2014), they likely have better information with which to differentiate high- and low-performing teachers. They also are more likely to act on that information in strategic ways; to this point, Donaldson (2013) found that principals with greater human capital in the form of ingenuity, motivation, and training in instructional improvement were more likely to circumnavigate obstacles to increasing teacher effectiveness in their buildings through hiring, dismissal, and other talent management functions.

More generally, even in the absence of specific targeted strategies for retaining high performers and removing low performers, effective instructional leaders establish school cultures marked by high expectations and consistent, useful feedback to teachers on their performance (e.g., Hallinger, 2005; Neumerski, 2013). Such school cultures likely are especially conducive to increasing job satisfaction and commitment among teachers who excel and to signaling a lack of fit to ineffective teachers (TNTP, 2012).

DATA AND MEASURES

This study analyzes administrative data from Tennessee, a state made up of 146 districts operating roughly 1,800 schools that serve 996,000 students. Thirty-one percent of the state's students are black or Hispanic, and 58% are eligible for the federal subsidized lunch program.[3] Tennessee was a first-round winner of the Obama administration's Race to the Top competition and instituted a number of educational reforms under its auspices. Particularly useful for this study, these reforms included implementation of a statewide educator evaluation system, the Tennessee Educator Acceleration Model (TEAM), beginning in the 2011–12 school year.[4] We make use of a multiple measures from the TEAM system to measure the effectiveness of teachers and principals, as we describe below.

An important piece of Tennessee context is that dismissal or release of teachers is substantially easier for districts than in most other states. A law passed in April 2011 ensured that districts have wide latitude to dismiss teachers deemed ineffective (Tenn. Code. Ann. § 49-5-501–515). During the period we study, teachers in the first five years of their teaching careers in Tennessee ("probationary" teachers) could be let go without cause; districts simply could elect

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

[4] A small number of districts use an alternative rubric for teacher observations, though we find that average scores and standard deviations are very similar across rubrics. Our results are robust to the exclusion of these teachers.

not to renew their yearly employment contract. Even after entering the "non-probationary" period, teachers could be reverted to year-to-year contract status with two consecutive years of low overall evaluations. Permanent tenure status for teachers who already had tenure at the time of the law change grandfathered, meaning those teachers could only be let go for cause and after due process.[5] Districts differ in how they approach non-renewal of teacher contracts, but principals have relatively broad discretion to release teachers—particularly early-career teachers—they deem ineffective or a poor fit.

For this analysis, the Tennessee Department of Education (TDOE) provided us with longitudinal administrative data files covering all public education personnel in the state from 2011–12 to 2016–17 via the Tennessee Education Research Alliance (TERA) at Vanderbilt University. These files contain information about employees' personal and professional characteristics, including job positions, gender, race and ethnicity, years of experience, and highest degree earned. We use these files to construct additional experience measures, such as years employed in their current school.[6] We then merge these data with information on the characteristics of the schools and districts in which teachers and principals currently work from annual student demographic, enrollment, and achievement data from TDOE.

Table 1 contains descriptive statistics for teachers and schools, including our key measures that we discuss below. The teaching force in Tennessee is predominantly female and white, with a majority having a Master's degree or higher. More than a quarter of teachers have fewer than five years of teaching experience. The average teacher works in a school where 24%

[5] One "cause" defined by statute is low performance, so even teachers with permanent tenure status technically can be let go for persistently low evaluation ratings.
[6] In fact, we use personnel files going back to 2002 to construct these experience measures.

of the students are black and 59% percent qualify for free/reduced price lunch. More than half of Tennessee's teachers work in elementary schools.

Measuring Teacher Effectiveness

Under Tennessee's statewide evaluation system, teachers' overall summative evaluation scores are comprised of the weighted average of three components: scores from formal classroom observations, student test score growth, and student achievement. Classroom observations are conducted by trained raters—typically, the school's principal—throughout the school year using the TEAM rubric, which defines levels of performance on 19 instructional indicators and four additional professionalism indicators.[7] The test score growth measure comes from the Tennessee Value-Added Assessment System (TVAAS), estimated for TDOE by the SAS Institute.[8] For teachers of tested grades and subjects, an individual growth score is determined by their students' test score gains. Teachers of untested classrooms receive a growth score based on schoolwide growth. Achievement measures are locally determined and are not consistent across schools.[9] Scores on these individual components are aggregated using weighting formulas that vary based on a teacher's subject assignment (i.e., tested or non-tested subjects) and the availability of prior student achievement data; the typical split over the years of the data is 50% from observations, 35% from student growth, and 15% from the achievement measure.[10] The weighted sum is used to assign each teacher a "Level of Overall Effectiveness" (LOE) rating on a discrete scale of 1 to 5, where 1 is "Significantly Below Expectations" and 5 is

[7] Alternative rubrics approved for use in a handful of districts cover different domains, though with substantial overlap with the content of the TEAM rubric.
[8] For teachers of tested grades and subjects, the growth score is the teacher-level value-added score. For untested teachers, the growth score is a school-level value added score. Additionally, certain teachers in tested subjects use the school-level score because they had too few tested students to produce an individual value-added score.
[9] Achievement scores come from measures of school performance, which vary widely by school. A schoolwide achievement composite is the most common measure in elementary and middle schools, while high schools most often use graduation rates or average ACT scores.
10 For more information about the TEAM evaluation system for both teachers and leaders, see http://team-tn.org.

"Significantly Above Expectations." Growth scores and LOE ratings typically are returned to teachers and schools in the early fall of the following school year.

For teachers of tested grades and subjects (approximately 44% of teachers in Tennessee), both the average classroom observation and growth scores represent individualized measures of teacher effectiveness. For the remaining majority of teachers, the average classroom score is the only available measure of individualized teacher performance. In this analysis, we use both observation and growth scores as measures of teacher effectiveness. Additionally, we examine "individual growth scores" for the relevant subset of teachers.[11] Growth scores, which range from Level 1 to Level 5, are taken directly from the evaluation files.[12] To facilitate the interpretation of non-linear relationships between observation scores and turnover, as well as to match the construction of the growth and achievement measures, we convert the continuous score to a categorical indicator with five groups.[13] The lowest category (1.00-2.75) includes teachers who are roughly two or more standard deviations below the mean. These measures of teacher effectiveness are available for approximately 58,000 teachers in each year.[14]

Table 1 shows that the average observation score is 3.91, and the standard deviation (0.58) is much smaller than that of the growth or achievement score. Relatively few teachers

[11] In comparison to most states, Tennessee has many teachers who receive individual value added scores due to yearly testing in all core subjects (math, reading, science, social studies) in grades 3-8, in addition to end-of-course tests in Algebra 1 and 2, Geometry, English 1-3, Biology, Chemistry, and US History.
[12] These categorical scores are assigned according to a teacher's underlying TVAAS index, which is continuous. The index is obtained by dividing the value-added estimate by its standard error.
[13] We operationalize the average observation as categorical to facilitate examination of non-linear associations and to match the construction of the other teacher effectiveness measures. The choice of cutoff for the bottom category categorizes the lowest 3% of teachers, which corresponds to the percentage of teachers statewide who receive a Level 1 rating for their summative evaluation score. Moving the cutoff to a lower or higher score (e.g., bottom 2% or bottom 5%) yields similar results, though lower cutoffs diminish power. Similarly, moving the cutoff for the highest category of teachers does not appreciably change the results. We also explored cutoffs (e.g., 2+ SD below the mean, 1–2 SD below the mean), and the results were substantively similar.
[14] In 2015-16, Tennessee implemented a new statewide testing system called TNReady. Due to implementation challenges, testing was cancelled in grades 3–8 and growth scores are unavailable. Thus, analyses including growth scores are limited to 2011–12 to 2014–15.

receive very low (3%) or very high (16%) observation scores. The number of teachers in the lowest observation category is roughly equal to the number of teachers who receive the lowest summative evaluation rating from the state each year. There is much greater variation in growth scores, though nearly 42% of teachers receive the highest rating (Level 5). Additionally, far more teachers receive very low growth ratings (20%) than receive very low observation scores. Consistent with prior work, there is a modest positive correlation between observation and growth scores ($r = 0.18$).[15] Finally, the average achievement score is 4.12, with nearly 57% of teachers receiving a Level 5. Given that achievement scores come exclusively from school performance,[16] we do not include them in our analysis as a measure of teacher effectiveness.

Measuring Principal Effectiveness

We employ two distinct measures of principal effectiveness. Our primary measure comes from summative supervisor practice ratings assigned to each principal each year as part of the TEAM system. These ratings, which comprise 50% of each principal's overall evaluation score, are assigned by district leaders—the superintendent or his or her designee—who have been trained to rate principals using a rubric derived from the Tennessee Instructional Leadership Standards. The rubric contains 22 indicators of effective leadership practice grouped (as of 2015–16) into 4 domains: Instructional Leadership for Continuous Improvement, Culture for Teaching and Learning, Professional Learning and Growth, and Resource Management. Importantly, the number of content of these indicators and domains has changed several times since 2011–12, though prior research on these scores shows that, regardless of the specific

[15] Unsurprisingly, the correlation is stronger for teachers with individual growth scores ($r = 0.34$).
[16] Roughly 88% of teachers receive an achievement score based on a school-based measure. The most common achievement measures are schoolwide average test scores (~25% of teachers), graduation or promotion rates (10%), and average ACT scores (5%). The remaining 12% of teachers receive an achievement score based on a grade- or classroom-level achievement score (e.g., average grade-level TCAP reading score).

indicators included on the rubric in a year, the indicators are so highly inter-correlated that they can be reduced to a single underlying performance score using factor analysis (Grissom, Blissett, & Mitani, 2018). That is, the supervisors' ratings do not differentiate areas of principal performance and instead identify a single underlying principal effectiveness construct. Given this evidence, we calculate the predicted score from this single-factor model[17] to use as a summary measure of effective leadership practice— at least from the point of view of the principal's supervisor. We refer to this score, which we standardize, as the *TEAM rating*.[18]

We supplement the TEAM supervisor rating with an additional measure of principal effectiveness from the perspective of teachers in the school. This measure is similar to those employed in other studies linking principal effectiveness to teacher turnover (e.g., Grissom, 2011; Ladd, 2011). We construct this measure from responses to the First to the Top (FTTT) Survey, conducted by researchers at Vanderbilt University in cooperation with TDOE as part of evaluation efforts associated with the state's FTTT initiatives. Teachers statewide were invited to participate in the survey in the spring of the 2011–12 through 2013–14 school years. Response rates ranged across years from 25% to 40%. A random subset of responding teachers completed a module containing a battery of questions designed to assess their principal's leadership. Items ask, for example, whether the school's principal consistently monitors student academic progress, communicates a clear school vision, and sets high standards. Using factor analysis, we again found that responses captured one latent construct, which we take to be *perception of principal effectiveness* (Grissom, Blissett, & Mitani, 2018). To obtain a principal-level score, we

[17] The average of the items is correlated with the factor score at 0.97.
[18] Due to missing data on other variables, a few principals are dropped from the analytic sample after standardization, which is why the TEAM rating reported in Table 1 is not exactly mean 0 and standard deviation 1. Also, a handful of school districts opted to use different principal evaluation processes in certain years, so approximately 20% of teacher-year observations do not have a principal with a TEAM rating.

averaged the teacher-level factor scores at each principal's school and standardized them. We refer to this score as the principal's *FTTT score*. Approximately 19% of teacher-year observations do not have an FTTT score because none of the school's teachers responded to the requisite survey items, and most schools have responses from only a few teachers, so results should be interpreted with caution.[19] The correlation between principal TEAM rating and FTTT score is a modest 0.20, suggesting that supervisors and teachers judge principals by related but distinct criteria.

