Essays in Principal Improvement, Quality, and Turnover

By

Brendan Bartanen

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Leadership and Policy Studies

June 30, 2019

Nashville, Tennessee

Approved:

Jason A. Grissom, Ph.D.

Robert D. Ballou, Ph.D.

Christopher A. Candelaria, Ph.D.

Christopher S. Carpenter, Ph.D.

To my parents, who provide endless love, wisdom, and encouragement.

To my students from C.O. Greenfield School, who are a reminder of why I do this work.

ACKNOWLEDGMENTS

There are many people—family, friends, colleagues, and teachers—who were instrumental in helping me to reach this milestone. While I cannot thank each of them in this limited space, I want to acknowledge that my successes are not possible without their kindness, patience, and sacrifice. There are a few people, however, whom I wish to highlight here.

I want to thank collectively my committee members for their questions, suggestions, and criticisms throughout the process. I am also grateful to the Tennessee Education Research Alliance and Tennessee Department of Education, who provided data access for each project.

It is difficult to express the extent of my appreciation for the mentorship from my advisor and dissertation chair, Jason Grissom. From the moment we started working together at the end of my first year, Jason treated me as a colleague and collaborator. Our projects over the past three years not only provided me with real-world applications for what I was learning in coursework, but also taught me important lessons about publishing academic research and effectively communicating research to policymakers and practitioners. Above all, he cared about my development and was my steadfast advocate. I cannot thank him enough.

I took five courses taught by Dale Ballou and I count him among the best teachers I have ever seen. Nobody pushed me more than he did, and I thank him for always engaging with my work and ideas, even when I was wrong. Chris Candelaria helped keep me sane during the dissertation and job market process, and I appreciate that his door was always open. Aliza Husain provided encouragement that kept me going on the toughest days.

None of this would have been possible without my family. My parents made sure I received a world-class education, but also taught me to work hard, be humble, and serve others. My brother and sister-in-law, Peter and Julie, are two of the most selfless people I know and inspired me to be a teacher and work to improve public education. My wife, Paulina, sacrificed on my behalf and provided unconditional love and support through both the peaks and the valleys. Thank you all for everything you have done for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Education scholars have long recognized the importance of high-quality leadership for the success of schools. However, our understanding of the labor market for principals remains quite shallow. In particular, there is a lack of rigorous quantitative evidence linking principals to student, teacher, and school outcomes. This dissertation helps to fill this gap by examining three aspects of principal quality: principals' on-the-job improvement, principal effects on student attendance, and the impact of principal turnover.

Recognition of the importance of school leaders has spurred attention towards increasing principal quality. Yet we have relatively little understanding of why some principals are more effective than others, which makes it unclear which policy approaches are likely to be effective in increasing the average quality of leadership. In particular, almost no work examines the extent to which principals improve over time. In the first chapter, I use statewide data from Tennessee over more than a decade to estimate the job performance returns to principal experience as measured by student, teacher, and principal outcomes. I find that principals improve substantially over time, evidenced by higher student achievement, higher ratings from supervisors, and lower rates of teacher transfer. However, improvement in student achievement as principals gain experience does not carry over when principals change schools, suggesting that the returns to experience are driven by school-specific rather than general skills. Finally, the returns to school-specific experience are largest for principals in high-poverty schools, highlighting the potential benefits of policies to improve the recruitment and retention of high-quality leaders.

An emerging body of research examines the extent to which schools affect students' non-test-score outcomes. However, prior work almost exclusively focuses on teachers, and no studies have explicitly examined principal effects on non-test-score outcomes. The second chapter begins to fill this gap by estimating principal value-added to student attendance. Examining attendance is particularly timely given increased attention to chronic absenteeism as part of the recent reforms in the Every Student Succeeds Act (ESSA). Drawing on statewide data from Tennessee over more than a decade, I find that principal effects on student attendance are comparable or larger in magnitude than effects on student achievement. Moving from the 25th to 75th percentile in principal value-added raises student attendance by 0.14 standard deviations, which corresponds to 2.2 additional instructional days. Comparing across school contexts, principals have larger effects on attendance in urban, high-poverty, and high schools. Additionally, attendance and achievement value-added are only weakly correlated, demonstrating that principals who excel at improving student non-test-score outcomes may not be those who excel at increasing student test scores. High-stakes accountability measures, such as supervisor ratings, fail to identify principals who improve student attendance.

Nationally, 18% of principals turn over each year, yet research has not yet credibly established the effects of this turnover on student and teacher outcomes. In the third chapter, I draw on statewide data from Missouri and Tennessee to employ a difference-in-differences model with a matched comparison group to estimate arguably causal effects. I find that principal turnover lowers school achievement by 0.03 SD in the next year, on average. Effects vary by transition type, with larger negative effects for transfers to other schools but no or even positive later effects of demotions of (presumably lower-performing) principals. Principal turnover also increases teacher turnover, but this does not explain the drop in student achievement. Replacement with an experienced successor can largely offset negative principal turnover effects.

Collectively, these chapters speak to the importance of effective school leadership and for the need to craft policies that identify high-quality leaders and give them the necessary supports to improve and remain in their schools. Further, these studies highlight the multidimensional nature of principal quality and extend beyond student test scores as the sole measure of school performance. Finally, each chapter pushes the envelope in terms of applying rigorous quantitative methods to large-scale administrative datasets.

CHAPTER I

Identifying Principal Improvement

Effective leadership is an important ingredient in school performance. High-quality principals are linked to a variety of school and student outcomes, including higher student achievement (Grissom et al., 2015a; Chiang et al., 2016; Dhuey and Smith, 2018; Coelli and Green, 2012), lower teacher turnover (Boyd et al., 2011; Grissom, 2011; Grissom and Bartanen, 2019b; Ladd, 2011), and better school climate (Sebastian and Allensworth, 2012; Burkhauser, 2017). At the same time, there exists substantial variation in principal quality, and disadvantaged schools tend to be led by principals with less experience (Grissom et al., 2019; Loeb et al., 2010; Branch et al., 2012) and lower effectiveness ratings (Grissom et al., 2019).

While it is clear that principals matter and that some principals are more effective than others, we have less knowledge of what drives the variation in principal quality. In particular, little research has considered on-the-job improvement. Despite the conventional wisdom that more experienced principals tend to be more effective, we have little evidence that explicitly identifies the job performance returns to principal experience. This lack of evidence stands in contrast to the robust literature on teacher improvement and worker productivity, more broadly. Numerous studies, for example, document substantial returns to experience for teachers, particularly in the first few years of teaching (e.g., Ladd and Sorensen, 2017; Harris and Sass, 2011; Kane et al., 2008; Papay and Kraft, 2015). While most of these studies focus on student achievement, some have also shown positive effects of teacher experience on non-test outcomes, such as student attendance (Ladd and Sorensen, 2017).

Understanding the returns to principal experience is an important issue for both policymakers and researchers, particularly given increased investments in school leadership at the district, state, and federal level. Prior studies have found a correlation between principal experience and performance (e.g., Clark et al., 2009; Grissom et al., 2018; Bastian and

Henry, 2015), but it is not clear whether this is driven by within-principal improvement or the selective attrition of less effective principals. Distinguishing between improvement and selection matters, as they lead to different policy prescriptions. If the returns to principal experience are small, for instance, resources may be better spent on identifying high-quality candidates for school leadership and removing ineffective leaders, rather than focusing on the development of the existing stock of leaders. However, if the returns to principal experience are large relative to the overall distribution of principal quality, policies focused on retaining leaders and providing them with opportunities for development would be more effective at raising the average quality of school leadership.

Using longitudinal administrative data from Tennessee, this study fills an important gap in the literature by estimating the job performance returns to principal experience. Specifically, my primary research question examines the extent to which principals improve at raising student achievement as they gain experience. I supplement this analysis with other outcomes, such as rubric-based ratings of principals' practice from their supervisors, student attendance and discipline, and teacher turnover. Examining these outcomes is important because they potentially capture different dimensions of principal performance, but they may also help to explain the mechanisms that drive the relationship between principal experience and student achievement.

My second research question examines whether the returns to principal experience are portable–i.e., whether improvement in a principal's ability to raise student test scores carries over when the principal changes schools. Specifically, I estimate the extent to which the returns to principal experience are driven by *total* experience (i.e., number of years as principal) versus *school-specific* experience (i.e., number of years as principal in the same school). Understanding the difference between the returns to total versus school-specific principal experience is important for research and policy. Conceptually, this distinction can provide insight into the nature of principal improvement. What are the skills that principals build that lead to greater student performance? For example, principals may become more effective over time at using data to drive school-level policies (e.g., identifying students who need additional support), which could lead to higher test scores, even when the principal changes schools. Alternatively, principals may become more effective over time through fostering relationships with teachers, families, or the broader community, which ultimately improve student learning. Upon moving to a new school, these relationships have to be rebuilt, such that the accumulated "improvement" in the prior school does not help to increase student achievement in the new school.

From a policy perspective, understanding whether principal improvement reflects returns to total versus school-specific experience informs debates regarding the allocation of principals and the importance of promoting stability in school leadership. If improvement is largely school-specific, for instance, a policy that moves experienced principals into struggling schools may have unanticipated costs to both sending and receiving schools.

My final research question examines heterogeneity in the returns to principal experience by school context. Prior work establishes that principals in schools serving larger percentages of low-income and low-achieving students have less experience and lower ratings from supervisors, on average (Grissom et al., 2019). One potential driver of this pattern is that principals in these schools improve at lower rates, which could lead them to receive lower ratings and turn over more frequently. Further, understanding the average improvement trajectory of principals across different types of schools helps to quantify the extent to which differences in principal turnover rates may contribute to disparities in access to high-quality principals.

Estimating the causal effect of principal experience presents an empirical challenge. To overcome this, I estimate models that include both principal and school fixed effects. Because principal quality varies and likely influences whether principals remain in the principalship (Grissom and Bartanen, 2019a), it is critical to isolate within-principal variation in experience to ameliorate selection bias. Similarly, principals are not randomly assigned to schools and prior work shows they tend to sort to more advantaged schools over time (Loeb et al., 2010;

Béteille et al., 2012). The inclusion of school fixed effects helps ensure that the estimated returns to experience are not conflated with differences in school quality across a principal's career.

One existing paper has used longitudinal administrative data to examine the relationship between principal experience and student test scores. Clark et al. (2009) use a school fixed effects model, but their models do not contain principal fixed effects, meaning that they do not explicitly estimate the extent to which principals improve over time. Rather, their experience estimates capture both the returns to experience and "ability" bias generated by non-random attrition from the principalship. They find that, within a given school, performance is higher when led by a principal with more experience.

To my knowledge, this is the first study to estimate the returns to principal experience using both principal and school fixed effects. I find substantial returns to experience for student achievement. The returns are largest for math achievement—relative to a principal's first year in the principalship, the average student in the principal's school scores 0.065 SD higher on statewide exams when the principal has 5 years of experience. Student test scores continue to improve up to 14 years—the highest value of experience I can observe in my data. Additionally, principals' ratings from supervisors increase by more than 0.45 standard deviations in their first five years on the job, on average, which moves the typical principal from the 35th to the 53rd percentile in the statewide distribution of scores.

Leveraging principals who work in multiple schools across the study period, I find that the returns to principal experience for student achievement are driven by school-specific rather than total experience. Put differently, improvement as measured by higher student test scores does not transfer across schools—principals effectively "start over" at their new school. The non-portability of improvement highlights the cost of frequent principal turnover, as it may take several years for leaders—even those with prior principal experience—to drive improvements in student learning.

I proceed first by reviewing the existing literature that relates principal experience to school performance, which helps to frame the contribution of this study. I also provide a framework for examining the portability of principals' accumulated skills. Next, I describe the data, measures, and methods used to estimate the returns to principal experience. I then describe the results for the returns to experience, the portability of improvement, and heterogeneity in improvement by school context. The final section concludes with the implications of the study for policy and suggestions for future research.

I.1 Connecting Principal Experience and School Performance

An increasing body of work demonstrates that principals have substantial effects on student learning. For instance, recent estimates of principal effects on math achievement range from 0.05 to 0.20 student-level standard deviations (Branch et al., 2012; Grissom et al., 2015a; Dhuey and Smith, 2018). In other words, a principal who is 1 standard deviation above the mean in terms of quality increases student growth by 0.05 to 0.20 standard deviations. While these estimates tend to be smaller in magnitude than teacher effects, principals affect the learning of *every* student in the school, which further underscores the importance of high-quality leadership for student success.