Operationalizing Teacher Turnover

From the longitudinal administrative data files, we create both binary and categorical teacher turnover variables. The binary variable takes a value of 1 if a teacher leaves his or her school during or immediately following year t (i.e., they are not in the same school in year $t+1$). The categorical indicator includes four turnover types (not including teachers who stay in their schools): moving to another school in the same district, moving to a different district, changing positions (e.g., instructional coach, assistant principal, and leaving the education system.[20]

Table 2 shows simple mean yearly turnover rates for Tennessee teachers by each measure of teacher effectiveness. Panel A shows that the yearly teacher turnover rate is 13.3%. About a third of turnover events are teachers who leave the education system (4.5%), with a somewhat higher proportion of teachers moving to another school in the same district (5.5%). In comparison to exits and within-district moves, across-district moves and position changes (e.g., becoming an instructional coach or administrator) are less common (2.3% and 1.0%).

[19] Approximately 21% have complete leadership module data from only 1 teacher; only 3% have more than 5 respondents. Noise introduced by these small samples likely attenuates the association between principal effectiveness and teacher turnover. Re-estimation of the main results with different thresholds for minimum numbers of responding teachers (2+, 3+, 4+, 5+) found qualitatively similar results to those reported in the main text.
[20] Additionally, a small number of schools closed during this period. We drop teachers in these schools from the analysis in the year of the closure (~1,500 teacher-year observations). Results are not sensitive to including them.

Unsurprisingly, less effective teachers are more likely to turn over. For each of the three measures of teacher effectiveness (average observation score, growth score, and individual growth score), there is a (generally monotonic) negative relationship between effectiveness and turnover. However, the difference in turnover rates between teachers in the highest and lowest effectiveness categories is much larger for observation scores. Teachers whose average observation scores are 1.00–2.75 are more than three times more likely to turn over than teachers scoring 4.50–5.00. By comparison, for growth, the turnover rate of Level 1 teachers is only 24% greater than Level 5 teachers. When constraining growth scores to teachers in tested classrooms, the turnover gap between Level 1 and Level 5 teachers increases to 49%. Even when comparing teachers with the highest observation scores to those with merely below-average observation scores (2.75–3.50), the turnover gap is larger than between Level 1 and Level 5 growth.

METHODS

Our main analysis examines the extent to which principal effectiveness is associated with higher or lower turnover probabilities among teachers with different performance ratings. The base model is as follows:

$$\Pr(Teacher\ turnover)_{ijt} = P_{jt}\beta_1 + T_{ijt}\phi + X_{ijt}\gamma + S_{jt}\eta + \delta_j + \tau_t + \epsilon_{ijt} \quad (1)$$

where the probability that teacher i leaves their position in year t is a function of the effectiveness rating P of the principal in school j and year t, and the teacher's own effectiveness T. Additionally, we control for a vector of teacher demographics X (race, gender, age, experience, and education) and school characteristics S (achievement index, enrollment, proportion of black and Hispanic/Latino students, proportion of students qualifying for free/reduced price lunch, proportion of gifted students, and proportion of students with

disabilities).[21] The parameter of interest β_1 indicates the predicted change in the probability of teacher turnover for a one standard deviation increase in principal effectiveness (as measured by either the TEAM supervisor rating or FTTT score).

The inclusion of school fixed effects δ_j and year fixed effects τ_t is critical to obtaining unbiased estimates of β_1. Although we control for a variety of observable school characteristics, there are likely unobserved school factors that affect both principal performance and teacher turnover. To the extent that these unobserved factors are not accounted for with school characteristics, their omission from the model introduces bias. For example, a school located in a highly engaged community may attract particularly high-quality principals and teachers who are committed to remaining in the school. Unless fully accounted for by school demographic covariates, lower average teacher turnover rates will be over-attributed to the effectiveness of the principal. Therefore, we focus on the within-school variation in principal effectiveness across years[22], which controls for any time-invariant school-level heterogeneity that would otherwise bias the estimated relationship between principal effectiveness and teacher turnover. We also include year fixed effects to control for common shocks that would affect teacher turnover rates statewide in a given year, such as changes in the economy or to policies governing evaluation.

This base model estimates the adjusted correlation between principal effectiveness and turnover among all teachers in a school. To examine whether this estimated relationship changes

[21] We do not include other principal characteristics, such as experience. Principals tend to become more effective with more years of experience, such that controlling for experience might attenuate any relationship between principal effectiveness and teacher turnover. In practice, however, all results are robust to the inclusion of principal demographic characteristics (i.e., experience, tenure in school, race, and gender).

[22] Within-school variation in principal effectiveness comes both from changes in the same principal's performance across years and from principal turnover. Across the analysis period, roughly 50% of schools had a single principal, 40% had two principals, and 10% had more than two principals. The within-school standard deviation of principal TEAM rating and FTTT score is 0.62 SD and 0.71 SD, on average. The adjacent-year correlation of TEAM ratings is 0.61 (0.29 for FTTT scores).

based on a teacher's effectiveness, we modify equation 1 to include an interaction term between principal effectiveness and teacher effectiveness:

$$\Pr(Teacher\ turnover)_{ijt} = P_{jt}\beta_1 + T_{ijt}\phi + (P_{jt} * T_{ijt})\beta_2 + X_{ijt}\gamma + S_{jt}\eta + \delta_j + \tau_t + \epsilon_{ijt} \quad (2)$$

where the linear combination of β_1 and β_2 is the estimated relationship between principal effectiveness and teacher turnover across the range of teacher effectiveness. Because our measures of teacher effectiveness, T_{ijt} are categorical indicators with five groups, $\beta_1 + \beta_2$ yields five estimated marginal effects of a one standard deviation increase in principal effectiveness for teachers in each group.

We also investigate the extent to which the relationship between principal effectiveness and teacher turnover changes when we differentiate between categories of turnover. Here, we adjust equations 1 and 2 to the multinomial case and estimate the probability of each category of turnover outcome (e.g., move to another school in the district, exit from the profession) relative to the same base category, teachers who stay in their positions. Estimating a series of models separately for each of the five turnover categories is preferred to other methods (e.g., multinomial logit) because we include a large number of fixed effects (Angrist & Pischke, 2009).

RESULTS

Principal Effectiveness and Average Teacher Turnover

Table 3 displays the results from estimating versions of equation 1 with observation and growth scores as measures of teacher effectiveness. For each measure, we estimate four models. Columns 1 and 5 include only year fixed effects, columns 2 and 6 add district fixed effects, and columns 3 and 7 replace district fixed effects with school fixed effects (our preferred

specification). Additionally, columns 4 and 8 show the school fixed effects results using the FTTT score as the measure of principal effectiveness.[23]

Across models, the estimated coefficients for the principals' TEAM ratings are relatively stable. On average, principals with higher TEAM ratings experience lower rates of average teacher turnover, which is consistent with prior findings (e.g., Boyd et al., 2011; Grissom, 2011; Ladd, 2011). Columns 3 and 7, which include school fixed effects, suggest that a 1 s.d. increase in principal TEAM rating is associated with a decrease in average teacher turnover in the school of half a percentage point, or roughly 5 percent of the average turnover rate. When replacing TEAM ratings with FTTT scores, the results are very similar.

Table 3 also shows that the descriptive pattern from Table 2—that less effective teachers turn over more often—holds when adjusting for other teacher and school characteristics. The predicted turnover rate of teachers who receive very low observation scores is roughly 23 percentage points greater than teachers with the highest observation scores, a very large difference. This difference drops to roughly 2 percentage points when comparing the lowest-growth and highest-growth teachers. Turning to other teacher characteristics, Table 3 shows that teachers with higher levels of education are more likely to leave their positions, and that black teachers are less likely to leave their schools than white teachers. Similar to previous studies of teacher turnover, we find a "U-shaped" relationship between teacher age/experience and the likelihood of turnover; teachers at each end of the age and experience distributions have higher turnover rates than teachers in the middle. The bottom of Table 3 also shows results for school characteristics; these results should be interpreted with caution given that several of the measures are highly correlated (e.g., racial/ethnic and socioeconomic composition).

[23] We omit the FTTT results with only year fixed and district fixed effects. The patterns are nearly identical to the TEAM models.

Principal Effectiveness and Strategic Retention

Table 3 demonstrates that more effective principals experience lower teacher turnover rates *on average*. Next, we allow for the possibility that schools with effective principals experience higher turnover rates among teachers at different levels of effectiveness. Table 4 shows the results for our school fixed effects model (control variables are omitted for brevity), with each column using a different measure of teacher effectiveness (observation, growth, and individual growth). The linear combination of the main effect and an interaction term yields a distinct marginal effect for teachers with different effectiveness levels.[24]

For each column in Table 4, the teachers with the highest effectiveness are the reference group, so the TEAM rating and FTTT score coefficients each represent a change in the predicted probability of teacher turnover among highly effective teachers. In every model, the estimated impact of principal effectiveness on the most effective teachers is negative; a 1 s.d. increase in principal effectiveness translates into approximately a 1 percentage point decrease in turnover for these teachers ($p < .01$ in each case).

The interaction terms in columns 1 and 4, however, suggest that the association between principal effectiveness and teacher turnover varies by teacher observation score. In contrast, for growth and individual growth, Table 4 shows little evidence of meaningful interaction. Table 5 makes these conclusions more readily apparent by showing the predicted change in turnover rates for a one standard deviation increase in principal effectiveness (TEAM rating or FTTT score) across the teacher effectiveness distribution, based on the results in Table 4. For the observation models, the evidence is consistent with strategic retention behavior—principals who receive higher ratings on the administrator evaluation rubric are *more* likely to retain teachers

[24] We also estimated models with district fixed effects instead of school fixed effects. Results were very similar.

with high observation scores but *less* likely to retain teachers with very low observation scores. To be more specific, a one standard deviation increase in the TEAM rating is associated with a 1.3 percentage point *decrease* in turnover among teachers scoring 4.50–5.00 but a 2.3 percentage point *increase* in the likelihood of turnover among teachers scoring 1.00–2.75. Compared to the average turnover rates for high-observation and low-observation teachers, these differences correspond to a 12% decrease and 6% increase, respectively.[25]

Figure 1 plots the predicted probability of teacher turnover by average observation score across the range of principal TEAM ratings. As noted previously, teachers with low observation scores have substantially higher turnover rates than teachers with average or above-average scores, irrespective of principal effectiveness, evidenced by the large gaps between the lines for teachers with high and low scores. However, low-scoring teachers in schools with effective principals are more likely to leave than low-scoring teachers in schools with ineffective principals, demonstrated by the upward sloping line for teachers in the lowest category. This finding is consistent with the idea that effective leaders seek to improve the quality of their school's teaching staff by pushing out low performers. Moreover, more effective principals are better at retaining high-scoring teachers. Moving from low to high principal TEAM ratings, we see increasingly large turnover gaps between effective and ineffective teachers.

[25] Both differences in effectiveness within the same principals over time (due, for example, to increases in effectiveness as principals gain experience) and between principals in the same school in different years appear to be relevant. For example, when we re-estimate these models with school-by-principal fixed effects, which isolate within-principal variation, observation results are very similar to those shown in Tables 4 and 5. When we re-estimate the observation models for the subset of schools who had a single principal across the analysis period, we in fact find stronger patterns of increased turnover among teachers with low observation scores. Lastly, when we attempt to isolate the variation from different principals in the same school by averaging principals' TEAM ratings across years (within the same schools) and re-estimating the observation models using the average rating instead of the year-specific rating, we find qualitatively similar results. Restricting this analysis to schools that had multiple principals attenuates the estimated marginal effects, particularly for teachers with very low observation scores, but this attenuation appears to be driven by principal turnover events. When we exclude principal turnover years, the marginal effects are again similar to the main results.