One limitation of existing work is that principal quality typically is treated as fixed, and principals are either "effective" or "ineffective." However, given the complex nature of school leadership, it is likely that effectiveness comes from a broad set of skills that principals develop over time. For instance, principals have responsibilities across many domains, including administrative tasks (e.g., managing student discipline, fulfilling compliance requirements), instruction management (e.g., conducting classroom observations), and internal relations (e.g., developing relationships with staff members).[1] Further, many of the skills important for effective leadership are not necessarily those developed from experience as a classroom teacher. While most principals complete certification programs and serve as assistant principals, these training experiences likely do not cover the full range of a principal's responsibilities. There is strong reason to expect that novice principals have yet to develop

[1]See, for example: Gates et al. (2003); Horng et al. (2010); Urick and Bowers (2014)

the full range of skills required for effective leadership, such that they become more effective as they gain experience on the job. Further, while the issue of principal improvement has received little attention, ample evidence demonstrates substantial job performance returns to experience for teachers, particularly in their first few years in the classroom (e.g., Rockoff, 2004; Papay and Kraft, 2015; Ladd and Sorensen, 2017).

Despite a strong conceptual basis for expecting that principals improve over time, there is little empirical evidence documenting returns to principal experience, and the findings from the handful of prior studies are mixed. One major reason for this inconsistency is that differences in empirical approaches among studies produce parameters that have different interpretations. For instance, using cross-sectional data, Eberts and Stone (1988) find a strong positive correlation between principal experience and student achievement, while Brewer (1993) find no evidence of a relationship. However, it is difficult to draw good inferences about principal improvement from these studies since they exploit both across- and between-principal variation in experience. Instead, any correlation between principal experience and student achievement likely conflates three processes: (1) the returns to principal experience (i.e., within-principal improvement), (2) systematic sorting of principals to certain types of schools over their careers, and (3) nonrandom attrition of less (or more) effective principals.

More recent studies have addressed nonrandom sorting of principals to schools using a school fixed effects approach (Clark et al., 2009; Bastian and Henry, 2015; Grissom et al., 2018). Inclusion of school fixed effects effectively compares the performance of principals who lead the same school in different years, which alleviates bias from unobserved school characteristics (to the extent they are fixed over time). Models using school fixed effects have consistently found a positive association between experience and principal effectiveness (Clark et al., 2009; Bastian and Henry, 2015; Grissom et al., 2018). As noted by Clark et al. (2009), however, these estimates reflect the combined effect of any returns to principal experience and "ability" bias induced by less effective leaders leaving the principalship. To my knowledge, no studies have employed models that account for both nonrandom attrition and principal-school sorting to isolate the returns to principal experience.[2]

I.1.1 The Portability of Principal Improvement

Human capital theory distinguishes between general and firm-specific or industry-specific skills (Becker, 1962). General human capital increases worker productivity not only in the current firm but in any other firm, while firm-specific human capital only increases productivity only in the current firm. In the case of principals, distinguishing between the returns to general versus firm-specific (school-specific) experience is important for two reasons. First, it tells us something about the nature of principal improvement. For instance, what are the actual skills that principals are building that lead them to better job performance? Second, distinguishing between general and school-specific returns to principal experience informs policy debates around the distribution of experienced principals. If the returns to school-specific experience are large relative to general experience, policies aimed at reallocating highly experienced principals to struggling schools may be less effective than those that promote the retention and development of existing leaders in such schools.

Assuming some amount of on-the-job learning for principals, it is unclear, *a priori*, the extent to which this improvement constitutes general or firm-specific (school-specific) human capital. On one hand, principals' responsibilities may not vary widely across schools, particularly schools of the same level (e.g., elementary schools), which would suggest that productivity increases from on-the-job learning carry over when the principal changes schools. On the other hand, schools are complex ecosystems that have specific strengths and challenges that principals must adapt to and learn from over time.

As an illustrative example, consider one of the primary channels through which principals affect student learning: hiring and retention of teachers (Jacob, 2011; Loeb et al., 2012). Prior work finds that effective principals may improve the quality of their school's

[2]Grissom et al. (2018) demonstrate a within-principal correlation between supervisor ratings and experience, which is suggestive of positive returns to experience. However, the results are not definitive with respect to improvement because the models do not contain any other covariates.

teachers by engaging in strategic hiring and retention of effective teachers (Loeb et al., 2012; Grissom and Bartanen, 2019b). Improving the composition of the teaching staff likely requires both general and school-specific human capital. For example, principals needs to be able to identify high and low performers (Grissom and Bartanen, 2019b). Even with the widespread adoption of multiple-measure teacher evaluation systems that explicitly aim to facilitate differentiation of teacher quality, principals must rely on their own judgment in weighing different signals of teacher performance (e.g., formal classroom observations, informal walk-throughs, value-added measures). As principals gain experience with hiring teachers and making retention decisions, they may learn which signals are more reliable in terms of predicting future effectiveness, which ultimately leads to a more effective teaching staff and increased student learning. In this situation, on-the-job learning constitutes an increase in principals' general human capital, as the ability to accurately predict teacher effectiveness should increase principal job performance at any school.

Alternatively, an important facet of retaining effective teachers, particularly in disadvantaged schools, is building a positive school climate—an atmosphere where teachers feel a sense a collegiality, trust, and support (e.g., Brown and Wynn, 2009; Johnson and Birkeland, 2003). Creating and maintaining a positive school climate may require principals to form individual relationships with teachers over time. Additionally, the dynamics of schools—even those in the same neighborhood—may vary widely. Even an experienced principal who changes schools must build relationships with the new teaching staff and adapt to the specific context over time. Here, on-the-job learning is school-specific, as the relationships and trust built among teachers in one school do not carry over to the next school.

To summarize, there are several gaps in our understanding of the returns to principal experience. First, prior work documents a positive correlation between principal experience and school performance. However, it is unclear the extent to which this relationship is explained by principal improvement as opposed to systematic attrition of less effective leaders from the principalship. Additionally, prior studies have almost exclusively focused on student achievement,[3] despite agreement that principal effects on test scores are indirect. Finally, the issue of portability has been almost completely ignored. The remainder of the study focuses on helping to fill these gaps.

I.2 Data, Sample, and Measures

This study analyzes longitudinal administrative data from Tennessee covering the 2001–02 through 2016–17 school years, provided by the Tennessee Department of Education (TDOE) via the Tennessee Education Research Alliance at Vanderbilt University. The Tennessee data contain detailed information about all employees in the K–12 public school system, including job title, school placement, and demographic information. I connect these staff data to student files beginning in 2006–07. The student data include demographic, enrollment, attendance, and discipline information, as well as achievement scores on statewide end-of-year exams for grades 3–8 and end-of-course exams for high school students.

I.2.1 Measuring Principal Experience

Like many statewide administrative datasets, the Tennessee data do not contain measures of job-specific experience. This means that while I can observe how long a given principal has worked in K–12 public education in Tennessee, I cannot observe how long they have been a principal if they entered the principalship in 2002 or earlier, which is the first year of the staff data files. Similarly, I cannot observe years of school-specific experience for individuals who were a principal in 2002 until they move to a different school. There are two ways to address this data limitation. My preferred approach is to treat principal experience as missing in cases where I cannot definitively determine the true experience value. This effectively drops principals who entered the principalship in 2002 or earlier. The primary disadvantage of this approach is that it potentially limits external validity, because the returns to principal experience are identified only from those individuals who entered the principalship in 2003 or later. Relatedly, I am only able to estimate the returns to experience up to 14 years. An

[3] Clark et al. (2009) and Grissom et al. (2018) being the most notable exceptions.

alternative approach is to retain the full sample of principals and top-code experience (e.g., 10+ years). While this avoids the external validity issue, the trade-off is increased risk of bias, particularly if principal effectiveness continues to increase (or decreases) beyond the first 10 or 15 years.[4] Additionally, because I cannot observe the true experience value for these principals, there is no clear way to estimate a model that would recover the returns to experience after 14 years even when keeping the entire sample.

Figure I.1 shows the statewide distribution of total principal experience and school-specific experience. In 2017, only 6.7% of principals in Tennessee had 15 or more years of total principal experience, and only 3.9% had been in the same school (as a principal) for 15 or more years. Roughly two-thirds of principals had six or fewer years of prior principal experience, and the typical principal was in their fifth year in their current school. Figure I.1 demonstrates that the restriction of estimating the returns to experience up to 14 years is not a major limitation, as most principals do not remain in the principalship for very long.

Table I.A1 shows, for each analysis year, the proportion of principals that have non-missing total and school-specific experience. In the earliest years, a majority of principals entered the principalship prior to 2002, and thus have missing values for total principal experience. For instance, only 29% of principals working in Tennessee schools in 2007 have observable prior experience. However, this proportion increases steadily over time, as the rate of principal attrition is high. By 2012, I can observe prior experience for two-thirds of principals, and 84% of principals in 2017. For school-specific experience, the missing data problem is less severe, as principals change schools relatively frequently. In 2007, I can observe school-specific experience for half of principals, up to 93% in 2017.

Table I.1 shows descriptive statistics for Tennessee principals between 2006–07 and 2016–17. Among those for whom I can observe prior experience, the average principal has 3.1 years of prior principal experience and has been in her school as a principal for 2.7 years. A large portion of principals are both new to the principalship and new to their schools. Thirty-five

[4] I explain this reasoning more thoroughly in the methods section when discussing the censored growth model.

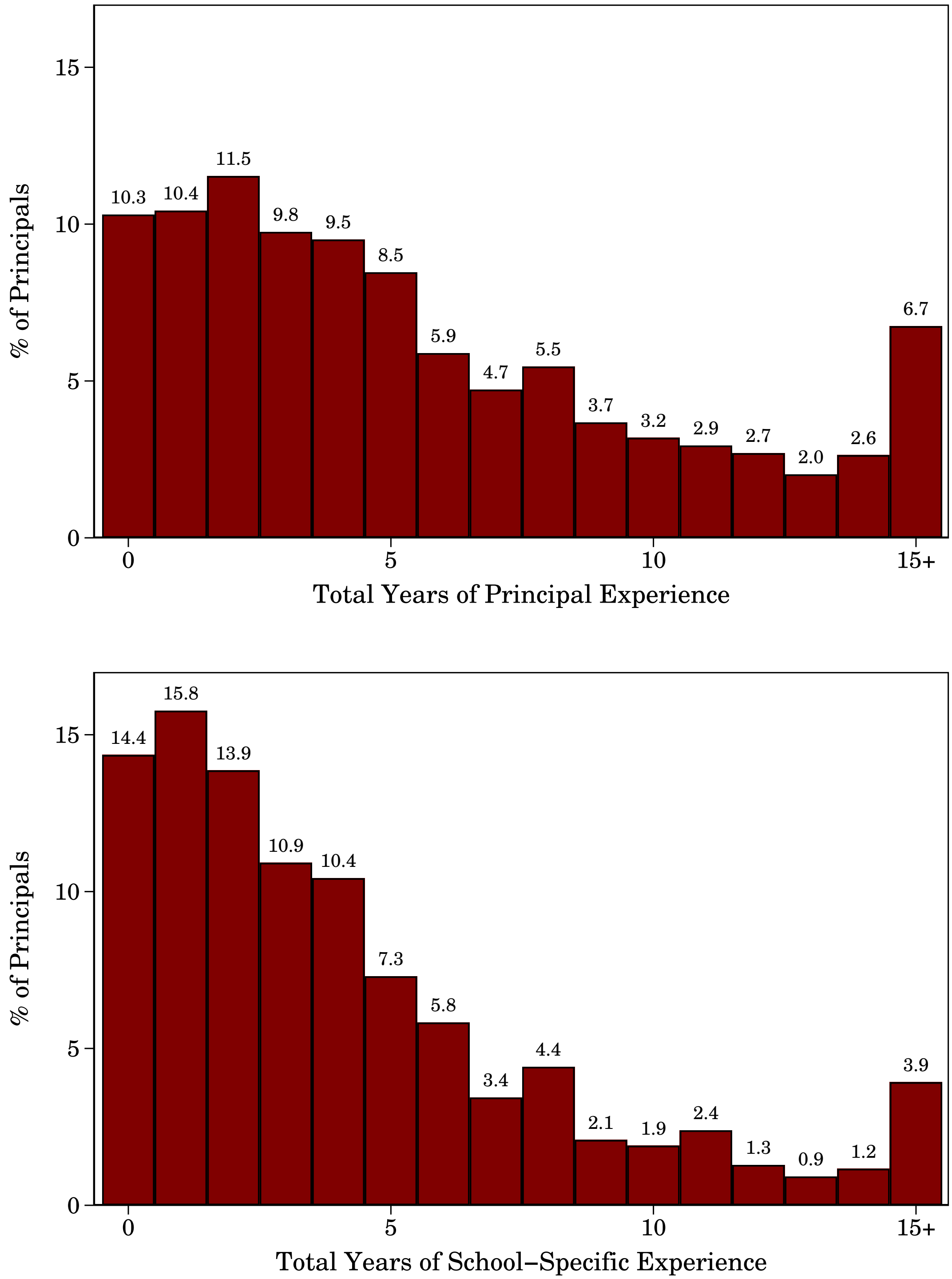


Figure I.1: Distribution of Principal Experience in Tennessee (2017)

Notes: Zero experience refers to a principal in their first year as a principal. School-specific experience only counts years served as the principal.

percent of principals have fewer than two years of prior principal experience, and 41% have been in their current school (as a principal) for fewer than two years. The average principal in Tennessee is 50 years old and has worked in the public school system for roughly 23 years. Almost 60% of principals work in elementary schools, and more than half of Tennessee's schools are located in areas classified as town or rural.