Returning to Table 5, the FTTT score results for teacher observations show similar patterns, though the evidence of increased teacher turnover among teachers with low observation scores is weaker; the marginal effect is positively signed but not statistically distinguishable from zero. An important consideration in comparing TEAM and FTTT is that they likely reflect different aspects of principal performance. While TEAM ratings come from supervisors who evaluate principals throughout the district, FTTT scores reflect teacher perceptions of principal performance within a given school. For the sake of simplicity and to further explore the strategic retention patterns we observe for this measure, our subsequent analyses focus on TEAM ratings as the measure of principal effectiveness.

Teacher Observation Scores Drive Strategic Retention Patterns

The teacher observation score results in Tables 4 and 5 stand in contrast to those for the two growth measures. Although Table 4 shows that more effective teachers by these metrics are more likely to stay in the school, Table 5 shows no evidence of an interaction with principal effectiveness and no strategic retention pattern. Instead, marginal effects of similar magnitude suggest that teachers' turnover probabilities decline roughly similarly as principal effectiveness increases across levels of teacher effectiveness.

Up to this point, however, we have considered teacher observation scores and growth scores separately. In reality, principals receive both signals, to some degree, for each teacher. Although growth scores are not available until the following fall, they may monitor proxies for this growth—for example, from student benchmark or interim assessment data—throughout the year. To parse out the relative importance of different signals of teacher performance to principal human capital decision-making, next we include observation and growth in the same model. We also include interactions between principal TEAM rating and both observation and growth

scores.[26] We focus on teachers with individual growth scores, since principals are unlikely to take the school-level scores assigned to non-tested teachers as signals of individual performance.

Table 6 shows the estimated marginal effects. The combination of two interaction terms with five levels produces a total of 25 coefficients. To isolate the relative importance of individual growth scores, we can compare the estimated marginal effects between teachers with different individual growth scores, holding observation scores constant. For example, the top row shows the estimated change in the probability of turnover for teachers with low observation scores with Level 1 through Level 5 individual growth scores. Each of the marginal effects is positive and similar in magnitude.[27] For teachers with high observation scores, the marginal effects are negative regardless of level of growth, with no clear evidence that the marginal effects increase for teachers with higher individual growth.[28]

We interpret these results as suggesting that observation scores, rather than measures of teacher effectiveness derived from student test score growth, drive the patterns of strategic retention. To see this interpretation more easily, Figure 2 plots the predicted probability of turnover for teachers with different observation and individual growth scores, based on Table 6. For simplicity, we compare teachers in the highest and lowest observation score categories who have individual growth scores of Level 1, Level 3, or Level 5. The probability of turnover among teachers with very low observation scores differs greatly from teachers with very high

[26] For completeness, we also include the teacher's achievement measure from the evaluation system in this model. Typically, however, this metric is chosen to be the same for most or all teachers in the school, so it is unlikely that principals use it in making decisions about teacher retention or other human capital processes. Thus, we do not include an interaction with achievement score in the models. Additionally, its inclusion increases the number of marginal effects from 25 to 125, with no substantive difference in the patterns of observation and growth scores.
[27] Most are just outside statistical significance at conventional levels, given the smaller sample sizes.
[28] We estimated models that substitute growth scores from the prior school year for current-year growth scores. The results show no evidence of differential retention by the prior-year growth scores. Additionally, expanding to all teachers (i.e., school and individual growth scores) instead of those with individual growth scores produces very similar results to those shown in Table 6.

observation scores, and the difference grows larger as principal effectiveness increases. Holding observation scores constant, however, turnover differences among teachers with low, medium, and high individual growth scores are minimal. There is, perhaps, some visual evidence of separation between high- and low-growth teachers in schools with highly-rated principals, but these differences are small and not statistically significant. In other words, as principal effectiveness increases, low observation scores predict greater turnover, and high observation scores predict lower turnover, nearly regardless of whether growth scores are high or low.

Strategic Retention or Strategic Scoring?

So far, we have discussed observation scores as an independent source of information principals use to identify high- and low-performers for the purpose of strategic retention, as if they are collected by an external source. However, these measures are not externally created; although some districts incorporate observations from external evaluators, principals (or other school leaders) usually assign them themselves. Given the weak evidence of an interaction with growth scores, a potential alternative explanation of our findings is that effective principals systematically assign lower observation scores to teachers whom they want to push out, regardless of the actual performance of those teachers. Under this interpretation, the observation score represents a removal mechanism—either by triggering a formal administrative process, such as tenure denial, or by providing a discouraging signal to a teacher that leads them to seek employment elsewhere—rather than a source of information regarding teacher effectiveness.

We conduct a number of analyses to investigate this possibility. First, we examine the extent to which more effective principals are more likely to assign low ratings. In Appendix Figure 1, we plot histograms of the average teacher observation score across each quintile of principal TEAM rating. Overall, the distribution of teacher observation scores is similar across

levels of principal effectiveness. In fact, the mean observation score *increases* in schools with more effective principals. This approach, however, examines both within- and between-school variation, whereas our identifying variation comes only from within-school variation in principal effectiveness. We more formally test the relationship between within-school changes in principal effectiveness and the distribution of teacher observation scores in Appendix Table 1. Here, we collapse the data into school-by-year cells and estimate the relationship between principal TEAM rating and (1) the school-by-year mean of observation scores, (2) the school-by-year standard deviation of observation scores, and (3) the proportion of teachers who fall into each of our five categories of average observation scores. We find that while within-school increases in principal TEAM rating are associated with small increases in the average teacher observation score, there is no change in the dispersion of observation scores, nor in the proportion of teachers who receive very low observation scores.

As an additional check, we investigate the extent to which lower scoring teachers are less effective according to other measures, and whether these patterns vary by principal TEAM rating. In Appendix Table 2, we show descriptive statistics for teachers across the range of observation scores and by quintile of their principal's TEAM rating. In addition to two teacher experience measures (first-year teacher and new-to-school teacher), we examine three types of observation score and value-added (TVAAS) measures: prior-year score, average of all prior scores, and career-average scores. Appendix Table 2 illustrates two important descriptive facts. First, across each measure, there is a monotonic relationship between experience/effectiveness and a teacher's average observation score in the current year. In other words, low-scoring teachers are the lowest performers across all measures, and high-scoring teachers are the highest performers. Second, teachers with low (high) observation scores in schools with effective

principals are observationally similar to teachers with low (high) observation scores in schools with ineffective principals.

Finally, we replicate our main results using ratings provided only by raters other than the principal himself or herself.[29] In most schools, assistant principals (APs) also conduct teacher observations, and we observe a somewhat smaller number of observations conducted by teacher-leaders or someone from central office. In Appendix Table 3, we show marginal effects from turnover models interacting principal TEAM ratings with observation scores, first for observations conducted by APs and then for other (non-school leader) raters. Results are consistent with those shown in Table 5; the highest-rated teachers are less likely to turn over as principal effectiveness increases, while the lowest-rated teachers' turnover probabilities increase.

Taken together, these analyses suggest that our main results are not driven by differential or strategic scoring of teachers by more effective principals. Instead, they are consistent with the interpretation that more effective principals are more successful at retaining teachers they (and others) observe to be more effective and at encouraging turnover among less effective teachers.

Strategic Retention across Different Teachers and School Contexts

Our third research question asks to what extent patterns of strategic retention change when looking at different types of teachers and across school contexts? Understanding differences across teachers and schools helps to provide insight into the mechanisms that underlie the descriptive patterns in our main results.

In Table 7, we examine differences in teacher retention across categories of teachers and schools. Each column shows the estimated marginal effects from a separate regression. The first

[29] This analysis relies on observation-level files, which are not available for the full sample of teachers. Specifically, we can conduct this analysis for teachers in districts that use the TEAM rubric for teacher observations, which covers roughly 85% of the analytic sample. Additionally, we do not include observations from 2011–12 because the data do not reliably identify who conducted observations in that year.

two columns compare results for schools whose principal stays following year *t* versus schools whose principal leaves following year *t*. We would expect that principal turnover matters for strategic retention behavior. Principals who do not plan to remain in the school next year may not prioritize prospective human capital management. Alternatively, new-to-school principals may be less equipped to distinguish between high- and low-performing teachers and/or have the means to strategically target retention efforts. Given that most principals do not receive value-added scores prior to the start of the next school year, new principals may also simply have fewer sources of information about the effectiveness of their inherited teachers.

The results in Table 7 (columns 1 and 2) are consistent with the idea that principal turnover disrupts strategic retention behavior. When limiting the sample to schools without a principal transition in a given year, patterns of strategic retention become stronger in comparison to the pooled models. In contrast, limiting to schools whose principals leave at the end of the year, the marginal effects across categories of observation scores are statistically similar. In other words, patterns of strategic retention are found solely in schools in which the principal returns next year, where the effects are even more pronounced than in the pooled sample.

Next, we examine whether there are differences between novice teachers (0–4 years of experience) and veteran teachers (5+ years of experience). In Tennessee, administrative means are more available for early-career teachers who have fewer due process protections, particularly in the wake of tenure reforms in 2011 (Wesson, 2012). Effective principals may be more adept at utilizing these administrative options to remove low-performing novice teachers from their schools. Novice teachers may also be more responsive to informal means, such as "counseling out" by the principal. Columns 3 and 4 show the results for novice and veteran teachers. We find

little evidence that the observed patterns differ between these two groups of teachers, which suggests that our results are not completely driven by administrative removal.

The ability or willingness of principals to strategically target teachers for retention or turnover may also vary across school contexts. Differences in school characteristics, for example, may proxy for differences in the labor market supply of teachers or in the accountability context for the school. For instance, principals who consistently face teacher shortages may not seek to "push out" ineffective teachers, but rather focus their efforts on retaining high-performers. To investigate these differences, we estimate separate models across three context measures: average school achievement, percentage of students qualifying for free/reduced price lunch, and locale type.

Columns 5 and 6 show the results for schools in the lowest and highest quintile of average school achievement, respectively. Here, we find important differences between high-achievement and low-achievement schools. The marginal effect of increasing principal effectiveness on teacher turnover for teachers with very low observation scores is positive and significant in high-achievement schools, but virtually zero in low-achievement schools. Furthermore, the marginal effect in high achievement schools is roughly three times greater in magnitude than the pooled model estimate in Table 5. Among teachers with the highest observation scores (4.50–5.00), neither of the marginal effects is statistically significant. For the middle of the observation score distribution, we find consistent evidence of greater retention in low achievement schools with more effective principals. In comparison to the pooled models, the estimated relationship between principal effectiveness and teacher retention is substantially larger for low-achievement schools, which highlights the importance of school leadership in

promoting stability in these schools. On the other hand, we find limited evidence that more effective principals in high achievement schools experience lower teacher turnover rates.

Columns 7 through 9 group schools by the percentage of students who qualify for free/reduced price lunch, a proxy for student poverty. While FRPL and school achievement are highly correlated (r = -0.80), school achievement may not fully capture differences between high-poverty and low-poverty schools. However, our results are similar to those in columns 5 and 6; effective principals in high-poverty schools are particularly adept at retaining effective teachers, while effective principals in low-poverty schools experience greater turnover among the lowest-scoring teachers.

School locale may also be an important determinant of teacher labor market dynamics. Urban and suburban schools, through their proximity to population centers, may have greater access to replacement teachers than rural schools. They may also have greater access to data management systems (e.g., data dashboards) and other supports principals to facilitate the use of teacher effectiveness data in making talent management decisions (Goldring et al., 2014; Grissom et al., 2017).

Columns 10 through 12 group teachers by school locale: urban, suburban, and town/rural.[30] The results are most striking for suburban schools—the marginal effects are positive and significant for the lowest-scoring teachers. A 1 s.d. increase in principal TEAM rating predicts an 8.8 percentage point increase in the likelihood of turnover among very low-scoring teachers in suburban schools. Effective principals in suburban schools also experience lower rates of turnover among very highly-rated teachers. Results for urban schools are similar in direction, though the marginal effect for the lowest scoring teachers is not statistically significant

[30] These designations come from the Common Core of Data (CCD) for each respective year. We combine town and rural schools for the sake of simplicity. Their individual patterns are very similar.

at conventional levels. Finally, effective principals in town/rural schools experience lower rates of turnover among high-scoring teachers, but no differences for low-scoring teachers.

Together, these results demonstrate that school context is an important aspect of the relationship between principal effectiveness and teacher turnover. In terms of pushing out low-performing teachers, the patterns are strongest in low-poverty, suburban schools, which could reflect labor market dynamics; principals in more advantaged schools may worry less about finding quality replacements for teachers who leave, leading them to focus on pushing out low performers. In contrast, principals in rural settings may instead focus on retaining teachers, since the cost or uncertainty associated with finding and hiring replacement teachers is higher.

Strategic Retention: Administrative Action or "Counseling Out"?

In Table 7 we showed that patterns of strategic retention are similar when comparing novice and veteran teachers, which suggests that formal processes may not be the primary driver of increased turnover among low-performers. To further examine this distinction, we restrict our model to teachers who received Level 3 or higher on their summative evaluation rating. Because we exclude teachers who are most vulnerable to administrative removal (Level 1 and Level 2), any remaining evidence of strategic retention would suggest that less formal means, such as counseling out, likely contribute to the patterns of strategic retention found in the main results.

We show these results in Table 8. When limiting the analysis to Level 3 through 5 teachers, effective principals' patterns of strategic retention are consistent with the main results. For example, a 1 s.d. increase in principal TEAM rating predicts a 2.9 percentage point increase in the likelihood of turnover among low-scoring teachers, which is 8% of the average turnover rate for such teachers. The results are nearly identical when restricting the analysis to medium- and high-growth teachers, including when we examine teachers in tested grades and subjects.

These patterns suggest that strong principals' strategic retention behavior—specifically, the failure to retain ineffective teachers—is not completely attributable to administrative means of removal. That is not to say that principals abstain from administrative means to remove teachers, given the substantially higher turnover rates among ineffective teachers. Rather, more effective principals may be better able to remove ineffective teachers through informal means, while also retaining effective teachers.

Strategic Retention by Types of Teacher Turnover

The prior analyses operationalize teacher turnover as a binary outcome. We next examine whether the relationship between principal effectiveness and teacher turnover changes when differentiating among types of turnover. This analysis is important for more fully understanding the consequences of strategic retention. If effective principals release low-performing teachers only to see them move to other district schools, for example, they may be contributing to a "dance of the lemons" rather than improving the overall quality of teachers in the school district.

Table 9 shows the results of the multinomial analysis. Each of the four turnover outcomes is estimated in reference to the base category (teachers who stay in their school). Panel A shows the regression results (again, with control variables omitted for brevity), while Panel B shows the estimated marginal effects of a 1 s.d. increase in principal effectiveness for teachers with different average observation scores.

Our earlier analysis showed that low-scoring teachers are more likely to turn over than average- or above average-scoring teachers. Table 9 shows that this pattern holds for all types of turnover, but that the largest increase is in the probability of exiting the education system. Additionally, the higher overall turnover rate of Level 1 teachers with effective principals is largely explained by exits and across-district moves (Panel B). A 1 s.d. increase in principal

TEAM rating raises the predicted probability of exiting the education system by 1.9 percentage points for low-scoring teachers and the probability of across-district transfer by a similar amount. In contrast, there is no evidence of an interaction in the within-district transfer column, and no evidence that effective principals are merely successful at shuffling teachers with lower observation scores to other schools in the district.

Among high-scoring teachers, effective principals have lower rates of within-district teacher transfers. A 1 s.d. increase in principal TEAM rating lowers the predicted probability of within-district transfer by 0.9 percentage points for teachers with the highest observation scores. This decrease is substantial—roughly 20% of the within-district transfer rate for these teachers.

DISCUSSION AND CONCLUSIONS

In recent years, many states have altered the policy environment around teacher evaluation and dismissal to explicitly encourage retention of effective teachers and "deselection" of ineffective teachers (Hanushek, 2009). Yet principals drive talent management (Goldring et al., 2014), and making the policy environment more conducive to strategic retention does not guarantee that it will be used to shape the teacher workforce.

Our results suggest that strategic retention is evidenced among more effective principals. As in prior work (e.g., Boyd et al., 2011), increases in principal effectiveness are associated with lower average teacher turnover rates, but, we find, focusing on turnover responses for the average teacher masks important differences between effective and ineffective teachers. High-performing teachers, measured both by classroom observation and value-added scores, are less likely to leave schools with effective principals. Yet teachers who receive low observation scores are in fact *more* likely to leave schools with effective principals, regardless of whether they have high or low value-added scores.

This evidence suggests that principals may rely more heavily on classroom observation scores than student test score growth to target high- and low-performers, which is consistent with prior qualitative investigations of principals' human capital decisions (Goldring et al., 2015). Principals conduct observations themselves, and can easily access scores assigned by themselves or other members of their leadership team throughout the school year. Moreover, observation information is available for every teacher. In contrast, although principals can monitor student progress through benchmark testing or other low-stakes assessments during the school year, teachers' growth scores are not directly observable until well into the next school year, and, in any case, are only relevant for teachers in tested classrooms. In this light, the primacy of observation scores for strategic retention is perhaps unsurprising.

A key implication of our analysis is that the school principal is an important—and, we would argue, largely ignored—component of the theory of action linking large-scale teacher evaluation systems to the reshaping of the teacher workforce through selective teacher retention, which often is used as justification for these systems (Hanushek, 2009). Inattention to variation in the capabilities of principals to collect actionable observation data and—more to the point of our analysis—act strategically in response to those data may be undercutting states' and districts' investments in teacher evaluation systems. On the other hand, our analyses also speak to conclusions that because few teachers receive low overall evaluation ratings that trigger dismissal or other administrative action, investment in teacher evaluation systems are resulting in little change (see Kraft & Gilmour, 2016). In fact, effective principals who deem a teacher low-performing via classroom observations systematically see those teachers leave their schools, and often the profession altogether, even when aggregation with other components of the evaluation system means that the overall level of effectiveness does not trigger administrative action.

An important addendum to these conclusions is that school context matters for strategic retention. For example, suburban schools appear better positioned than their urban and especially their rural counterparts to strategically retain teachers of different performance levels. With fewer applicants for open positions, rural schools may have fewer options for replacing low performers, meaning that it may in fact be strategic—albeit by a different definition than the one we have employed—for some rural principals to retain low performers rather than leave a position open or fill it with an undesirable alternative. Schools where the principal turns over also behave differently, with strategic retention patterns essentially driven only by schools that retain their principal from one year to the next, suggesting one potential reason that principal stability is associated with more positive school outcomes (Miller, 2013).

Attention to state context is also important. Our results come from Tennessee, a state which has invested heavily in a statewide evaluation system for teachers and principals. Such a system may be a necessary condition for strategic retention (TNTP, 2012). Meaningful teacher evaluation systems create opportunities to observe and evaluate representative teaching; the absence of such a system creates barriers to dismissal of low-performing teachers and to targeting resources, such as useful feedback or opportunities for teacher leadership, towards retention of high performers. Our results may not be generalizable to other state contexts, particularly those without such developed educator evaluation systems. An additional limitation concerns the simultaneity of principal evaluation and teacher retention decisions. Our results potentially could be driven by principal supervisors who observe strategic retention behaviors, thereby informing higher principal ratings. While Tennessee's administrator rubric does not explicitly evaluate strategic retention practices, it does include some indicators related to human capital and resource management. However, we believe it unlikely that any such bias is very

large, given that district evaluators do not differentiate among aspects of principal performance on the observation rubric, which covers many aspects of principal leadership beyond talent management (Grissom, Blissett, & Mitani, 2018).

Understanding the mechanisms of strategic retention is an important avenue for future work. While we present evidence that effective principals pursue such strategies, we cannot speak directly to their specific strategic behaviors because we cannot observe them. As we show, increased turnover among ineffective teachers under effective principals likely cannot be fully explained by administrative means (e.g., contract nonrenewal), but we need to know more about how principals use informal means to encourage (directly or indirectly) low-performing teachers to leave. At the other end of the distribution, future work might also investigate the strategies through which effective principals influence the retention of high-performing teachers. Such strategies may be particularly helpful to principals in low-achieving and high-poverty schools that have substantial challenges in retaining their most effective teachers.

More generally, our findings speak to the importance of achieving a better understanding of principals' human resources or talent management roles in their schools. Future work might explore the teacher hiring, assignment, and development strategies of effective principals. Prior research shows, for example, that high-growth schools hire teachers and place them across tested and untested classrooms differently than other schools (e.g., Grissom, Kalogrides, & Loeb, 2017; Loeb et al., 2012), but research has only begun to examine the specific ways principals—especially effective principals—shape and engage in their schools' talent management processes.

REFERENCES

Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2016). Teacher turnover, teacher quality, and student achievement in DCPS. *Educational Evaluation and Policy Analysis*, *39*(1), 1–23.

Allensworth, E., Ponisciak, S., & Mazzeo, C. (2009). *The schools teachers leave: Teacher mobility in Chicago Public Schools*. Chicago: Consortium on Chicago School Research.

Alliance for Excellent Education. (2014). *On the path to equity: Improving the effectiveness of beginning teachers*. Washington D.C. Retrieved from http://all4ed.org/wp-content/uploads/2014/07/PathToEquity.pdf

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Borman, G. D., & Dowling, N. M. (2008). Teacher attrition and retention: A meta-analytic and narrative review of the research. *Review of Educational Research*, *78*(3), 367–409.

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., Rockoff, J., & Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management*, 27(4), 793-818.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). Estimating the effect of leaders on public sector productivity: The case of school principals. *CALDER Working Paper*, (January), 1–50.

Burkhauser, S. (2017). How Much Do School Principals Matter When It Comes to Teacher Working Conditions? *Educational Evaluation and Policy Analysis*, 39(1), 126–145.

Cannata, M., Rubin, M., Goldring, E., Grissom, J. A., Neumerski, C. M., Drake, T. A., & Schuermann, P. (2017). Using teacher effectiveness data for information-rich hiring. *Educational Administration Quarterly*, *53*(2) 180–222.

Chingos, M. M., & West, M. R. (2011). Promotion and reassignment in public school districts: How do schools respond to differences in teacher effectiveness? *Economics of Education Review*, *30*(3), 419-433.

Cohen-Vogel, L. (2011). "Staffing to the test": Are today's school personnel practices evidence based? *Educational Evaluation and Policy Analysis*, *33*(4), 483–505.

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

Donaldson, M. L. (2013). Principals' approaches to cultivating teacher effectiveness: Constraints and opportunities in hiring, assigning, evaluating, and developing teachers. *Educational Administration Quarterly*, *49*(5), 838-882.

Drake, T. A., Goldring, E., Grissom, J. A., Cannata, M. A., Neumerski, C., Rubin, M., & Schuermann, P. (2016). Development or dismissal? Exploring principals' use of teacher effectiveness data. In *Improving teacher evaluation systems: Making the most of multiple measures*, eds. J. A. Grissom and P. Youngs. New York: Teachers College Press, 169-183.

Goldhaber, D., Gross, B., & Player, D. (2011). Teacher career paths, teacher quality, and persistence in the classroom: Are public schools keeping their best? *Journal of Policy Analysis and Management*, 30(1), 57-87.