To separate the returns to total experience as a principal and school-specific experience, there must be a sufficient number of principals who work in multiple schools across the study period. Figure I.A2 shows the distribution of total principal experience versus school-specific experience. Each dot is a principal-year observation, with random jitter added to show relative density. The diagonal (i.e., equal amounts of total experience and school-specific experience) have the most observations, demonstrating that many principals in the sample are observed in just a single school. Nevertheless, there are a fair number of principals who move between schools, and thus have different amounts of total and school-specific experience.

I.2.2 Measuring Principal Effectiveness

I.2.2.1 Student Outcomes

The main outcome of interest in this study is student achievement. Specifically, I draw on achievement scores in math, reading, and science for students in grades 3–8 and end-of-course (EOC) exams for high school students. The grade 3–8 exams are required for every student across each year of the study period, while the EOC exams vary by year. In 2016–17, students took exams for Algebra I, Geometry, Algebra II, English I, English II, English III, Chemistry, and Biology. Earlier years had fewer tested subjects in high school. I construct a common measure of student achievement by standardizing exam scores within subject, grade, and year for grades 3–8. For EOC exams, which can have students from multiple grades (e.g., the Algebra I exam includes large numbers of ninth and tenth grade students), I standardize scores within each course and year.

Table I.1: Descriptive Statistics

	Mean	SD	Min	Max	N
Principal Experience					
Total years	3.1	2.7	0	14	11577
0 years	0.17				11577
1 year	0.18				11577
2–3 years	0.28				11577
4–6 years	0.24				11577
7–9 years	0.10				11577
10–14 years	0.03				11577
Tenure in School					
Total years	2.7	2.6	0	14	14475
0 years	0.21				14475
1 year	0.20				14475
2–3 years	0.29				14475
4–6 years	0.21				14475
7–9 years	0.07				14475
10–14 years	0.02				14475
Principal Demographics					
Black	0.19				18079
Male	0.45				18079
Age	50.0	9.1	19	93	17788
Experience in TN system	22.6	9.4	0	66	17992
School Demographics					
Enrollment size (100s)	6.45	3.82	0.14	40.65	18044
Proportion FRPL	0.57	0.26	0.00	1.00	18017
Proportion Black	0.25	0.31	0.00	1.00	18017
Proportion Hispanic	0.06	0.09	0.00	0.74	18017
Proportion Gifted	0.02	0.03	0.00	0.56	18017
Proportion SPED	0.15	0.08	0.00	1.00	18017
School Level					
Elementary	0.59				17992
Middle	0.19				17992
High	0.18				17992
Other	0.05				17992
School Locale					
Urban	0.31				17995
Suburban	0.15				17995
Town	0.16				17995
Rural	0.39				17995

Notes: Includes principals in Tennessee from 2006–07 to 2016–17. Unit of observation is principal-by-year.

In addition to student test scores, I also examine two non-test outcomes: attendance and discipline. The Tennessee data contain daily attendance information, with absences categorized as "excused" or "unexcused." To account for differences in the number of school days and students who are not enrolled at a single school the entire year, I construct attendance rates that represent the percentage of school days (for which a student was enrolled at the given school) that a student was present at school. I also disaggregate absences to construct excused and unexcused absence rates. The discipline data contain details about the type of offense, whether the student was suspended in-school or out-of-school, and the length of the suspension. For this analysis, I construct a binary outcome for whether a student was ever suspended in the current school year.

I.2.2.2 Ratings from Supervisors

As an alternative to using changes in student outcomes as a proxy for principal performance, I also draw on (plausibly) more direct measure of principals' practice: ratings from their supervisors. These ratings are rubric-based scores that principals receive as part of Tennessee's statewide educator evaluation system (TEAM) implemented in 2011–12. Fifty percent of the TEAM evaluation for principals comes from ratings of principal performance on a rubric derived from the Tennessee Instructional Leadership Standards.[5] These ratings are based on formal observations conducted by the principal's supervisor. Prior work shows that principals' ratings across indicators are highly inter-related and can be reduced to a single underlying performance score using factor analysis (Grissom et al., 2018). In this analysis, I use principals' average yearly observation scores—the exact measure used by the state to calculate summative evaluation ratings. I refer to this measure as "supervisor ratings."[6]

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

[6] Using the average observation score instead of the factor score described in Grissom et al. (2018) allows me to include principals in districts that used alternative observation rubrics (approximately one-quarter of principals in the state), as these districts do not report domain-specific scores for principals. However, for principals for whom I can calculate factor scores, the average observation score and the factor score are correlated at 0.95 or higher each year.

I.2.2.3 Teacher Outcomes

The final outcome I examine is teacher turnover. Prior work demonstrates that high-quality principals retain teachers at higher rates (Boyd et al., 2011; Ladd, 2011; Grissom and Bartanen, 2019b). Therefore, I also examine the extent to which principals improve at retaining teachers as they gain experience. More specifically, I construct a binary and multinomial measure of teacher turnover. The binary measure takes a value of one if teacher i in school s in year t is no longer a teacher in school s in year $t+1$, and zero otherwise. The multinomial measure categorizes three types of teacher turnover: exits from the state education system, moves to a teaching position in a different school, and changes to a non-teaching position (e.g., assistant principal, instructional coach).

I.3 Methods

I.3.1 Research Question 1: To what extent do principals become more effective as they gain experience?

My first research question seeks to estimate the job performance returns to principal experience. I estimate via ordinary least squares models of the general form:

$$Y_{ist} = \delta Experience_{it} + \gamma \mathbb{X}_{st} + \mu_i + \psi_s + \tau_t + \epsilon_{ist} \tag{I.1}$$

where Y is the performance of principal i in year t and δ is the marginal effect of principal experience. As discussed above, in addition to direct measures of principal performance, I also examine whether there are returns to principal experience for student- and teacher-level outcomes. These models follow the same form as principal-level models but include additional covariates, which I explain below. Equation 1 also includes a vector of school characteristics ($\mathbb{X}$): enrollment size and average student demographics (race/ethnicity, free/reduced-price lunch eligibility, gifted status, special education status). Finally, I include fixed effects for principal (μ_i), school (ψ_s), and year (τ_t).

The inclusion of principal, school, and year fixed effects are critical to the identification of δ. Principal fixed effects isolate within-principal variation in experience, such that the effects of additional experience are identified by comparing student outcomes under the same principal across years. An unbiased estimate of δ in a model without principal fixed effects requires that the accumulation of experience is uncorrelated with any fixed differences in principal quality. Prior work demonstrates that less effective principals are more likely to exit the principalship (Grissom and Bartanen, 2019a), which highlights the importance of including principal fixed effects.

School fixed effects control for time-invariant differences between schools, such as the quality of facilities or neighborhood effects. If principals were randomly assigned to schools, accounting for school heterogeneity would not be necessary. However, prior work demonstrates that principals may seek to sort to more advantaged schools over time (e.g., Béteille et al., 2012). Including school fixed effects helps to ensure that the returns to experience are not conflated with sorting to higher-quality schools.[7]

Year fixed effects account for any state-level factors that are correlated with both the given outcome and the accumulation of experience. In particular, I must account for any systematic *trends* in the outcome, which I would otherwise attribute to the returns to experience. The nature of the year fixed effects—and what they actually account for—depends on the particular outcome variable and whether it has been standardized within year. In the case of an unstandardized variable, such as teacher turnover, the year fixed effects will capture any time-varying factors that change the turnover propensity all teachers in the state, such as the implementation of a high-stakes educator evaluation system, law changes affecting teachers' tenure and due-process protections, or labor market conditions. For outcome variables that have been standardized within year, the year fixed effects account for average changes in the distribution of principal quality over time. If, for instance, the quality of new principals

[7]Here, I refer to differences in school quality in terms of factors that the principal cannot control. Clearly, principals themselves are an input to school quality. However, there are many school-level factors that are more less fixed over time, cannot be controlled by the principal, and contribute to student learning. Some examples are the neighborhood in which the school is located, the amount of resources the principal can access, and the quality of school facilities.

is increasing over time and the outcome variable is standardized within year, estimates of the returns to principal experience in a model without year fixed effects will be biased downwards.[8]

I.3.2 Separately Identifying Year Fixed Effects and the Returns to Experience

An important consideration in estimating equation 1 is how to parameterize principal experience. Modeling within-person returns to experience (i.e., including principal fixed effects) means that experience is perfectly collinear with year for most principals.[9] This collinearity means that identifying both the returns to experience and year fixed effects requires additional identification assumptions or sample restrictions. Papay and Kraft (2015) discuss three approaches for identifying identifying year fixed effects in the context of teacher fixed effects models, which I outline below.

I.3.2.1 Approach 1: Place Restrictions on the Experience Profile

The first approach—and the one that is most common in the literature—is to exploit regions of the experience profile where the marginal returns to experience are zero (or small in magnitude).[10] For instance, Rockoff (2004) implements a "censored growth model" whereby the returns to experience beyond a teacher's tenth year are restricted to be zero. Using this restriction, Rockoff (2004) identifies the year fixed effects from the subset of teachers who have more than ten years of experience. A related approach is the "indicator variable model," which places restrictions throughout the experience profile by constructing experience "bins"

[8]To see why, consider a simplified example where principal performance improves by x with each additional year of experience and the average quality of entering principals also improves by x each year. In this scenario, as long as principals leaving the profession are not systematically above average in terms of effectiveness, the distribution of principal quality increases across years. However, the outcome variable does not measure true principal performance, but rather reflects a principal's performance relative to the average principal in that year. A given principal, then, who improves by x each year, appears to improve less than x because of the global mean shift in the standardized outcome.

[9]More specifically, experience and year are perfectly collinear for principals who do not have discontinuous careers. For individuals who leave the principalship (i.e., move to central office or take a year off) and then return, experience and year will not be perfectly correlated. I discuss this case further below.

[10]See Rockoff (2004); Kraft and Papay (2014); Papay and Kraft (2015); Harris and Sass (2011); Ladd and Sorensen (2017) for examples.

(e.g., 0, 1–2, 3–5, 6–10 years, etc.). The censored growth model is effectively a special case of the indicator variable model that uses a single bin (10+ years). The identifying assumptions of these models are similar; unbiased estimates of the returns to experience require that the marginal effect of experience is zero within the specified bins. Any productivity growth (decline) within these bins will lead to upward (downward) bias in the estimated year fixed effects, which will downwardly (upwardly) bias the estimated returns to experience.

The choice of how to construct the experience bins is arbitrary, though researchers typically draw on prior empirical findings. For example, numerous studies demonstrate that teachers improve most rapidly in their first few years on the job (e.g., Papay and Kraft, 2015; Harris and Sass, 2011; Ladd and Sorensen, 2017), which suggests that placing restrictions on productivity growth towards the beginning of the experience profile will lead to conservative estimates of the returns to experience. Of course, the true shape of the experience profile is unknown and could vary substantially across contexts (e.g., state-level versus district-level datasets, urban versus rural schools), which means that relying on prior findings is not a perfect solution. Papay and Kraft (2015) propose two checks for the plausibility of the restrictions on the experience profile. The first is to simply examine the estimates near the cutoff points; evidence of productivity growth near the censoring point(s) would suggest that the returns to experience within these bins are not zero.[11] The second check is to split the bins (e.g., 5–10 years becomes 5–7 and 8–10 years) and compare the estimates to the initial model. If the zero growth assumption holds, narrow bins should produce an estimated experience profile similar to wider bins.