Goldring, E., Neumerski, C. M., Cannata, M., Drake, T. A., Grissom, J. A., Rubin, M., & Schuermann, P. (2014). Principals' use of teacher effectiveness data for talent management decisions. Nashville, TN: Peabody College, Vanderbilt University. Retrieved from http://www.principaldatause.org/assets/files/reports/Summary-Report-201405.pdf

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, *44*(2), 96–104.

Gordon, R. J., Kane, T. J., & Staiger, D. (2006). *Identifying effective teachers using performance on the job*. Washington, DC: Brookings Institution.

Griffith, J. (2004). Relation of principal transformational leadership to school staff job satisfaction, staff turnover, and school performance. *Journal of Educational Administration*, *42*, 333–356.

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*, *1*, 1–44.

Grissom, J. A., Blissett, R. S. L., & Mitani, H. (2018). Evaluating school principals: Supervisor ratings of principal practice and principal job performance. *Educational Evaluation and Policy Analysis*, *40*(3), 446–472.

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., Kalogrides, D., & Loeb, S. (2017). Strategic staffing? How performance pressures affect the distribution of teachers within schools and resulting student achievement. *American Educational Research Journal*, *54*(6), 1079–1116.

Grissom, J. A., & Loeb, S. (2017). Assessing principals' assessments: Subjective evaluations of teacher effectiveness in low- and high-stakes environments. *Education Finance and Policy*, *12*(3), 369–395.

Grissom, J. A., Loeb, S., & Doss, C. (2015). The multiple dimensions of teacher quality: Does value-added capture teachers' nonachievement contributions to their schools? In *Improving teacher evaluation systems: Making the most of multiple measures*, eds. J. A. Grissom and P. Youngs. New York: Teachers College Press, 37-50.

Grissom, J. A., Loeb, S., & Nakashima, N. A. (2014). Strategic involuntary teacher transfers and teacher performance: Examining equity and efficiency. *Journal of Policy Analysis and Management*, *33*(1), 112–140.

Grissom, J. A., Rubin, M., Goldring, E., Cannata, M., Drake, T., Neumerski, C. M., & Schuermann, P. (2017). Central office supports for data-driven talent management decisions: Evidence from the implementation of new systems for measuring teacher effectiveness. *Educational Researcher*, 46(1), 21–32.

Grissom, J. A., Viano, S. L., & Selin, J. L. (2016). Understanding employee turnover in the public sector: Insights from research on teacher mobility. *Public Administration Review*, *76*(2), 241–251.

Grissom, J. A., & Youngs, P (Eds.). (2016). *Improving teacher evaluation systems: Making the most of multiple measures.* New York: Teachers College Press.

Guarino, C. M., Santibanez, L., & Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research*, *76*(2), 173–208.

Hallinger, P. (2005). Instructional leadership and the school principal: A passing fancy that refuses to fade away. *Leadership and Policy in Schools*, 4(3), 221-239.

Hanushek, E. A. *(*2009*).* Teacher deselection*. In* D. Goldhaber *&* J. Hannaway *(Eds.),* Creating a new teaching profession *(pp.* 165–180*).* Washington, DC*: Urban Institute Press.*

Hanushek, E. A., Rivkin, S. G., & Schiman, J. C. (2016). Dynamic effects of teacher turnover on the quality of instruction. *Economics of Education Review*, 55, 132-148.

Harris, D. N., & Sass, T. R. (2014). Skills, productivity and the evaluation of teacher performance. *Economics of Education Review*, *40*, 183–204.

Harrison, C., & Cohen-Vogel, L. (2012). The politics of teacher reform in Florida: Analyzing causal narratives surrounding state adoption of performance-based evaluations, Performance Pay, and Tenure Elimination. *Peabody Journal of Education*, *87*(5), 517–534.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, *38*(3), 499–534.

Jacob, B. (2011). Do principals fire the worst teachers? *Educational Evaluation and Policy Analysis*, *33*(January), 403–434.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, *26*(1), 101–136.

Johnson, S. M., & Birkeland, S. E. (2003). Pursuing a "sense of success": New teachers explain their career decisions. *American Educational Research Journal*, *40*(3), 581–617.

Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-needs schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record*, *114*(10), 1–39.

Kalogrides, D., Loeb, S., & Béteille, T. (2013). Systematic sorting: Teacher characteristics and class assignments. *Sociology of Education*, *86*(2), 103-123.

Kraft, M. A., & Gilmour, A. F. (2016). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. Brown University working paper.

Kraft, M. A., Marinell, W. H., & Yee, D. (2016). School organizational contexts, teacher turnover, and student achievement: Evidence from panel data. *American Educational Research Journal*, *53*(5), 1411–1449.

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.

Loeb, S., Beteille, T., & Kalogrides, D. (2012). Effective schools: Teacher hiring, assignment, development, and retention. *Education Finance and Policy*, *7*(3), 269–304.

Master, B. (2014). Staffing for success: Linking teacher evaluation and school personnel management in practice. *Educational Evaluation and Policy Analysis*, 36(2), 207-227.

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

Neumerski, C. M. (2013). Rethinking instructional leadership, a review: What do we know about principal, teacher, and coach instructional leadership, and where should we go from here? *Educational Administration Quarterly*, 49(2), 310-347.

Painter, S. R. (2000). Principals' perceptions of barriers to teacher dismissal. *Journal of Personnel Evaluation in Education*, *14*(3), 253–264.

Player, D. (2010). Nonmonetary compensation in the public teacher labor market. *Education Finance and Policy*, *5*(1), 82-103.

Reinhorn, S. K., Johnson, S. M., & Simon, N. (2017). Investing in development: six high-performing, high-poverty schools implement Massachusetts' teacher evaluation policy. *Educational Evaluation and Policy Analysis, 39*(3), 383-406.

Rockoff, B. J. E., Staiger, D. O., Kane, T. J., & Taylor, E. S. (2012). Information and employee evaluation: Evidence from a randomized intervention in public schools. *American Economic Review*, *102*(7), 3184–3213.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2012). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*(1), 4–36.

Sartain, L., Stoelinga, S. R., & Brown, E. R.. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation*. Chicago: Consortium on Chicago School Research.

Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives*, *24*(3), 97–118.

Steinberg, M. P., & Donaldson, M. L. (2016). The new educational accountability: Understanding the landscape of teacher evaluation in the post-NCLB era. *Education Finance and Policy, 11*(3), 340-359.

Stockard, J., & Lehman, M. B. (2004). Influences on the satisfaction and retention of 1st-year teachers: The importance of effective school management. *Educational Administration Quarterly*, *40*(5), 742–771.

The New Teacher Project. (2012). The irreplaceables: Understanding the real retention crisis in America's urban schools. Retrieved from http://tntp.org/assets/documents/TNTP_Irreplaceables_2012.pdf.

Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., & Morgan, K. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness.* New Teacher Project.

Wesson, L. (2012). Recent teacher policy changes in Tennessee: Achieving and maintaining tenure (OREA Report). Nashville, TN: Offices of Education Research and Accountability.

Figure 1: Predicted Probability of Teacher Turnover by Average Observation Score and Principal Effectiveness

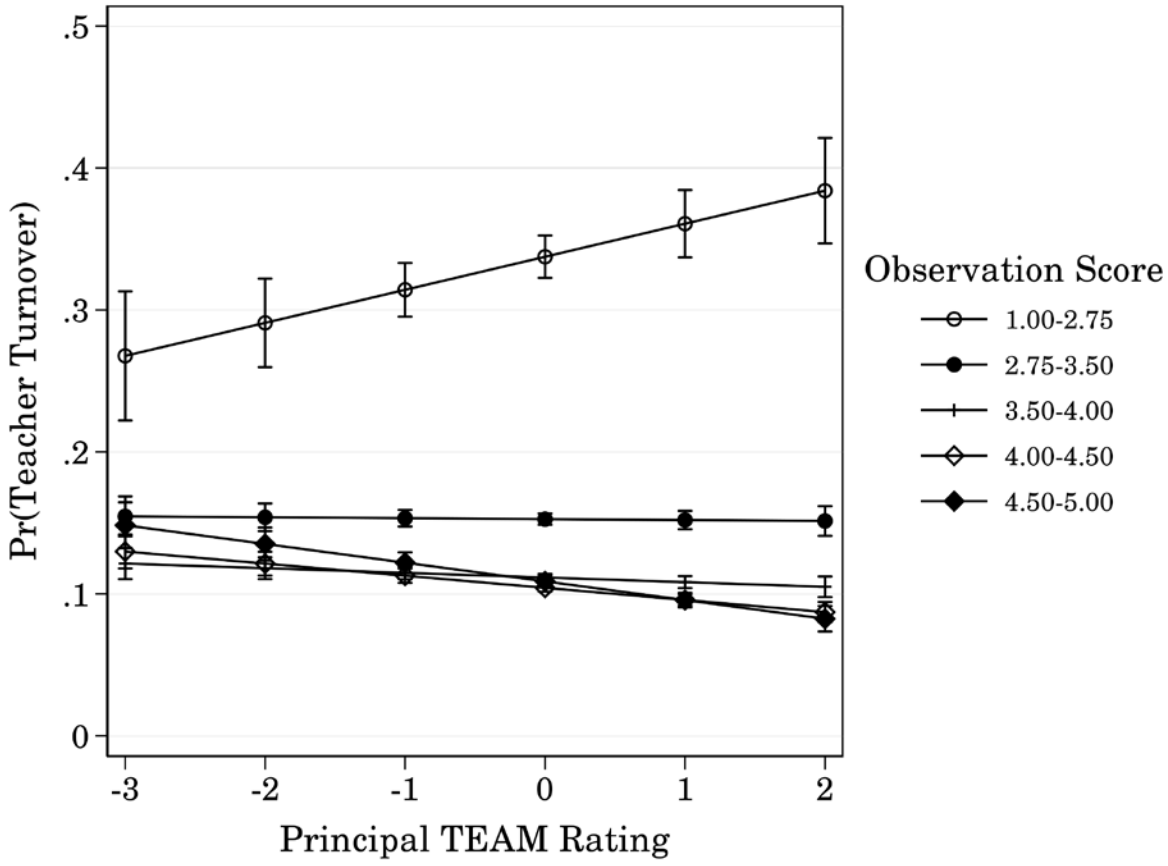


Figure 2: Probability of Teacher Turnover by Observation and Individual Growth Scores