I.3.2.2 Approach 2: Leverage Discontinuous Careers

Whereas approach 1 relies on assumptions about the shape of the experience profile, a different approach is to circumvent the perfect collinearity between experience and year by leveraging individuals who have "discontinuous careers." Some teachers temporarily leave

[11] Rockoff (2004) also implements this check and finds that his censoring point at 10 years is reasonable for most, though not all, of his outcomes.

the profession such that they do not always accumulate additional experience each year. Without placing restrictions on the experience profile, one can identify both the returns to experience and year fixed effects. As noted by Papay and Kraft (2015), the discontinuous career approach faces both internal and external validity concerns. Teachers with discontinuous careers tend to be a small subset of the sample, which raises concerns that the returns to experience (or, equivalently, the year fixed effects) for these teachers are not generalizeable to teachers with continuous careers. In terms of internal validity, this approach assumes that temporarily leaving the profession has no effect on the returns to experience. This assumption is violated if, for instance, taking medical leave has a negative shock on teacher effectiveness upon returning.[12] An additional limitation of this approach is that even if the assumptions holds, the estimates can be very imprecise given the small number of individuals with discontinuous careers.

I.3.2.3 Approach 3: Leverage Between-Person Variation

A final approach proposed by Papay and Kraft (2015) is a two-stage model that produces estimates of the year fixed effects in the first stage then applies these coefficients to the second-stage model when estimating the returns to experience. Specifically, the first-stage model omits teacher fixed effects, such that the year fixed effects are identified from between-teacher variation. The key assumption of this approach is that there is no change in the quality of the teachers entering the profession (among those in the particular sample) over time. As Papay and Kraft (2015) discuss, there are many plausible reasons why this assumption would not hold. In particular, they suggest that policy reforms that have lowered barriers to entry through alternative certification and improvements in teacher preparation programs could have led to increases in the quality of new teachers. Papay and Kraft (2015) show that under this scenario, the estimated returns to experience in the two-stage approach

[12] Papay and Kraft (2015) find evidence of negative productivity shocks for teachers with temporary absences from the profession. Specifically, they estimate modified versions of the discontinuous career model with indicators for the year immediately before and after a discontinuity, finding that teacher effectiveness is lower in both of these years.

would be biased downwards.[13]

Given the dearth of evidence on principals, it is unclear which of these approaches is best-suited for estimating the returns to principal experience, which leads me to estimate models using each of them. Specifically, I estimate an indicator variable model using the following experience ranges: 0, 1, 2, 3, 4–6, 7–9, 10–14 years. While this choice of bins is based on the assumption that the returns to principal experience are largest in the first few years, I test the sensitivity of the estimates to different bins. Additionally, I estimate a model where growth is censored after five years. Using a cutoff at a higher experience level is not feasible given data limitations and how few principals remain in the principalship over time. I also implement the two-stage model proposed by Papay and Kraft (2015) by omitting principal fixed effects in the first-stage (but still including school fixed effects) and applying the estimated coefficients for the year fixed effects to the second-stage model to estimate the returns to principal experience. For the discontinuous career model, I include a fully non-parametric specification of principal experience up to 14 years. However, this approach yields imprecise estimates and, for the sake of brevity, I simply provide the results in Appendix Table I.A2.

To examine the returns to experience for student outcomes, I estimate the following specification of equation 1:

$$Y_{igjst} = \delta Experience_{it} + \gamma \mathbb{X}_{st} + \eta \mathbb{Z}_{jt} + \mu_i + \psi_s + \sigma_g + \tau_t + \epsilon_{igjst} \tag{I.2}$$

where Y_{igjst} is the achievement score (or, alternatively, attendance rate or suspension rate) of student j in grade g, with principal i, in school s, in year t. In addition to school characteristics, these models also adjust for student characteristics ($\mathbb{Z}_{jt}$): race/ethnicity, gender, free/reduced-price lunch eligibility, gifted and special education status, an indicator

[13]Both Papay and Kraft (2015) and Ladd and Sorensen (2017) both find evidence that the key identifying assumption of the two-stage model is not met for at least some subjects. In both cases, they find that the quality of new teachers is increasing over time, which biases downwards the estimated returns to experience in the two-stage model.

for grade repetition, and an indicator for whether the student was previously enrolled at a different school in the current year. Note that I do not control for students' prior achievement scores. Because most students remain in the same school between year $t-1$ and year t, they also tend to have the same principal in both years. The inclusion of prior-year achievement, then, is a violation of strict exogeneity in a model with principal fixed effects, as the principal in year t often affects the prior-year score. As a check, however, I also estimate (a) models that adjust for prior-year test scores and (b) models that adjust for a student's most recent test score in a prior school and find qualitatively similar results for the returns to experience.[14] I cluster standard errors at the principal-by-school level.

The model for principals' ratings from supervisors is:

$$Rating_{ist} = \delta Experience_{it} + \gamma \mathbb{X}_{st} + \mu_i + \psi_s + \tau_t + \epsilon_{ist} \tag{I.3}$$

where *Rating* is a principal's average score in year t, with scores standardized across the full sample of principals within each year to have a mean of zero and standard deviation of one. Besides the fixed effects, I also include controls, S_{st}, for time-varying school characteristics (enrollment size and school-level averages of student demographics). I cluster standard errors by school district.

Finally, I estimate the following teacher-level model:

$$Y_{ijst} = \delta Experience_{it} + \gamma \mathbb{X}_{st} + \eta \mathbb{Z}_{jt} + \mu_i + \psi_s + \tau_t + \epsilon_{ijst} \tag{I.4}$$

where Y is a binary indicator for teacher turnover (i.e., takes a value of one in year t if teacher j does not remain a teacher in the same school in year $t+1$). As in the student models, I control for personal (teacher) characteristics, $\mathbb{Z}_{jt}$, which include race, gender, age, experience, and highest education level. I cluster standard errors at the principal-by-school level.

[14] These results are shown in Table I.A3.

I make one very important modification when estimating the returns to principal experience for teacher turnover: I include an indicator for principal turnover. Prior work demonstrates that teacher turnover is greater in years where schools change principals (Miller, 2013; Bartanen et al., 2019), and principals' efforts to retain teachers may be less effective if the principal is not returning to the school in the following year. Additionally, the likelihood of principal turnover increases the longer that principals remain in the school (Grissom and Bartanen, 2019a), such that including these principal turnover years may lead to the conclusion that more experienced principals are less effective at retaining teachers. Given that this analysis focuses on identifying improvement, adjusting for the final year of a school spell ensures that identification comes only from years when the principal should be actively working to retain teachers.

I.3.3 Research Question 2: Are the returns to principal experience driven by total or school-specific experience?

My second research question seeks to examine whether principal improvement is driven by improvements in general or school-specific skills. Here, I exploit the fact that some principals work in multiple schools over their careers to separately identify the returns to *total* principal experience and *school-specific* principal experience:

$$Y_{ist} = \delta Experience_{it} + \theta ExperienceSchool_{ist} + \gamma \mathbb{X}_{st} + \mu_i + \psi_s + \tau_t + \epsilon_{ist} \tag{I.5}$$

If principal improvement over time reflects an increase in skills that are fully portable across schools, controlling for school-specific experience should not appreciably change estimates of δ relative to equation 1. Conversely, if improvement is not portable, estimates of δ will be attenuated while estimates of θ will be positive. As mentioned above, successfully separating the returns to total and school-specific experience requires principals who I observe in multiple schools. Across the study period (2007–2017) and sample (principals for whom I can determine total experience), 19% of the 2,500 unique principals worked in more than one school.[15]

I.3.4 Research Question 3: To what extent is there heterogeneity in the returns to principal experience across school contexts?

My third research question examines heterogeneity in the returns to principal experience across school contexts. To be specific, I examine whether principals in certain types of schools improve more or less rapidly over time. Because principals may sort to (or away from) certain types of schools over their careers (e.g., moving from elementary to high schools), I focus on heterogeneity in the returns to school-specific experience. I examine three contextual variables: student poverty (low-, medium-, and high-poverty),[16] school locale type (urban, suburban, and town/rural), and school level (elementary, middle, high).[17] To test for heterogeneity, I include an interaction between the given school contextual variable and school-specific experience:

$$\begin{aligned} Y_{igjst} = {} & \theta ExperienceSchool_{ist} + \eta(ExperienceSchool_{ist} \times Context_s) \\ & + \gamma \mathbb{X}_{st} + \mu_i + \psi_s + \omega(Year_t \times Context_s) + \epsilon_{ist} \end{aligned} \tag{I.6}$$

η represents the difference in the returns to principal experience relative to the arbitrary holdout group. Positive (negative) estimates of η would indicate that principals in the given school category improve more (less) rapidly than principals in the omitted school category. Note that the school context variables are time-invariant and thus the main effects are absorbed by the school fixed effect. I also replace year fixed effects with year-by-context effects, which allows any trends in unobserved determinants of principal effectiveness to be different

[15] Specifically, 16% worked in two schools, 2.5% worked in three schools, and 0.5% worked in more than three schools. Figure I.A2 plots total experience versus school-specific experience for each principal-by-year observation, with random jitter added to show density at discrete values of experience.

[16] I construct these categories to be time-invariant across the study period by taking the median of the proportion of students that qualify for free/reduced price lunch (FRPL) at the school in each year. The low-poverty group includes schools with fewer than 30% FRPL students, medium-poverty includes 30–80% FRPL, and high-poverty includes 80% or higher FRPL.

[17] I drop from the heterogeneity analysis the small number of schools that are classified as "other" school level by NCES.

by school contextual category. This modification is critical to avoid misattributing heterogeneity in productivity trends to heterogeneity in the returns to principal experience. If, for example, changes in school accountability systems affect principal quality in high-poverty schools more than in low-poverty schools, including year fixed effects (which assume that time-varying shocks affect all principals/schools equally) will lead to bias in the estimates of heterogeneity in the returns to experience.

I.4 Results

The analysis proceeds in four parts. First, I present estimates of the returns to principal experience for student outcomes and supervisor ratings. Based on the findings for student achievement, I then propose an alternative approach to estimating the returns to experience that places restrictions on the year fixed effects rather than the experience profile. Second, I show results from models that separate the returns to total principal experience and school-specific experience. Third, I examine the relationship between principal experience and teacher turnover. Finally, I examine heterogeneity in the returns to experience for principals working in different school contexts.

I.4.1 The Returns to Principal Experience for Student Achievement

Table I.2 shows estimates of the returns to experience for student achievement in math, English/language arts (ELA), and science. For each outcome, I show results from the indicator variable, censored growth, and two stage models. While I also estimated the discontinuous career model (see Appendix Table I.A2), there was an insufficient number of principals to produce precise estimates.[18]

[18] Further, the documented relationship between school/principal performance and principal turnover (Bartanen et al., 2019; Grissom and Bartanen, 2019a) strongly suggests that the discontinuous career approach is problematic both in terms of internal and external validity. Specifically, principal turnover (which is the primary reason why a principal would fall into the discontinuous career group) is preceded by a drop in school/principal performance and principals who are less effective are more likely to turn over. Thus, almost by definition, discontinuous career principals are systematically less effective than the typical principal in Tennessee. While this is not problematic, per se, it suggests that the improvement trajectory among these principals may also be unrepresentative of the population. In terms of internal validity, the identifying assumption of the discontinuous career model is that the performance of these principals when they returned

Beginning with the estimates from the indicator variable model (IVM), I find positive returns to principal experience in math and science. For ELA, the coefficients are positive and increasing over time, but they are not statistically significant at conventional levels. Similar to findings for teachers, the IVM results show that principals improve most rapidly in the first few years. However, I find that the marginal returns to experience are positive throughout the experience profile. This implies that the identifying assumption of both the IVM and censored growth model——that there is a "flat" region of the experience profile which can be used to identify year fixed effects—is not met. By consequence, both the IVM and censored growth models should produce conservative estimates of the returns to principal experience. The censored growth model results in columns 4–6 are consistent with this expectation. Placing a restriction of zero returns to experience after a principal's fifth year leads to upward bias in the year fixed effects and downward bias in the returns to experience.[19] I return to this issue below.

Columns 7–9 show the estimated returns to principal experience from the two-stage model. Compared to the IVM model, the coefficients are substantially smaller in magnitude. The identifying assumption of the two-stage is that the quality of new principals between 2007 and 2017 (the study period) is unchanging. That the returns to experience are smaller compared to the IVM results suggests that this assumption does not hold. Specifically, that the estimates appear to be biased downwards suggests the quality of new principals *increased* over time.

to the principalship is the same (in expectation) as it would have been had they not left. Whereas teachers are more likely to have discontinuous careers for reasons plausibly orthogonal to performance (e.g., child-bearing), the reasons for principals temporarily exiting the principalship are more likely performance-related. Finally, the gap in experience typically corresponds to a change in school, making it difficult to assume that principal performance would have been the same in the absence of the move.

[19] Appendix Table I.A4 shows the results of the IVM using narrower experience bins. As expected given the apparent presence of within-bin growth in Table I.2, narrowing the bins leads to larger estimates of the returns to experience.