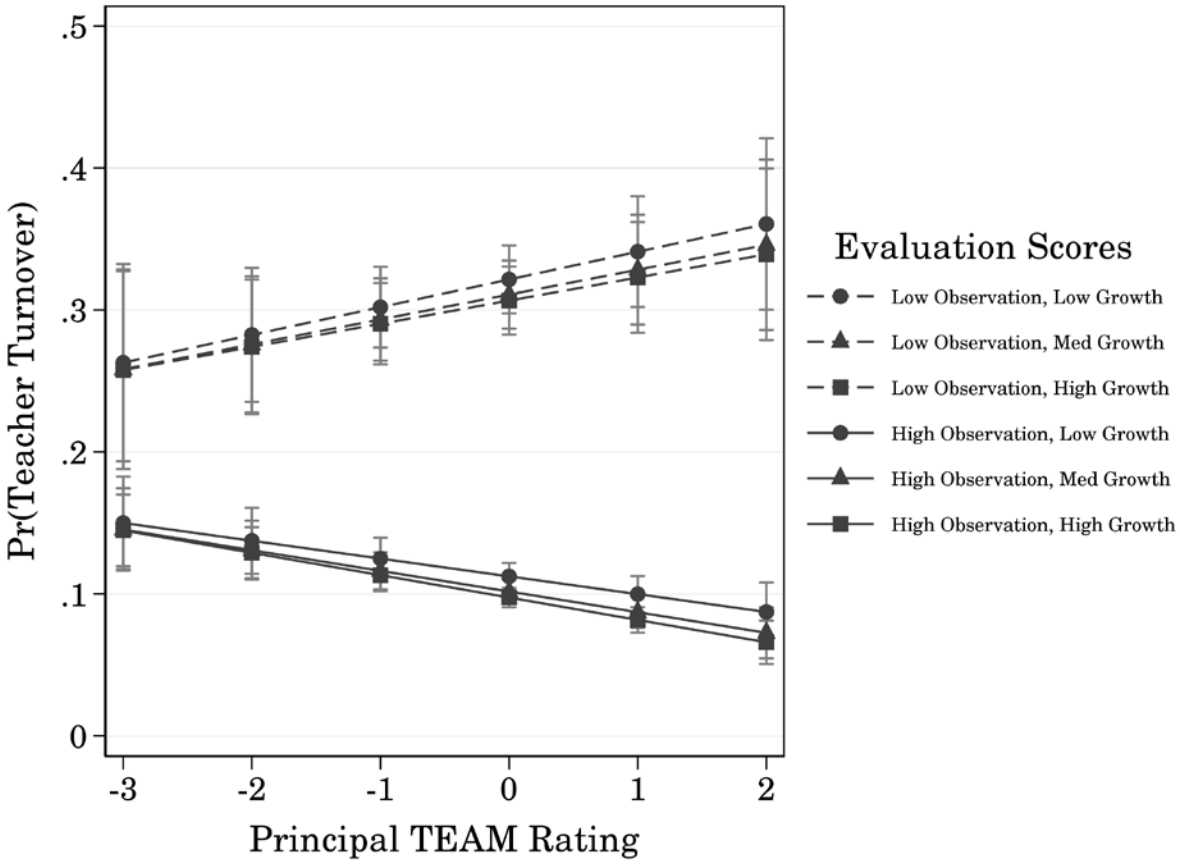


Table 1: Descriptive Statistics

	Mean	SD	Min	Max	N
Teacher Characteristics					
Female	0.79				292708
White	0.86				292708
Black	0.13				292708
Age	42.5	11.5	18	86	287261
Highest Level of Education					
BA degree	0.42				292699
MA degree	0.49				292699
EDS or PhD degree	0.08				292699
Experience					
0-4 years	0.28				291601
5-14 years	0.38				291601
15-24 years	0.21				291601
25-39 years	0.13				291601
40+ years	0.01				291601
School Characteristics					
Enrollment (100s)	8.25	4.65	0.01	55.20	286525
Proportion FRPL	0.59	0.24	0.00	1.00	286525
Proportion Black	0.24	0.29	0	1	286525
Proportion Hispanic/Latino	0.08	0.09	0	1	286525
Proportion gifted	0.02	0.03	0	1	286525
Proportion w/ disabilities	0.15	0.05	0	1	286525
School Level					
Elementary	0.53				285955
Middle	0.20				285955
High	0.24				285955
Other	0.04				285955
Teacher Effectiveness					
Average observation score	3.91	0.58	0	5	261804
Observation score 1.00-2.75	0.03				261804
Observation score 2.75-3.50	0.21				261804
Observation score 3.50-4.00	0.32				261804
Observation score 4.00-4.50	0.28				261804
Observation score 4.50-5.00	0.16				261804
Growth score	3.46	1.57	1	5	217033
Individual growth score	3.36	1.50	1	5	100070
Achievement score	4.12	1.24	0	5	215965
Principal Effectiveness					
TEAM rating	0.06	0.96	-3.96	2.25	226084
FTTT score	-0.02	0.96	-4.18	1.53	138030

Notes: All variables are shown at the teacher-by-year level covering the analysis period 2011–12 to 2015–16. Data from 2016–17 is used only to construct a turnover outcome for teachers in 2015–16. FTTT scores are available from 2011–12 to 2013–14. Teacher growth and achievement scores are unavailable in 2015–16 because Tennessee cancelled statewide testing in that year.

Table 2: Teacher Turnover Rates by Effectiveness Measures

Panel A: By Average Observation Score

	Pooled	1.00-2.75	2.75-3.50	3.50-4.00	4.00-4.50	4.50-5.00
Teacher turnover	0.133	0.372	0.173	0.122	0.109	0.110
Within-district move	0.055	0.110	0.069	0.052	0.047	0.045
Across-district move	0.023	0.078	0.037	0.022	0.016	0.012
Position change	0.010	0.004	0.004	0.007	0.011	0.020
Exit	0.045	0.179	0.062	0.040	0.034	0.033
N	261804	6820	55453	82470	73895	43166

Panel B: By Growth Score

	Pooled	Level 1	Level 2	Level 3	Level 4	Level 5
Teacher turnover	0.136	0.156	0.144	0.135	0.131	0.126
Within-district move	0.057	0.066	0.062	0.056	0.055	0.052
Across-district move	0.023	0.027	0.026	0.024	0.020	0.020
Position change	0.010	0.010	0.008	0.008	0.010	0.011
Exit	0.046	0.053	0.047	0.046	0.045	0.042
N	217033	43930	16977	42637	21544	91945

Panel C: By Individual Growth Score

	Pooled	Level 1	Level 2	Level 3	Level 4	Level 5
Teacher turnover	0.133	0.167	0.150	0.133	0.126	0.112
Within-district move	0.054	0.065	0.061	0.053	0.055	0.047
Across-district move	0.025	0.035	0.031	0.026	0.022	0.020
Position change	0.008	0.006	0.005	0.007	0.009	0.010
Exit	0.045	0.061	0.052	0.047	0.041	0.035
N	100070	18759	9344	24737	11573	35657

Notes: Panel A includes teachers from 2011–12 to 2015–16. Panels B and C include teachers from 2011–12 to 2014–15.

Table 3: Principal Effectiveness and Teacher Turnover (Baseline Model)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Principal Effectiveness								
TEAM Rating	-0.006***	-0.006***	-0.005***		-0.008***	-0.006***	-0.005***	
	(0.001)	(0.001)	(0.001)		(0.001)	(0.002)	(0.002)	
FTTT Score				-0.005***				-0.007***
				(0.002)				(0.002)
Teacher Effectiveness								
Observation score 1.00-2.75	0.234***	0.229***	0.226***	0.203***				
	(0.008)	(0.008)	(0.008)	(0.009)				
Observation score 2.75-3.50	0.046***	0.045***	0.045***	0.032***				
	(0.003)	(0.003)	(0.003)	(0.004)				
Observation score 3.50-4.00	0.003	0.004*	0.005*	-0.001				
	(0.002)	(0.002)	(0.002)	(0.003)				
Observation score 4.00-4.50	-0.002	-0.003	-0.003	-0.006*				
	(0.002)	(0.002)	(0.002)	(0.003)				
Level 1 growth					0.019***	0.018***	0.018***	0.023***
					(0.003)	(0.003)	(0.003)	(0.003)
Level 2 growth					0.010***	0.009**	0.010***	0.015***
					(0.004)	(0.004)	(0.004)	(0.004)
Level 3 growth					0.006**	0.006**	0.007***	0.005*
					(0.002)	(0.002)	(0.002)	(0.003)
Level 4 growth					0.003	0.003	0.004	0.004
					(0.003)	(0.003)	(0.003)	(0.003)
Teacher Characteristics								
Female	-0.008***	-0.008***	-0.003*	-0.004	-0.015***	-0.016***	-0.011***	-0.009***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Black	-0.011***	-0.009**	-0.016***	-0.012***	-0.009**	-0.004	-0.012***	-0.008**
	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Other race	0.022*	0.023*	0.017	0.013	0.017	0.018	0.012	0.015
	(0.012)	(0.012)	(0.012)	(0.014)	(0.014)	(0.013)	(0.013)	(0.014)
Highest Education								
MA degree	0.010***	0.011***	0.011***	0.009***	0.006***	0.007***	0.007***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
MA+ degree	0.021***	0.024***	0.025***	0.021***	0.018***	0.022***	0.023***	0.020***

	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
EDS degree	0.027***	0.033***	0.032***	0.028***	0.025***	0.031***	0.030***	0.025***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
PhD degree	0.056***	0.055***	0.057***	0.042***	0.045***	0.046***	0.048***	0.039***
	(0.009)	(0.009)	(0.009)	(0.011)	(0.010)	(0.010)	(0.010)	(0.011)
Teacher Age								
30-39	-0.007**	-0.007**	-0.005*	-0.003	-0.004	-0.003	-0.002	-0.002
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
40-49	-0.029***	-0.029***	-0.026***	-0.020***	-0.024***	-0.023***	-0.021***	-0.018***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
50-59	-0.037***	-0.037***	-0.035***	-0.035***	-0.030***	-0.029***	-0.028***	-0.029***
	(0.003)	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)
60 and above	0.046***	0.045***	0.047***	0.047***	0.061***	0.060***	0.063***	0.062***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)
Teacher Experience								
0-4 years	0.065***	0.063***	0.056***	0.057***	0.076***	0.074***	0.066***	0.065***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)
5-14 years	0.020***	0.020***	0.018***	0.018***	0.020***	0.019***	0.018***	0.019***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
25-39 years	0.029***	0.029***	0.029***	0.034***	0.026***	0.027***	0.027***	0.033***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
40+ years	0.066***	0.068***	0.071***	0.070***	0.058***	0.060***	0.064***	0.062***
	(0.012)	(0.012)	(0.012)	(0.016)	(0.014)	(0.014)	(0.014)	(0.016)
School Characteristics								
Enrollment (100s)	-0.001***	-0.001***	0.006***	0.014***	-0.002***	-0.001***	0.008***	0.013***
	(0.000)	(0.000)	(0.002)	(0.003)	(0.000)	(0.000)	(0.002)	(0.003)
Proportion Black	0.129***	0.135***	0.109*	-0.015	0.134***	0.145***	0.158**	-0.059
	(0.007)	(0.012)	(0.061)	(0.119)	(0.007)	(0.013)	(0.076)	(0.118)
Proportion Hispanic/Latino	0.104***	0.041**	0.190**	0.020	0.102***	0.043*	0.160*	0.004
	(0.015)	(0.020)	(0.074)	(0.104)	(0.017)	(0.023)	(0.093)	(0.103)
Proportion gifted	-0.004	0.101**	0.361***	0.585**	-0.002	0.102**	0.380**	0.534**
	(0.033)	(0.042)	(0.134)	(0.249)	(0.038)	(0.049)	(0.154)	(0.247)
Proportion w/ disabilities	0.009	0.000	0.013	0.061	0.028	0.018	-0.041	0.088
	(0.021)	(0.023)	(0.058)	(0.093)	(0.025)	(0.027)	(0.071)	(0.095)
Proportion FRPL	-0.045***	0.004	0.014	-0.004	-0.047***	0.009	0.001	0.013
	(0.006)	(0.009)	(0.012)	(0.023)	(0.007)	(0.011)	(0.018)	(0.024)

School Fixed Effects	No	No	Yes	Yes	No	No	Yes	Yes
District Fixed Effects	No	Yes	No	No	No	Yes	No	No
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	210140	210140	210140	126883	167892	167892	167892	127500
R^2	0.040	0.045	0.067	0.073	0.027	0.033	0.058	0.064

School-by-year clustered standard errors in parentheses. The dependent variable is whether a teacher leaves their position in the current year.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: The Interaction between Principal and Teacher Effectiveness