Table I.2: The Returns to Principal Experience (Student Achievement)

	IVM			Cen. Growth			2-Stage		
	Math (1)	ELA (2)	Sci (3)	Math (4)	ELA (5)	Sci (6)	Math (7)	ELA (8)	Sci (9)
Total Principal Experience									
0 years (base)									
1 year	0.017***	0.005	0.010*	0.013**	0.003	0.003	0.009*	0.005*	0.004
	(0.006)	(0.004)	(0.005)	(0.006)	(0.004)	(0.005)	(0.005)	(0.003)	(0.004)
2 years	0.023**	0.005	0.024***	0.014	0.000	0.011	0.007	0.004	0.012**
	(0.010)	(0.007)	(0.008)	(0.009)	(0.006)	(0.008)	(0.006)	(0.004)	(0.006)
3 years	0.041***	0.012	0.039***	0.028**	0.005	0.022*	0.018**	0.011**	0.023***
	(0.013)	(0.010)	(0.012)	(0.012)	(0.009)	(0.012)	(0.007)	(0.005)	(0.007)
4–6 years	0.065***	0.015	0.048***				0.028***	0.013**	0.027***
	(0.019)	(0.013)	(0.017)				(0.009)	(0.006)	(0.009)
7–9 years	0.066**	0.015	0.059**				0.010	0.013	0.030**
	(0.027)	(0.020)	(0.024)				(0.013)	(0.008)	(0.013)
10–14 years	0.099***	0.024	0.071**				0.023	0.021	0.031
	(0.037)	(0.025)	(0.033)				(0.018)	(0.015)	(0.019)
4 years				0.048***	0.009	0.027*			
				(0.015)	(0.011)	(0.015)			
5 years				0.038*	-0.001	0.022			
				(0.020)	(0.015)	(0.019)			
N	3034743	3328312	3029986	3034743	3328312	2819174	3034743	3328312	2819174
R^2	0.301	0.319	0.341	0.301	0.319	0.339	0.302	0.319	0.339
Joint Test of Year FE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Notes: Standard errors clustered by principal-school shown in parentheses. The dependent variables are standardized student achievement scores. The omitted category is principals who have zero years of experience. Models include fixed effects for principal, school, year, and grade. Covariates include student characteristics and time-varying school characteristics. "Joint Test of Year FE" shows the p-value from an F-test that the estimated year fixed effects are jointly zero. For the 2-stage model, this test refers to the year FE from the first-stage model. IVM = Indicator Variable Model. N refers to the total number of student-by-year observations.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

I.4.2 An Alternative Approach to Estimating the Returns to Principal Experience

Table I.2 suggests that while the returns to principal experience for student achievement are positive, the estimates from each of the models are conservative. However, the magnitude of the bias is unclear. To further explore this problem, I propose an alternative approach that places restrictions on the year fixed effects instead of the experience profile. The logic of this approach is to exploit flat regions of the *time* profile to identify the returns to principal experience instead of using flat regions of the experience profile to identify year fixed effects. In essence, this is a modified version of the indicator variable model, where the "buckets" are groups of adjacent years. Here, the identifying assumption is that, conditional on principal experience, there is no positive or negative productivity growth within the specified bins. Of course, identifying *which* years should be grouped together is critical to this approach. Whereas the choice of experience bins in the IVM approach is motivated by the hypothesis of decreasing marginal returns to experience, there is no theory to guide the choice of year bins. Instead, I rely on empirical estimates of the year fixed effects from the models estimated in Table I.2. The validity of this approach relies on the assumption that the estimated year fixed effects from these models accurately identify the true *shape* of the productivity trend across years.[20]

Table I.A5 shows the estimated year fixed effects from the student achievement models. In contrast to the experience profile, there appear to be regions where the estimated time trend is flat. Guided by these estimates, I specify bins that group adjacent sets of years, which I use to re-estimate the achievement models. For example, in the math models I replace year fixed effects with the following year bins: 2007, 2008–2009, 2010, 2011, 2012–2013, 2014–2015, 2016–2017.[21] Table I.A6 shows the achievement results replacing year

[20]Here I distinguish *shape* from *level* in saying that it is not necessary that the year fixed effects from the IVM are unbiased as long as the magnitude and direction of the bias between adjacent years is small enough to make an accurate judgment of whether the year-to-year productivity growth is small or large.

[21]I use slightly different bins for ELA and science: (ELA) 2007–2008, 2009, 2010–2011, 2012–2013, 2014, 2015, 2016, 2017; (Science) 2007, 2008, 2009, 2010–2011, 2012–2013, 2014, 2015–2016, 2017. There is no reason to assume that the time trends should follow the exact same pattern across subjects. For instance,

fixed effects with the constructed year bins (but not changing the experience variables). The estimated returns to experience are very similar to the estimates from the models with year fixed effects, supporting the validity of the modified year indicators.

Next, I replace the principal experience bins with a fully flexible set of year indicators, with the results shown in Table I.3. As expected, placing restrictions on the year fixed effects effects rather than the experience profile leads to larger estimates of the returns to experience, suggesting that the estimates in Table I.2 are indeed conservative. Further, the magnitude of difference between the estimates in Table I.3 and Table I.2 is fairly substantial, particularly at higher levels of experience. The returns to experience at 1 year, for instance, are 25–40% larger (depending on the subject) using my modified approach as compared to the indicator variable model. At 4–6 years, the difference in magnitude is more than 30% in math and more than 60% in ELA and science.

At a minimum, the results from the IVM approach and the approach using modified year bins demonstrate that a principal's ability to raise student achievement increases substantially with experience. Even conservative estimates using the IVM approach imply that the magnitude of principal improvement is large. In other work I estimate that the standard deviation of principal effects on math (ELA, science) achievement (i.e., the amount that student math achievement increases for a 1 SD increase in principal quality) is roughly 0.20 (0.10, 0.17) SD in Tennessee. Thus, the amount of within-principal improvement relative to the overall distribution of principal quality is quite large. Put another way, these results imply that the average principal substantially increases their rank in the effectiveness distribution over time. Next, I examine the extent to which principals' apparent improvement over time is corroborated using alternative outcomes.

there may be changes in standards or the end-of-year exams that affect math achievement differently than reading achievement. Nevertheless, the patterns in the estimated year effects in Table I.A5 are substantially similar across subjects, suggesting that the year effects are picking up general productivity trends that affect test score performance similarly in all subjects.

Table I.3: The Returns to Principal Experience Using Modified Year Bins

	Math (1)	ELA (2)	Sci (3)
Total Principal Experience			
0 years (base)			
1 year	0.021***	0.007**	0.014**
	(0.006)	(0.004)	(0.005)
2 years	0.031***	0.009	0.033***
	(0.008)	(0.006)	(0.009)
3 years	0.053***	0.018**	0.054***
	(0.012)	(0.008)	(0.013)
4 years	0.083***	0.026***	0.071***
	(0.015)	(0.010)	(0.017)
5 years	0.082***	0.020	0.074***
	(0.018)	(0.012)	(0.021)
6 years	0.092***	0.029**	0.088***
	(0.021)	(0.014)	(0.024)
7 years	0.096***	0.027	0.110***
	(0.025)	(0.017)	(0.029)
8 years	0.098***	0.038**	0.105***
	(0.028)	(0.019)	(0.033)
9 years	0.105***	0.030	0.104***
	(0.032)	(0.022)	(0.037)
10 years	0.141***	0.044*	0.116**
	(0.037)	(0.026)	(0.045)
11 years	0.132***	0.031	0.148***
	(0.044)	(0.030)	(0.046)
12 years	0.179***	0.085**	0.183***
	(0.045)	(0.036)	(0.056)
13 years	0.230***	0.120	0.194***
	(0.080)	(0.109)	(0.060)
14 years	0.075	0.034	0.185**
	(0.086)	(0.055)	(0.083)
N	3034743	3328312	2819174
R^2	0.301	0.319	0.339

Notes: Standard errors clustered by principal-school shown in parentheses. The dependent variables are standardized student achievement scores. Models include fixed effects for principal, school, year, and grade. Covariates include student characteristics and time-varying school characteristics. The year fixed effects are replaced with the following year bins: (math) 2007, 2008–2009, 2010, 2011, 2012–2013, 2014–2015, 2016–2017; (ELA) 2007–2008, 2009, 2010–2011, 2012–2013, 2014, 2015, 2016, 2017; (Science) 2007, 2008, 2009, 2010–2011, 2012–2013, 2014, 2015–2016, 2017. N refers to the total number of student-by-year observations.
* p < 0.1, ** p < 0.05, *** p < 0.01.

I.4.3 The Returns to Principal Experience for Supervisor Ratings

Table I.4 shows estimates of the returns to principal experience for supervisor ratings. Different from the estimates for student achievement, the IVM, censored growth, and two-stage approaches produce very similar results for the returns to experience. Across each model, principals improve substantially in their first five years on the job (the returns to experience up to five years are roughly 0.50 SD), with little to no evidence of returns to experience beyond five years. Given the lack of growth beyond five years, it is no surprise that the censored growth and indicator variable models produce similar estimates, since they both assume that the returns to experience are largest in a principal's first few years.

Why are the results from the two-stage model similar to those from the censored growth and indicator variable models? Whereas the year fixed effects are statistically significant (and large in magnitude) in the student achievement models, they have no explanatory power in the supervisor ratings models, as evidenced by the large p-values for F-tests of their joint significance shown in the bottom of Table I.4.[22] Further, the year fixed effects in the first stage of the two-stage approach are small in magnitude and not jointly significant, meaning that the second stage model is effectively equivalent to estimating a model that simply omits year fixed effects.

As with student test scores, the magnitude of the estimated returns to experience for supervisor ratings are substantial. On average, principals in Tennessee move from the 35th percentile to the 53rd percentile in supervisor ratings between their first and sixth year in the principalship. That said, it is important to note that the analysis of supervisor ratings faces an important limitation: part of the observed effect is likely driven by raters' knowledge of a principal's experience level and their expectation that principals improve with experience and/or that experienced principals are more effective than inexperienced principals. Still, the improvement in supervisor ratings supports the findings from the student achievement

[22]This is not driven by large standard errors. The coefficient estimates are uniformly small in magnitude. For instance, the coefficients for the year fixed effects in the IVM are (ascending by year with 2017 omitted): 0.015, 0.018, -0.033, -0.018, -0.01 SD for 2012–2016.

Table I.4: The Returns to Principal Experience (Supervisor Ratings)

	IVM (1)	Cen. Growth (2)	2-Stage (3)
Total Principal Experience			
0 years (base)			
1 year	0.208***	0.218***	0.215***
	(0.039)	(0.039)	(0.034)
2 years	0.319***	0.340***	0.333***
	(0.064)	(0.053)	(0.054)
3 years	0.367***	0.398***	0.387***
	(0.071)	(0.065)	(0.065)
4–6 years	0.462***		
	(0.092)		
7–9 years	0.443***		
	(0.123)		
10–14 years	0.448**		
	(0.187)		
4 years		0.487***	0.471***
		(0.081)	(0.081)
5 years		0.534***	0.510***
		(0.108)	(0.099)
6 years			0.515***
			(0.105)
7 years			0.501***
			(0.132)
8 years			0.501***
			(0.147)
9 years			0.475***
			(0.170)
10 years			0.476**
			(0.188)
11 years			0.600***
			(0.218)
12 years			0.552**
			(0.249)
13 years			0.447*
			(0.270)
14 years			0.157
			(0.386)
N	6924	6986	6986
R^2	0.710	0.712	0.713
Joint Test of Year FE	0.9839	0.9797	0.9886

Notes: Standard errors clustered by school district shown in parentheses. The dependent variables is a principal's average rating from their supervisor, standardized within year. Models include fixed effects for principal, school, and year. Covariates include time-varying school characteristics. "Joint Test of Year FE" shows the p-value from an F-test that the estimated year fixed effects are jointly zero. For the 2-stage model, this test refers to the year FE from the first-stage model. IVM = Indicator Variable Model. N refers to the total number of principal-by-year observations.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table I.5: The Returns to Principal Experience (Attendance and Suspensions)

	Att (1)	Abs (U) (2)	Abs (E) (3)	Susp (4)	ISS (5)	OSS (6)
Total Principal Experience						
0 years (base)						
1 year	0.011	-0.024***	0.013	0.001	0.000	0.001
	(0.007)	(0.008)	(0.009)	(0.002)	(0.002)	(0.001)
2 years	0.017	-0.021	-0.000	0.005	0.004	0.001
	(0.011)	(0.013)	(0.013)	(0.003)	(0.003)	(0.002)
3 years	0.022	-0.022	-0.009	0.008*	0.009*	0.001
	(0.015)	(0.017)	(0.017)	(0.004)	(0.005)	(0.003)
4–6 years	0.020	-0.017	-0.014	0.007	0.011*	-0.002
	(0.020)	(0.023)	(0.023)	(0.006)	(0.006)	(0.004)
7–9 years	0.024	-0.023	-0.010	0.010	0.017**	-0.002
	(0.028)	(0.032)	(0.033)	(0.008)	(0.009)	(0.005)
10–14 years	0.040	-0.027	-0.034	0.016	0.018	0.004
	(0.039)	(0.045)	(0.045)	(0.010)	(0.011)	(0.007)
N	6432890	6432890	6432890	6449145	6449145	6449145
R^2	0.083	0.082	0.081	0.201	0.161	0.177
Joint Test of Year FE	0.5885	0.7695	0.6395	0.0000	0.0000	0.0000

Notes: Standard errors clustered by principal-school shown in parentheses. The omitted category is principals who have zero years of experience. Models include fixed effects for principal, school, year, and grade. Covariates include student characteristics and time-varying school characteristics. "Joint Test of Year FE" shows the p-value from an F-test that the estimated year fixed effects are jointly zero. For the 2-stage model, this test refers to the year FE from the first-stage model. IVM = Indicator Variable Model. N refers to the total number of student-by-year observations.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

models that there are substantial returns to principal experience.

I.4.4 The Returns to Principal Experience for Student Attendance and Discipline

Table I.5 shows results from the IVM approach for student attendance and discipline outcomes. The first three columns estimate the returns to principal experience for student attendance, unexcused absences, and excused absences, respectively. Each of these outcomes is standardized within grade and year. Column 1 shows that while the estimated coefficients are positive and generally increasing in magnitude over time, none are statistically significant at conventional levels. However, I cannot rule out meaningful effects given the large standard errors, particularly at higher levels of experience.

Columns 2 and 3 examine whether the returns to principal experience may be more apparent for specific types of absences. Here, negative estimates would indicate "improvement."

As with column 1, there is no clear evidence that principals improve at lower student absenteeism as they gain experience. However, the pattern in column 2 suggests that principals with no prior experience may have students with somewhat higher unexcused absence rates. The *1 year* estimate indicates that students' unexcused absences rates are 0.024 SD lower comparing a principal's first and second year in the principalship. The coefficients for higher levels of experience are similar in terms of magnitude but increasingly imprecise.

Columns 4–6 show the results for student suspensions. Specifically, the dependent variable in column 4 is a binary indicator for whether the student was suspended one or more times during the school year. Columns 5 and 6 distinguish between in-school suspensions (ISS) and out-of-school suspensions (OSS). I find some evidence that the probability of in-school suspensions increases as principals gain experience, though the magnitude is modest and the coefficients are not consistently statistically significant. To provide a sense of magnitude, the 1.1 percentage point increase in the probability that a student receives at least one in-school suspension is roughly 13% of the base rate. For out-of-school suspensions (column 6), I find no evidence of a relationship with principal experience, with precisely estimated null effects.

I.4.5 Is Principal Improvement General or School-Specific?

The previous section establishes that principals improve substantially over time as measured by increases in student test scores and ratings from supervisors. Next, I investigate the extent to which this improvement is driven by general or school-specific skills. Specifically, I leverage principals who have worked in multiple schools across the study period to separate the returns to *total* experience and *school-specific* experience for student achievement.

The odd columns in Table I.6 show the relationship between school-specific experience and student test scores in math, ELA, and science. The estimates are similar to those for the returns to total principal experience in Table I.2, which is expected given that the majority of principals are only observed in a single school.[23] The even columns show the parameters

[23]The correlation between total experience and school-specific experience is 0.78.

Table I.6: The Returns to Total and School-Specific Principal Experience

	Math		ELA		Sci	
	(1)	(2)	(3)	(4)	(5)	(6)
Total Principal Experience						
0 years (base)						
1 year		0.012		-0.002		0.004
		(0.011)		(0.007)		(0.009)
2 years		-0.002		-0.019*		0.001
		(0.017)		(0.011)		(0.013)
3 years		0.009		-0.016		0.012
		(0.022)		(0.015)		(0.018)
4–6 years		0.030		-0.014		0.011
		(0.027)		(0.019)		(0.024)
7–9 years		0.008		-0.036		0.008
		(0.037)		(0.026)		(0.031)
10–14 years		0.031		-0.039		0.010
		(0.048)		(0.033)		(0.039)
School-Specific Experience						
0 years (base)						
1 year	0.018***	0.008	0.009**	0.011*	0.012**	0.009
	(0.006)	(0.010)	(0.004)	(0.006)	(0.005)	(0.009)
2 years	0.032***	0.034**	0.016**	0.032***	0.031***	0.030**
	(0.010)	(0.016)	(0.007)	(0.010)	(0.008)	(0.012)
3 years	0.053***	0.045**	0.027***	0.040***	0.046***	0.037**
	(0.014)	(0.022)	(0.009)	(0.015)	(0.013)	(0.018)
4–6 years	0.075***	0.051*	0.032**	0.042**	0.061***	0.052**
	(0.019)	(0.027)	(0.013)	(0.019)	(0.017)	(0.025)
7–9 years	0.095***	0.090**	0.051**	0.080***	0.079***	0.074**
	(0.028)	(0.038)	(0.020)	(0.026)	(0.026)	(0.033)
10–14 years	0.129***	0.104*	0.069**	0.099***	0.095**	0.089*
	(0.042)	(0.054)	(0.029)	(0.037)	(0.038)	(0.046)
N	3034743	3034743	3328312	3328312	3029986	3029986
R^2	0.301	0.301	0.319	0.319	0.341	0.341

Notes: Standard errors clustered by principal-school shown in parentheses. The dependent variables are standardized student achievement scores. The omitted category is principals who have zero years of experience. Models include fixed effects for principal, school, year, and grade. Covariates include student characteristics and time-varying school characteristics.
* p < 0.1, ** p < 0.05, *** p < 0.01.

of interest. Across all three subjects, the returns to principal experience are driven by school-specific rather than total experience. Put another way, the results in Table I.6 suggest that principal improvement (as measured by changes in student achievement) is not portable—that the skills principals acquire with experience in one school do not help them raise test scores in a different school.

Given my reliance on principals who I observe in multiple schools to separate the returns to total and school-specific experience, it is important to understand the extent to which these principals are representative of the broader sample. In Table I.A8, I examine whether the returns to school-specific experience are different between principals observed in a single school versus those observed in multiple schools. I find no evidence that the experience trajectories differ between these groups of principals.

A related concern, which applies to the analysis more broadly, is the high rate of attrition among principals. As mentioned above, roughly 80% of principals in the sample are observed in only a single school and the median tenure length is three years. By consequence, estimates of the returns to experience rely on an increasingly small subset of principals as experience increases. This may undermine the generalizability of the estimates to the broader sample if those who remain in the principalship have systematically steeper (or flatter) experience profiles. This is relevant for the policy implications of the study. If, for instance, the returns to experience merely reflect the fact that principals who improve most rapidly are those who stay in the principalship, then policies that curb principal attrition may not yield the expected benefits. To explore this possibility, I estimate a series of models that test for heterogeneity in the experience trajectory by how long the principal stays in their school.[24] The results, shown in Table I.A7, show no evidence of heterogeneity. In other words, I do not find that those who only remain in the principalship a few years improve at different rates than those who stay longer.

[24] Specifically, I estimate models that interact the returns to school-specific experience with an indicator for whether the principal stays in the school for at least x years, with separate models for x=2,3,5,10. Further, I exclude from the model principal-by-school spells shorter than x if they are left-censored, since I cannot determine how long these principals will stay in their school.

I.4.6 Principal Improvement and Teacher Outcomes

A key channel through which principals affect student outcomes is human capital management—the hiring and retention of teachers. Given that teachers are the most important school-level input to student learning (e.g., McCaffrey et al., 2003; Rivkin et al., 2005), the positive returns to principal experience for student achievement are likely to be mediated (at least in part) by principals' effects on teacher-level outcomes. In this section, I explore this potential pathway by estimating the returns to principal experience (total and school-specific) for teacher outcomes.

Table I.7: The Returns to Principal Experience (Teacher Turnover)

	All Turnover			Transfer			Exit			Position Change		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Total Principal Experience												
0 years (base)												
1 year	-0.003		0.000	-0.002		0.005	-0.002		-0.004	0.001		-0.002
	(0.004)		(0.008)	(0.003)		(0.007)	(0.003)		(0.006)	(0.001)		(0.002)
2 years	-0.005		-0.005	-0.005		0.007	-0.001		-0.014*	0.000		-0.002
	(0.006)		(0.011)	(0.005)		(0.009)	(0.004)		(0.007)	(0.001)		(0.002)
3 years	-0.009		0.000	-0.006		0.010	-0.007		-0.011	0.001		-0.003
	(0.008)		(0.013)	(0.006)		(0.011)	(0.005)		(0.009)	(0.002)		(0.003)
4–6 years	-0.012		0.010	-0.004		0.024	-0.011		-0.013	-0.000		-0.003
	(0.011)		(0.019)	(0.009)		(0.016)	(0.008)		(0.013)	(0.003)		(0.004)
7–9 years	-0.012		0.008	-0.005		0.030	-0.012		-0.023	0.000		-0.003
	(0.017)		(0.026)	(0.013)		(0.023)	(0.012)		(0.019)	(0.004)		(0.006)
10–14 years	-0.022		0.012	-0.011		0.042	-0.020		-0.033	0.002		-0.000
	(0.026)		(0.039)	(0.020)		(0.034)	(0.018)		(0.027)	(0.005)		(0.008)
School-Specific Experience												
0 years (base)												
1 year		-0.004	-0.004		-0.006*	-0.010		-0.000	0.003		0.001	0.003**
		(0.004)	(0.008)		(0.003)	(0.007)		(0.003)	(0.005)		(0.001)	(0.002)
2 years		-0.007	-0.002		-0.012**	-0.017*		0.004	0.015**		0.001	0.002
		(0.006)	(0.011)		(0.005)	(0.009)		(0.004)	(0.007)		(0.001)	(0.002)
3 years		-0.014*	-0.014		-0.015**	-0.023**		-0.003	0.005		0.002	0.004
		(0.009)	(0.014)		(0.007)	(0.012)		(0.006)	(0.009)		(0.002)	(0.003)
4–6 years		-0.023*	-0.031		-0.022**	-0.040**		-0.006	0.003		0.001	0.004
		(0.012)	(0.021)		(0.010)	(0.017)		(0.008)	(0.013)		(0.003)	(0.004)
7–9 years		-0.022	-0.027		-0.028*	-0.050*		0.000	0.019		0.001	0.003
		(0.020)	(0.030)		(0.016)	(0.026)		(0.013)	(0.020)		(0.004)	(0.006)
10–14 years		-0.044	-0.053		-0.053**	-0.085**		-0.004	0.022		0.003	0.003
		(0.029)	(0.044)		(0.022)	(0.038)		(0.020)	(0.029)		(0.006)	(0.008)
N	350843	350843	350843	320877	320877	320877	318333	318333	318333	296924	296924	296924
R^2	0.073	0.073	0.073	0.085	0.085	0.085	0.068	0.069	0.069	0.025	0.025	0.025

Notes: Standard errors clustered by principal-school shown in parentheses. The dependent variable is a binary indicator for whether the teacher left their position in the current year, with the type of turnover listed in the column header. Models include fixed effects for principal, school, and year. Models include teacher and school characteristics, and an indicator for whether the principal left the school at the end of the year.
* p < 0.1, ** p < 0.05, *** p < 0.01.

Tables I.7 shows the results for teacher turnover. Prior work demonstrates that teacher turnover has negative effects on student achievement (Ronfeldt et al., 2013), which suggests that principals who are able to retain teachers at higher rates may see improvements in student learning. Further, an important factor in teachers' decisions to remain in a school is the quality of school leadership (Grissom, 2011; Boyd et al., 2011; Kraft et al., 2016). In addition to a binary measure of teacher turnover (i.e., whether the teacher remains in the school between year t and $t+1$), I also three types of turnover: exits from the education system, transfers to a different school, and position changes (e.g., becoming a school counselor or an assistant principal).[25] Motivated by the results in Table I.6, I estimate models examining both total and school-specific principal experience.