	(1)	(2)	(3)	(4)	(5)	(6)
Principal Effectiveness						
TEAM rating	-0.013***	-0.007***	-0.008***			
	(0.002)	(0.002)	(0.003)			
FTTT score				-0.010***	-0.007***	-0.011***
				(0.003)	(0.002)	(0.003)
Teacher Effectiveness						
Observation score 1.00-2.75	0.229***			0.204***		
	(0.008)			(0.009)		
Observation score 2.75-3.50	0.044***			0.033***		
	(0.003)			(0.004)		
Observation score 3.50-4.00	0.003			-0.001		
	(0.003)			(0.003)		
Observation score 4.00-4.50	-0.005*			-0.006*		
	(0.002)			(0.003)		
Level 1 growth		0.018***			0.023***	
		(0.003)			(0.003)	
Level 2 growth		0.009***			0.015***	
		(0.004)			(0.004)	
Level 3 growth		0.007***			0.005*	
		(0.002)			(0.003)	
Level 4 growth		0.003			0.004	
		(0.003)			(0.003)	
Level 1 individual growth			0.042***			0.043***
			(0.004)			(0.004)
Level 2 individual growth			0.026***			0.030***
			(0.005)			(0.005)
Level 3 individual growth			0.017***			0.015***
			(0.003)			(0.004)
Level 4 individual growth			0.011***			0.009**
			(0.004)			(0.004)
Interactions						
Obs 1.00-2.75 / Level 1 x TEAM	0.036***	0.005*	0.004	0.017*	0.002	0.002
	(0.008)	(0.003)	(0.004)	(0.009)	(0.003)	(0.005)
Obs 2.75-3.50 / Level 2 x TEAM	0.013***	0.001	0.008*	0.011***	-0.000	-0.001
	(0.003)	(0.004)	(0.005)	(0.004)	(0.005)	(0.006)
Obs 3.50-4.00 / Level 3 x TEAM	0.010***	0.003	0.002	0.004	-0.000	-0.001
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
Obs 4.00-4.50 / Level 4 x TEAM	0.005*	0.003	0.003	0.000	0.005	0.005
	(0.002)	(0.004)	(0.004)	(0.003)	(0.004)	(0.005)
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	210140	167892	76660	126883	127500	60024
R^2	0.067	0.058	0.081	0.074	0.064	0.089

School-by-year clustered standard errors in parentheses. The dependent variable is whether a teacher leaves their position in the current year. All models include teacher/school controls. "Obs" indicates observation score.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Marginal Effects of Principal Effectiveness on Teacher Turnover by Teacher Effectiveness

	TEAM Results			FTTT Results		
	Observation Score	Growth Score	Individual Growth Score	Observation Score	Growth Score	Individual Growth Score
Teacher Effectiveness						
Obs 1.00-2.75 / Level 1	0.023***	-0.002	-0.004	0.007	-0.005*	-0.009**
	(0.008)	(0.003)	(0.004)	(0.008)	(0.003)	(0.004)
Obs 2.75-3.50 / Level 2	-0.001	-0.005	0.000	0.000	-0.008*	-0.013**
	(0.002)	(0.004)	(0.005)	(0.003)	(0.004)	(0.006)
Obs 3.50-4.00 / Level 3	-0.003*	-0.004	-0.006*	-0.006***	-0.008***	-0.012***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)
Obs 4.00-4.50 / Level 4	-0.009***	-0.004	-0.005	-0.010***	-0.002	-0.007
	(0.002)	(0.004)	(0.004)	(0.002)	(0.003)	(0.004)
Obs 4.50-5.00 / Level 5	-0.013***	-0.007***	-0.008***	-0.010***	-0.007***	-0.011***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
N	210140	167892	76660	126883	127500	60024

Each column displays coefficients from a separate regression model interacting the principal's TEAM rating or FTTT score with teacher observation, growth, and individual growth scores. The marginal effects correspond to the results in Table 4. The dependent variable is whether a teacher leaves their position in the current year. "Obs" indicates observation score.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Identifying the Relative Importance of Different Measures of Teacher Effectiveness

	Individual Growth Score				
	Level 1	Level 2	Level 3	Level 4	Level 5
Observation Score					
1.00-2.75	0.020	0.023*	0.018	0.020	0.016
	(0.012)	(0.013)	(0.012)	(0.013)	(0.012)
2.75-3.50	-0.000	0.004	-0.002	0.000	-0.003
	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)
3.50-4.00	-0.000	0.003	-0.002	-0.000	-0.004
	(0.004)	(0.005)	(0.003)	(0.004)	(0.003)
4.00-4.50	-0.007*	-0.003	-0.009**	-0.007	-0.010***
	(0.004)	(0.005)	(0.004)	(0.005)	(0.003)
4.50-5.00	-0.013**	-0.009	-0.015***	-0.012**	-0.016***
	(0.005)	(0.006)	(0.004)	(0.005)	(0.004)
N	14451	7118	18759	8447	27300

School-by-year clustered standard errors in parentheses. Results shown are from estimating equation 2 with interactions between principal TEAM rating and both teacher observation scores and individual growth scores. Each cell shows the estimated marginal effect of a 1 s.d. change in principal TEAM rating on the probability of teacher turnover among teachers with the given observation (rows) and growth (column) scores. Models include school and year fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Exploring Heterogeneity in the Marginal Effects of Principal Effectiveness on Teacher Turnover

	Principal Turnover		**Teacher Experience**		**School Achievement**	
	Principal Stays	Principal Leaves	0-4 Years	5+ Years	Lowest Quintile	Highest Quintile
	(1)	(2)	(3)	(4)	(5)	(6)
Observation Score						
1.00-2.75	0.032***	-0.013	0.027**	0.019**	0.005	0.067***
	(0.009)	(0.018)	(0.012)	(0.010)	(0.022)	(0.022)
2.75-3.50	-0.000	-0.009	0.001	-0.002	-0.020**	-0.001
	(0.003)	(0.008)	(0.004)	(0.003)	(0.008)	(0.006)
3.50-4.00	-0.005**	-0.005	-0.006*	-0.002	-0.020***	-0.003
	(0.002)	(0.007)	(0.004)	(0.002)	(0.007)	(0.004)
4.00-4.50	-0.007***	-0.018**	-0.016***	-0.006***	-0.031***	-0.004
	(0.002)	(0.008)	(0.004)	(0.002)	(0.010)	(0.003)
4.50-5.00	-0.013***	-0.015*	-0.013**	-0.013***	0.000	-0.004
	(0.003)	(0.009)	(0.006)	(0.002)	(0.010)	(0.004)
N	175530	33164	54903	155237	20013	54335
	School FRPL Percentage			**School Locale**		
	80-100%	40-80%	0-40%	Urban	Suburban	Town/Rural
	(7)	(8)	(9)	(10)	(11)	(12)
Observation Score						
1.00-2.75	0.009	0.016*	0.041	0.021	0.088***	0.005
	(0.018)	(0.009)	(0.025)	(0.015)	(0.023)	(0.010)
2.75-3.50	-0.012*	-0.002	0.003	-0.006	0.008	-0.001
	(0.006)	(0.003)	(0.006)	(0.005)	(0.006)	(0.003)
3.50-4.00	-0.014**	-0.004*	0.002	-0.012***	0.005	-0.002
	(0.006)	(0.002)	(0.004)	(0.004)	(0.004)	(0.002)
4.00-4.50	-0.027***	-0.007***	-0.002	-0.015***	-0.002	-0.008***
	(0.007)	(0.002)	(0.004)	(0.005)	(0.004)	(0.002)
4.50-5.00	-0.023***	-0.011***	-0.008*	-0.020***	-0.011**	-0.011***
	(0.008)	(0.003)	(0.005)	(0.005)	(0.005)	(0.003)
N	28119	138283	43738	53614	43903	112023

School-by-year clustered standard errors in parentheses. Each column displays the marginal effects from a separate regression model interacting the principal's TEAM rating with a teacher's average observation score. Models include school and year fixed effects and the full set of teacher/school controls. The dependent variable is whether a teacher leaves their position in the current year.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Examining Patterns of Strategic Retention Among "Harder to Remove" Teachers

	LOE 3+	Growth 3+	Individual Growth 3+
Observation Score			
1.00-2.75	0.029***	0.028**	0.033*
	(0.011)	(0.011)	(0.018)
2.75-3.50	-0.003	-0.003	-0.004
	(0.003)	(0.003)	(0.005)
3.50-4.00	-0.002	-0.003	-0.003
	(0.002)	(0.002)	(0.003)
4.00-4.50	-0.009***	-0.009***	-0.009***
	(0.002)	(0.002)	(0.003)
4.50-5.00	-0.013***	-0.011***	-0.013***
	(0.003)	(0.003)	(0.004)
N	150649	120160	54623

School-by-year clustered standard errors in parentheses. Each column displays coefficients from a separate regression model interacting the principal's TEAM rating with a teacher's average observation score. Models include school and year fixed effects and the full set of teacher/school controls. The dependent variable is whether a teacher leaves their position in the current year.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

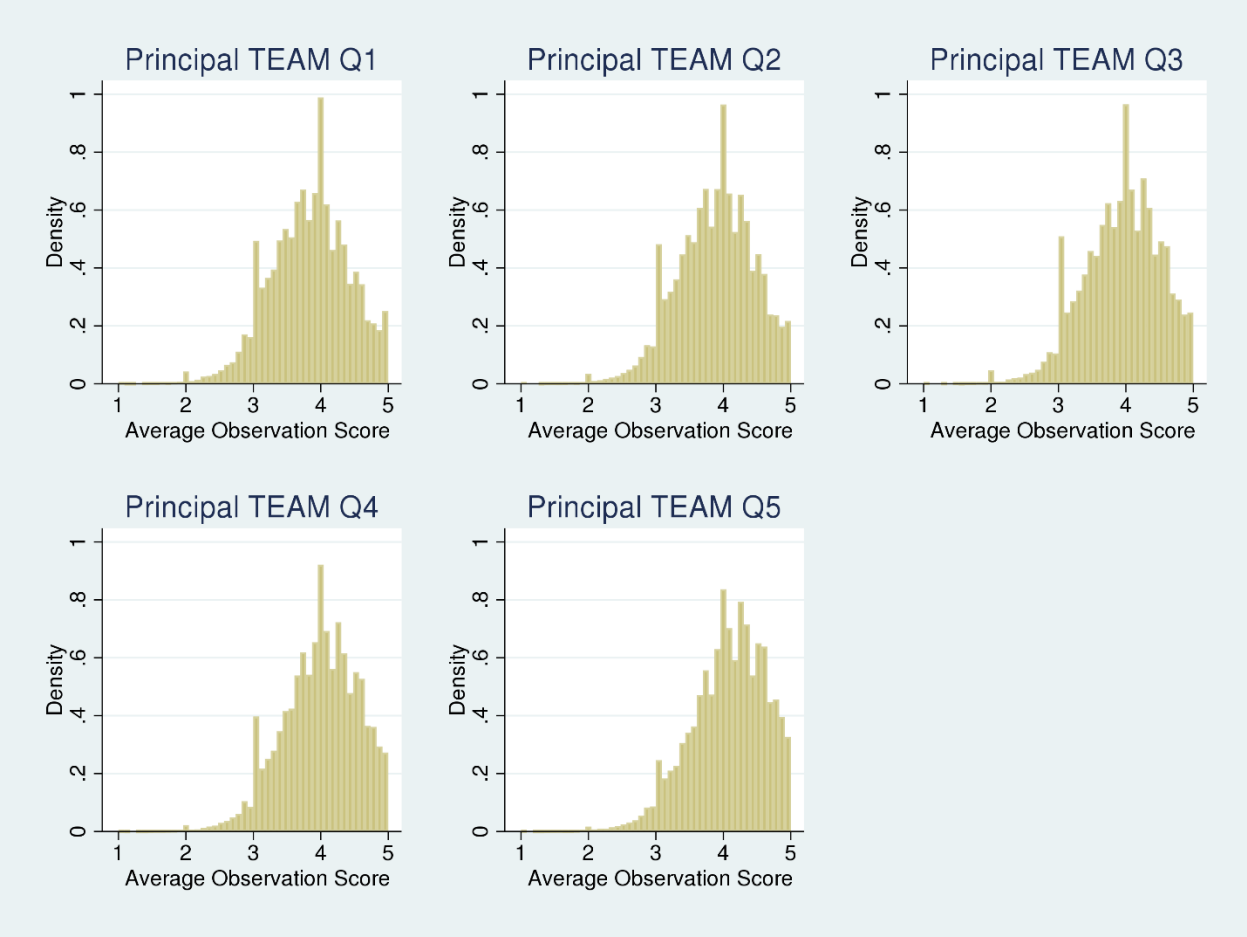
Table 9: Multinomial Teacher Turnover Results