Columns 1–3 show the results for a binary measure of teacher turnover. All models are estimated via OLS such that the coefficients reflect the marginal change in the probability that a teacher turns over. While the coefficients in column 1 (total experience) and column 2 (school-specific experience) are negative and increasing in magnitude at higher levels of principal experience (which would indicate lower teacher turnover rates), most are not statistically significant with large standard errors. Column 3, which includes both experience types, again suggests that the patterns are driven by school-specific rather than total principal experience.

The remaining columns examine specific types of teacher turnover. Column 5 shows that there are fairly substantial returns to school-specific experience in terms of lowering the probability that teachers transfer to another school. Compared to a principal's first year in the school, teachers are 0.6 percentage points less likely to transfer following the principal's second year, up to 2.2 and 5.3 percentage points at 4–6 and 10–14 years of school-specific experience, respectively. These effects correspond to 7%, 24%, and 59% of the average teacher transfer rate in Tennessee across the study period. For exits, the point estimates

[25] Specifically, I analyze this categorical outcome as a series of linear probability models where the base category is stayers. This is preferable to alternative modeling approaches (e.g., multinomial logistic regression) because of the inclusion of high-dimensional fixed effects.

are suggestive of a pattern for total experience in column 7, but they are not statistically significant. For position changes, the point estimates are consistently close to zero and not statistically significant.

An additional consideration in estimating the returns to principal experience for teacher turnover is that the composition of the teaching staff changes over time. New-to-school principals inherit the teaching staff of their predecessor(s), which they subsequently shape through hiring new teachers and retaining (or failing to retain) existing teachers. The returns to principal experience for teacher turnover, then, may be misleading (in terms of examining improvement) if the the average latent turnover propensity (i.e., individual factors that drive turnover decisions that the principal cannot control) of the teaching staff is correlated with a principal's length of tenure in the school. Further, any improvement in their ability to retain teachers may vary by whether the teacher was inherited or hired. Inherited teachers—who by definition have been in the school for a longer time and who tend to be more experienced—may be considerably less responsive to school leadership with respect to their mobility decisions.[26]

In Table I.8, I examine whether the relationship between principal experience and teacher turnover varies between inherited and hired teachers.[27] For ease of interpretation, I show the marginal effects of school-specific experience (of the principal) for these two groups of teachers.[28] Column 1 shows that while principals improve over time at lowering turnover rates of both inherited and hired teachers, the effect is much larger for hired teachers. As with the baseline models, these patterns are driven by lower rates of transfer among both hired and inherited teachers. However, column 3 shows that, as they gain experience in the

[26]Selection may also be a factor here. Inherited teachers presumably had the option to leave the school when the prior principal turned over. Those that remain may be systematically more committed to the particular school.

[27]I construct a binary indicator of "hired" teacher using the combination of teacher and principal's length of tenure in school. If a teacher entered the school in the same year or later than the principal, I code them as "hired," otherwise they are "inherited."

[28]The main effect plus interaction models are shown in Table I.A9. Except for position changes, the differences between hired and inherited teachers are statistically significant across the individual coefficients and in joint tests of the interaction terms.

school, principals see lower rates of exits among teachers that they hired.

While examining average teacher turnover rates is informative given the documented relationship between turnover and lower student performance, effective leaders may focus their retention efforts on effective teachers (Grissom and Bartanen, 2019b), for whom the costs of turnover are greater (Adnot et al., 2017). Next, I examine whether principals may become better at *strategic retention*—retaining effective teachers and "failing to retain" ineffective teachers—as they gain experience. To measure teacher effectiveness, I first estimate value-added (VA) models following the methodology of Chetty et al. (2014), then construct a categorical measure of teachers' rank in the statewide VA distribution (bottom 20%, middle 60%, top 20%).[29] In total, I can produce VA estimates for roughly 45% of teachers in the sample.

Table I.9 shows the estimated marginal effects of principal experience for turnover among low, middle, and high value-added teachers.[30] While the marginal effects are largest for high VA teachers, in general the estimates are noisy and I cannot reject the null hypothesis that they are equal across high, middle, and low VA teachers (see Table I.A10). Finally, Table I.A11 directly estimates the relationship between teacher value-added and principal experience. Perhaps unsurprisingly given the results in Table I.9, there is no apparent improvement in teacher quality as principals gain experience.

I.5 Heterogeneity in the Returns to Principal Experience by School Context

My final research question examines the extent to which principals improve more or less rapidly in certain types of schools. Specifically, I estimate whether the returns to school-

[29]The estimation steps are as follows. First, I residualize student test scores (separately by subject) on a vector of prior-year test scores, student characteristics (race/ethnicity, gender, FRPL eligibility, gifted status, special education status, lagged absences, grade repetition, and whether the student changed schools at least once during the year), school- and grade-level averages of these student characteristics, grade-by-year fixed effects, and teacher fixed effects. After computing the student residuals, I add back the teacher fixed effects and estimate the best linear predictor of a teacher's average student residuals in the current year based on their residuals from prior and future years. The coefficients from this best linear predictor are then used to predict a teacher's value-added in the current year. For teachers with value-added estimates in multiple subjects, I average these scores to construct a single measure of teacher effectiveness.

[30]Table I.A10 shows the main effects with interactions instead of marginal effects.

Table I.8: The Returns to Principal Experience for Turnover of Hired and Inherited Teachers

	All Turnover (1)	Transfer (2)	Exit (3)	Position Change (4)
School-Specific Experience				
Principal Inherited Teacher				
0 years (base)				
1 year	-0.008*	-0.007**	-0.002	0.000
	(0.004)	(0.003)	(0.003)	(0.001)
2 years	-0.014**	-0.016***	0.000	0.000
	(0.006)	(0.005)	(0.004)	(0.001)
3 years	-0.025***	-0.022***	-0.007	-0.000
	(0.009)	(0.007)	(0.006)	(0.002)
4–6 years	-0.030**	-0.025**	-0.010	0.000
	(0.013)	(0.010)	(0.008)	(0.003)
7–9 years	-0.032	-0.034**	-0.003	0.001
	(0.021)	(0.016)	(0.014)	(0.004)
10–14 years	-0.024	-0.035	0.006	0.001
	(0.031)	(0.023)	(0.022)	(0.007)
Principal Hired Teacher				
0 years (base)				
1 year	-0.023***	-0.021***	-0.010**	0.004**
	(0.006)	(0.005)	(0.005)	(0.002)
2 years	-0.030***	-0.029***	-0.010*	0.000
	(0.008)	(0.007)	(0.006)	(0.002)
3 years	-0.038***	-0.033***	-0.018***	0.003
	(0.010)	(0.008)	(0.007)	(0.002)
4–6 years	-0.055***	-0.046***	-0.025***	0.001
	(0.013)	(0.011)	(0.009)	(0.003)
7–9 years	-0.053**	-0.050***	-0.018	0.000
	(0.021)	(0.017)	(0.014)	(0.005)
10–14 years	-0.085***	-0.083***	-0.026	0.002
	(0.030)	(0.023)	(0.023)	(0.007)
N	350843	320877	318333	296924
R^2	0.078	0.089	0.070	0.026

Notes: Standard errors clustered by principal-school shown in parentheses. The dependent variable is a binary indicator for whether the teacher left their position in the current year, with the type of turnover listed in the column header. Models include fixed effects for principal, school, and year. Models include teacher and school characteristics, an indicator for whether the principal left the school at the end of the year, and an indicator for being a "hired" teacher. "Inherited" teachers are those who have more school-specific experience than the principal. "hired" teachers are those with the same or less school-specific experience than the principal.
* p < 0.1, ** p < 0.05, *** p < 0.01.

Table I.9: The Returns to Principal Experience for Turnover of Effective and Ineffective Teachers

	All Turnover (1)	Transfer (2)	Exit (3)	Position Change (4)
School-Specific Experience				
Low Value-Added Teacher				
0 years (base)				
1 year	-0.014*	-0.014**	-0.003	0.000
	(0.008)	(0.007)	(0.005)	(0.002)
2 years	-0.019*	-0.018**	-0.003	-0.001
	(0.010)	(0.009)	(0.006)	(0.002)
3 years	-0.023*	-0.027***	0.000	-0.001
	(0.012)	(0.010)	(0.008)	(0.003)
4–6 years	-0.028*	-0.027**	-0.007	-0.001
	(0.016)	(0.014)	(0.010)	(0.004)
7–9 years	-0.030	-0.036	0.004	-0.004
	(0.027)	(0.022)	(0.017)	(0.006)
10–14 years	-0.040	-0.062*	-0.001	0.013
	(0.046)	(0.034)	(0.031)	(0.014)
Middle Value-Added Teacher				
0 years (base)				
1 year	-0.002	-0.004	0.003	-0.001
	(0.005)	(0.004)	(0.003)	(0.001)
2 years	-0.007	-0.011*	0.005	-0.002
	(0.008)	(0.006)	(0.005)	(0.002)
3 years	-0.013	-0.013	-0.000	-0.002
	(0.010)	(0.009)	(0.007)	(0.003)
4–6 years	-0.024*	-0.023*	-0.002	-0.005
	(0.015)	(0.012)	(0.009)	(0.003)
7–9 years	-0.023	-0.026	0.004	-0.007
	(0.025)	(0.021)	(0.015)	(0.006)
10–14 years	-0.057*	-0.056**	-0.000	-0.014
	(0.033)	(0.028)	(0.024)	(0.009)
High Value-Added Teacher				
0 years (base)				
1 year	-0.022***	-0.018***	-0.006	0.001
	(0.007)	(0.006)	(0.004)	(0.002)
2 years	-0.021**	-0.020**	0.001	-0.005**
	(0.010)	(0.008)	(0.006)	(0.002)
3 years	-0.039***	-0.032***	-0.009	-0.005
	(0.012)	(0.010)	(0.008)	(0.003)
4–6 years	-0.043***	-0.031**	-0.011	-0.008**
	(0.016)	(0.013)	(0.010)	(0.004)
7–9 years	-0.042	-0.036	0.001	-0.014**
	(0.026)	(0.022)	(0.016)	(0.007)
10–14 years	-0.063	-0.074**	0.006	-0.006
	(0.039)	(0.031)	(0.025)	(0.011)
N	164585	154140	147697	140613
R^2	0.089	0.105	0.075	0.039

Notes: Standard errors clustered by principal-school shown in parentheses. The dependent variable is a binary indicator for whether the teacher left their position in the current year, with the type of turnover listed in the column header. Models include fixed effects for principal, school, and year. Models include teacher and school characteristics, an indicator for whether the principal left the school at the end of the year, and indicators for being high and low value-added. Value-added categories correspond to the top 20%, middle 60%, and bottom 20% of the statewide distribution.
* p < 0.1, ** p < 0.05, *** p < 0.01.

specific experience for student achievement vary across three measures of context: school poverty, school level, and school locale.

Table I.10 shows the estimated returns to experience for principals in low-poverty (0-30% FRPL), medium-poverty (30-70% FRPL), and high-poverty (70-100% FRPL) schools.[31] Across each subject, the returns to experience are largest in high-poverty schools. For instance, average math achievement improves by 0.042 SD between a principal's first and second year in a high-poverty school, compared to 0.002 SD (n.s.) and 0.009 SD (n.s.) in low- and medium-poverty schools, respectively. By 4–6 years of school-specific experience in high-poverty schools, student achievement has increased by 0.10 to 0.12 SD depending on the subject.

In Table I.A12, I show the results replacing year-by-poverty fixed effects with year fixed effects. The difference in the results between these two specifications is substantial. Whereas the returns to experience are driven by principals in high-poverty schools in my preferred models, the year fixed effects models lead to the opposite conclusion. This difference highlights the importance of allowing for flexibility in terms of identifying productivity trends that are correlated with the acquisition of experience. As an additional check, I estimate separate models by poverty group, which are shown in Table I.A13. Whereas the main models account for heterogeneity in the year fixed effects, separate models by poverty group account for heterogeneity in all of the covariates that might otherwise be conflated with heterogeneity in the returns to experience. Here, I find that the split sample results are very similar to the pooled models that include year-by-poverty fixed effects.

Tables I.A14 and I.A15 show estimates of heterogeneity by school level (elementary, middle, high) and locale (urban, suburban, town/rural), respectively. As with the poverty models, my preferred specification includes year-by-level or year-by-locale fixed effects, though I also show the results with year fixed effects. Again, accounting for heterogeneity in the year fixed effects matters, particularly for the results by school level. In my preferred mod-

[31] Table I.A12 shows the main effects and interactions that correspond to Table I.10.