Panel A: OLS Results	Exit	Within District	Across District	Position Change
Principal Effectiveness				
TEAM rating	-0.004***	-0.009***	-0.001	-0.002**
	(0.001)	(0.002)	(0.001)	(0.001)
Average Observation Score				
1.00-2.75	0.178***	0.066***	0.104***	-0.022***
	(0.007)	(0.006)	(0.006)	(0.002)
2.75-3.50	0.033***	0.015***	0.026***	-0.021***
	(0.002)	(0.002)	(0.002)	(0.001)
3.50-4.00	0.009***	0.003	0.009***	-0.018***
	(0.002)	(0.002)	(0.001)	(0.001)
4.00-4.50	0.003**	0.001	0.003***	-0.013***
	(0.001)	(0.002)	(0.001)	(0.001)
Interactions				
Observation score 1.00-2.75 x TEAM	0.023***	0.005	0.020***	0.005***
	(0.007)	(0.007)	(0.006)	(0.002)
Observation score 2.75-3.50 x TEAM	0.004*	0.006***	0.002	0.003***
	(0.002)	(0.002)	(0.002)	(0.001)
Observation score 3.50-4.00 x TEAM	0.003*	0.004**	0.001	0.004***
	(0.002)	(0.002)	(0.001)	(0.001)
Observation score 4.00-4.50 x TEAM	0.001	0.002	-0.000	0.002*
	(0.002)	(0.002)	(0.001)	(0.001)
Observations	194106	194025	189299	186612
R^2	0.064	0.069	0.042	0.024

Panel B: Marginal Effects	Exit	Within District	Across District	Position Change
Observation Score				
1.00-2.75	0.019***	-0.004	0.019***	0.003*
	(0.007)	(0.006)	(0.006)	(0.002)
2.75-3.50	-0.000	-0.003*	0.001	0.001**
	(0.002)	(0.002)	(0.001)	(0.001)
3.50-4.00	-0.001	-0.005***	0.000	0.001***
	(0.001)	(0.001)	(0.001)	(0.001)
4.00-4.50	-0.003***	-0.006***	-0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
4.50-5.00	-0.004***	-0.009***	-0.001	-0.002**
	(0.001)	(0.002)	(0.001)	(0.001)

School-by-year clustered standard errors in parentheses. Models include school and year fixed effects and the full set of teacher/school controls. The dependent variable is the type of teacher turnover listed at the top of the column.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Figure 1: Distribution of Teacher Observation Scores by Principal Team Rating (Quintiles)



Means and standard deviations by quintile: 3.82, 0.58 (Q1); 3.87, 0.57; 3.93, 0.57; 3.98, 0.56; 4.06, 0.55 (Q5)

Appendix Table 1: Predicting School-by-Year Aggregated Observation Scores

	Mean	SD	Prop 1.00-2.75	Prop 2.75-3.50	Prop 3.50-4.00	Prop 4.00-4.50	Prop 4.50-5.00
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Principal Effectiveness							
TEAM rating	0.035***	-0.003	-0.001	-0.019***	-0.011***	0.014***	0.017***
	(0.005)	(0.003)	(0.001)	(0.003)	(0.003)	(0.004)	(0.003)
School Characteristics							
Enrollment (100s)	0.008	-0.002	-0.001	-0.001	-0.006	0.002	0.006
	(0.007)	(0.003)	(0.001)	(0.004)	(0.004)	(0.005)	(0.004)
Proportion Black	-0.353	-0.050	-0.014	0.112	0.270*	-0.124	-0.244*
	(0.215)	(0.119)	(0.047)	(0.130)	(0.147)	(0.168)	(0.129)
Proportion	-0.780***	0.019	0.051	0.078	0.511***	-0.084	-0.556***
Hispanic/Latino	(0.241)	(0.140)	(0.053)	(0.149)	(0.158)	(0.155)	(0.153)
Proportion gifted	-0.058	0.386	0.005	0.106	-0.078	-0.087	0.054
	(0.539)	(0.267)	(0.099)	(0.347)	(0.346)	(0.365)	(0.337)
Proportion w/ disabilities	-0.329*	0.101	0.014	0.094	0.169	-0.111	-0.166
	(0.199)	(0.099)	(0.042)	(0.128)	(0.139)	(0.147)	(0.120)
Proportion FRPL	0.055	0.025	0.021**	-0.050	-0.024	0.014	0.039
	(0.048)	(0.021)	(0.010)	(0.035)	(0.035)	(0.030)	(0.024)
Teacher Characteristics							
Female	-0.032	0.014	0.008	0.064	-0.105	0.113	-0.079
	(0.111)	(0.061)	(0.027)	(0.095)	(0.099)	(0.094)	(0.049)
Black	-0.139	0.048	0.023	0.023	0.039	-0.037	-0.048
	(0.112)	(0.051)	(0.024)	(0.056)	(0.067)	(0.053)	(0.069)
Other race	0.634**	-0.180	-0.112	-0.098	-0.607**	0.783***	0.034
	(0.286)	(0.243)	(0.082)	(0.180)	(0.272)	(0.259)	(0.190)
Highest Education							
MA degree	0.110	-0.083**	-0.034**	-0.027	0.075	-0.054	0.040
	(0.072)	(0.036)	(0.016)	(0.058)	(0.065)	(0.060)	(0.034)
MA+ degree	-0.009	0.007	0.011	0.031	-0.101	0.018	0.041
	(0.151)	(0.086)	(0.029)	(0.110)	(0.118)	(0.122)	(0.100)
EDS degree	0.150	-0.128*	-0.044	0.015	-0.051	0.015	0.065
	(0.141)	(0.067)	(0.027)	(0.093)	(0.126)	(0.118)	(0.064)
PhD degree	0.082	0.155	0.002	0.095	-0.124	-0.307	0.334
	(0.366)	(0.183)	(0.064)	(0.217)	(0.261)	(0.238)	(0.239)
Teacher Age							
30-39	0.041	-0.041	0.026	-0.113	0.034	0.058	-0.005
	(0.097)	(0.043)	(0.022)	(0.072)	(0.072)	(0.076)	(0.052)
40-49	0.031	0.013	0.033	-0.108	0.059	0.026	-0.010
	(0.113)	(0.054)	(0.027)	(0.086)	(0.086)	(0.085)	(0.054)
50-59	-0.056	0.041	0.044	-0.076	0.020	0.077	-0.065
	(0.132)	(0.065)	(0.029)	(0.100)	(0.101)	(0.110)	(0.064)
60 and above	-0.125	0.106	0.057*	0.045	-0.047	-0.031	-0.024
	(0.156)	(0.079)	(0.034)	(0.121)	(0.120)	(0.122)	(0.082)
Teacher Experience							
0-4 years	-0.295***	0.030	0.026	0.168**	0.057	-0.144*	-0.107**
	(0.102)	(0.055)	(0.023)	(0.072)	(0.082)	(0.085)	(0.055)
5-14 years	-0.033	0.004	-0.010	0.053	0.010	-0.061	0.009
	(0.086)	(0.046)	(0.021)	(0.061)	(0.067)	(0.074)	(0.051)

25-39 years	0.126	-0.048	-0.024	0.045	-0.103	-0.116	0.198***
	(0.121)	(0.060)	(0.029)	(0.082)	(0.087)	(0.091)	(0.065)
40+ years	-0.263	-0.316*	0.011	0.067	0.008	-0.005	-0.081
	(0.346)	(0.191)	(0.056)	(0.288)	(0.212)	(0.223)	(0.213)
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6639	6590	6639	6639	6639	6639	6639
R^2	0.736	0.636	0.545	0.641	0.539	0.524	0.711

Observations are school-by-year averages. School-clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table 2: Descriptive Statistics by Teacher Observation Score and Quintile of Principal TEAM Rating

	Teacher Average Observation Score				
	1.00-2.75	**2.75-3.50**	**3.50-4.00**	**4.00-4.50**	**4.50-5.00**
Principal TEAM rating Q1					
First-Year Teacher	0.18	0.11	0.05	0.02	0.01
New-to-School Teacher	0.32	0.22	0.13	0.09	0.07
Prior-year Observation (1–5 scale)	3.04	3.39	3.74	4.06	4.43
All Prior Observations (SD)	-1.30	-0.78	-0.26	0.22	0.78
Career Average Observations (SD)	-1.86	-0.90	-0.19	0.33	0.93
Prior-year TVAAS Index (SD)	-0.83	-0.48	-0.15	0.17	0.49
All Prior TVAAS (SD)	-0.65	-0.46	-0.18	0.09	0.39
Career Average TVAAS (SD)	-0.63	-0.41	-0.16	0.08	0.34
Principal TEAM rating Q2–Q4					
First-Year Teacher	0.16	0.12	0.05	0.02	0.00
New-to-School Teacher	0.30	0.21	0.13	0.09	0.06
Prior-year Observation (1–5 scale)	3.05	3.42	3.75	4.05	4.42
All Prior Observations (SD)	-1.35	-0.76	-0.26	0.16	0.75
Career Average Observations (SD)	-1.86	-0.88	-0.20	0.32	0.90
Prior-year TVAAS Index (SD)	-0.74	-0.45	-0.13	0.13	0.49
All Prior TVAAS (SD)	-0.64	-0.41	-0.16	0.07	0.37
Career Average TVAAS (SD)	-0.60	-0.37	-0.14	0.06	0.34
Principal TEAM rating Q5					
First-Year Teacher	0.20	0.13	0.06	0.02	0.01
New-to-School Teacher	0.35	0.22	0.13	0.08	0.05
Prior-year Observation (1–5 scale)	3.08	3.40	3.74	4.06	4.42
All Prior Observations (SD)	-1.30	-0.76	-0.26	0.19	0.76
Career Average Observations (SD)	-1.87	-0.90	-0.19	0.34	0.90
Prior-year TVAAS Index (SD)	-0.80	-0.49	-0.09	0.20	0.54
All Prior TVAAS (SD)	-0.67	-0.45	-0.15	0.10	0.39
Career Average TVAAS (SD)	-0.62	-0.42	-0.15	0.09	0.36

Appendix Table 3: Principal Effectiveness and Teacher Turnover by Teacher Observation Score, Ratings Provided by Raters Other Than Principals (Marginal Effects)

	Assistant Principal Ratings	Non-School Leader Ratings
Average Observation Score		
1.00-2.75	0.024***	0.022
	(0.009)	(0.013)
2.75-3.50	0.000	0.016***
	(0.004)	(0.006)
3.50-4.00	-0.005	-0.005
	(0.003)	(0.006)
4.00-4.50	-0.006**	-0.015***
	(0.003)	(0.006)
4.50-5.00	-0.008**	-0.021***
	(0.004)	(0.007)
N	87209	26336

Each column displays coefficients from a separate regression model interacting the principal's TEAM rating with teacher observation scores. Models include school and year fixed effects and the full set of teacher/school controls. The dependent variable is whether a teacher leaves their position in the current year.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$