Table I.10: Heterogeneity in the Returns to Experience by School Poverty (Marginal Effects)

	Math (SD) (1)	ELA (SD) (2)	Sci (SD) (3)
School-Specific Experience			
Low-Poverty School			
0 years (base)			
1 year	0.002	-0.021*	-0.008
	(0.017)	(0.012)	(0.015)
2 years	0.005	-0.014	0.014
	(0.027)	(0.022)	(0.027)
3 years	0.028	-0.028	0.019
	(0.039)	(0.032)	(0.041)
4–6 years	0.064	-0.038	0.032
	(0.053)	(0.043)	(0.055)
7–9 years	0.054	-0.112	0.023
	(0.082)	(0.075)	(0.101)
10–14 years	0.052	-0.114	0.040
	(0.110)	(0.097)	(0.130)
Medium-Poverty School			
0 years (base)			
1 year	0.009	0.003	0.002
	(0.008)	(0.005)	(0.007)
2 years	0.023*	0.006	0.011
	(0.012)	(0.008)	(0.011)
3 years	0.035**	0.012	0.016
	(0.018)	(0.011)	(0.017)
4–6 years	0.053**	0.011	0.020
	(0.024)	(0.016)	(0.024)
7–9 years	0.079**	0.032	0.020
	(0.036)	(0.024)	(0.034)
10–14 years	0.120**	0.058	0.007
	(0.055)	(0.036)	(0.051)
High-Poverty School			
0 years (base)			
1 year	0.042***	0.036***	0.032***
	(0.012)	(0.007)	(0.010)
2 years	0.058***	0.048***	0.060***
	(0.019)	(0.012)	(0.016)
3 years	0.101***	0.082***	0.102***
	(0.027)	(0.017)	(0.023)
4–6 years	0.123***	0.103***	0.122***
	(0.037)	(0.024)	(0.033)
7–9 years	0.124**	0.151***	0.184***
	(0.053)	(0.037)	(0.052)
10–14 years	0.122	0.123*	0.262***
	(0.079)	(0.064)	(0.067)
N	3034717	3328229	2819110
R^2	0.302	0.319	0.339

Notes: Standard errors clustered by principal-school shown in parentheses. In each column, I show the estimated marginal effects for school-specific principal experience (i.e., main effect plus interaction term) by school-poverty level. Models include fixed effects for principal, school, year-by-poverty group, and grade. Covariates include student characteristics and time-varying school characteristics. The model results showing the main effect and interactions are in Table I.A12. High, medium, and low-poverty refer to the percentage of students in the school who qualify for free/reduced price lunch: 0-30%, 30-70%, 70-100%. These categories are time-invariant and absorbed by the school fixed effect.
* p < 0.1, ** p < 0.05, *** p < 0.01.

els, I find no evidence that the returns to experience vary by school level or locale. In each case, joint tests of the interaction terms are not statistically significant at conventional levels, though in some cases the standard errors on the interactions terms are quite large, particularly at higher experience levels.

I.6 Discussion

Increased recognition of the pivotal role of principals in driving school performance has spurred policy attention at the local, state, and federal levels to developing of high-quality school leadership. Yet we know little about what makes some principals more effective than others, and, by extension, the extent to which differences in principal quality are driven by on-the-job improvement.

My results demonstrate that principals become substantially more effective at raising student achievement as they gain experience. These returns to experience are largest for math achievement, where the difference between a principal's first and fifth year in the principalship is roughly 25% of a standard deviation in terms of the distribution of principal quality in Tennessee. The returns to principal experience for student achievement are corroborated by ratings from supervisors. Within-principal improvement in supervisor ratings is substantial—the average principal improves their average score by 0.20 SD between their first and second years in the principalship and an additional 0.25 SD between their second and sixth years.

An important addendum to these results is that the returns to experience for student achievement are driven by school-specific experience rather than total experience. Put another way, improvement in a principal's ability to raise test scores in one school does not carry over when they change schools. There are multiple potential explanations for this finding. The first is that a principal's skill acquisition over time (with respect to skills that lead to student test score growth) constitutes school-specific rather than general learning. Principals may be adapting to the unique needs of their school or forming relationships with teachers, students, and parents which ultimately leads to improvements in student learning. Such improvements in principals' practice may not help them in a new school context.

An alternative explanation—and one that I cannot definitively rule out given my data and method—is that these results reflect the benefits of stable leadership over time. More specifically, principals may not substantially improve (in terms of concrete changes in their practice) with additional experience in the school, but student achievement may increase simply because the principal has time to have an impact. Prior work, for instance, has pointed out that new principals inherit the school conditions shaped by their predecessor, such that any effects they might have will take time to manifest in student test score gains (e.g., Grissom et al., 2015a; Coelli and Green, 2012). Without direct measures of principal skills and behaviors, it is very difficult to test which of these explanations holds. However, the improvement in supervisor ratings and the decrease in teacher turnover with experience are consistent with some amount of on-the-job learning. From a policy perspective, both explanations lead to the prescription that stable leadership is important for school performance.

These results have implications for policy and research. First, this study suggests that the increased investments to principal development and coaching are warranted, as the average principal improves quite substantially over time. To the extent that these supports also increase the likelihood that principals remain in their schools, the findings here suggest that increased stability in the principal's office is likely to benefit student learning. Unfortunately, a large number of principals in Tennessee and nationally have been in their schools for only a few years (Grissom et al., 2019; Fuller and Young, 2009), meaning that only a small percentage of schools are reaping the benefits of having a highly experienced principal. Second, in light of the finding that the returns to principal experience are largely school-specific, district administrators should consider policies that aim to reduce the shuffling of principals across schools. This is particularly relevant for larger urban districts, which have the highest rates of within-district principal transfers and more schools led by inexperienced leaders (Grissom et al., 2019). Finally, this study suggests that we need to think differently about the nature of principal effects on student outcomes. Previous work most often treats principal quality as a fixed quantity, whereas the results here demonstrate that a given principal's effectiveness varies substantially as they gain experience and change schools.

This study has some important limitations. Perhaps most importantly, estimating the returns to principal experience involves addressing several obstacles to identification using imperfect methods. As with any analysis using observational data, there is no guarantee that all confounding factors have been addressed. Of particular note in this study is that isolating the returns to experience from unobserved time-varying productivity trends requires imposing strong identification assumptions. While others have proposed approaches for navigating these obstacles in the context of teacher productivity (e.g., Papay and Kraft, 2015), my results suggest that these approaches may not be suitable for principals.

An additional issue is that the majority of principals I observe only worked in one school during the study period, and many are observed for a small number of years. While this is a reflection of the nature of the principal labor market, it has two implications for my findings. First, identification of the returns to experience are only based on principals who remain in the principalship. As principal experience increases, this subset of "stayers" becomes increasingly small, which raises the possibility that the estimates are not representative of the full sample. This is particularly concerning for school-specific experience, where my estimates could be driven by principals who are particularly well matched to their schools. That said, I do not find evidence of differences in the early-career returns to experience for principals who exit early, which suggests they would have improved at similar rates to those who stayed in the principalship longer.

The second implication is that the returns to total experience and school-specific experience are identified only from principals who changed schools. To the extent that this mobility is influenced by general or school-specific improvement, these findings may not generalize to the broader population of principals. Finally, because the data do not contain job-specific experience measures, I cannot estimate the returns to principal experience beyond the length of the panel—14 years in this case. Future work should seek to leverage increasingly available longitudinal datasets that span across many years and include job-specific measures of experience.

In sum, the findings of this study suggest that principals have substantial capacity for improvement, such that the quality of the school leadership could be increased through promoting leadership stability and professional growth. Future work should continue to focus on identifying ways to better support principals, particularly in disadvantaged or under-resourced schools. Additionally, this study raises important questions about the nature of principal improvement. What are the actual skills that principals build as they accumulate more experience as leaders? Why do some principals improve at greater rates than others? What training or preparation experiences lead to faster on-the-job learning? Answers to these questions will inform policies that can increase the quality of school leadership.

I.7 Appendix

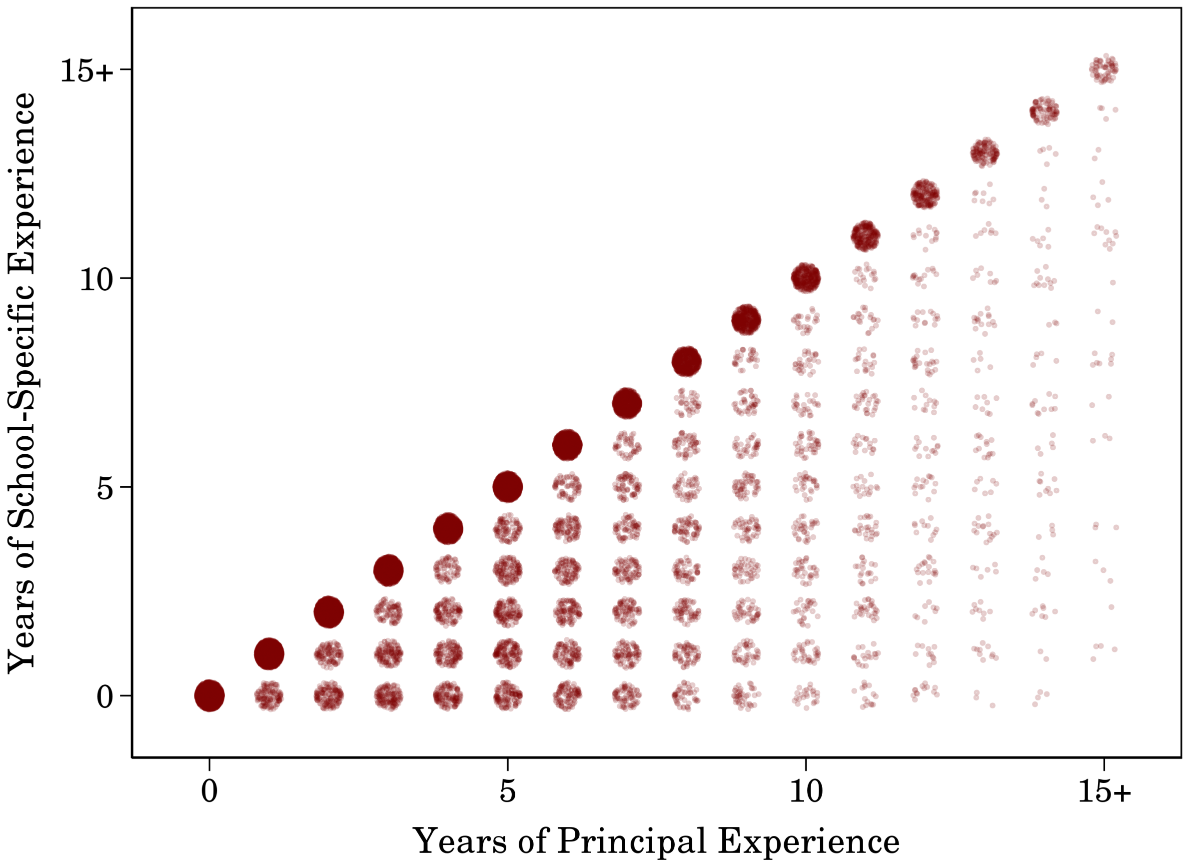


Figure I.A2: The Distribution of Total vs. School-Specific Principal Experience

Notes: Experience and tenure are discrete values from 0 to 15+. Random jitter added to show density.

Table I.A1: Proportion of Principals with Observed Experience and Tenure by Year

Year	# of Principals	Experience is Non-missing	Tenure is Non-missing
2007	1620	0.29	0.50
2008	1643	0.38	0.59
2009	1651	0.47	0.67
2010	1677	0.55	0.74
2011	1696	0.61	0.79
2012	1703	0.67	0.83
2013	1698	0.72	0.87
2014	1707	0.75	0.88
2015	1679	0.79	0.91
2016	1681	0.82	0.92
2017	1670	0.84	0.93

Table I.A2: Returns to Principal Experience from Discontinuous Career Model

	Math (1)	ELA (2)	Sci (3)	Sup. Ratings (4)
Total Principal Experience				
0 years (base)				
1 year	0.102***	0.024**	0.023	0.341
	(0.032)	(0.012)	(0.016)	(0.247)
2 years	0.192***	0.042*	0.051	0.587
	(0.064)	(0.024)	(0.032)	(0.506)
3 years	0.294***	0.067*	0.080*	0.768
	(0.096)	(0.035)	(0.048)	(0.758)
4 years	0.400***	0.088*	0.104	0.980
	(0.128)	(0.047)	(0.064)	(1.013)
5 years	0.477***	0.098*	0.116	1.147
	(0.160)	(0.059)	(0.080)	(1.257)
6 years	0.572***	0.122*	0.135	1.282
	(0.191)	(0.070)	(0.096)	(1.499)
7 years	0.658***	0.137*	0.163	1.396
	(0.224)	(0.082)	(0.112)	(1.764)
8 years	0.740***	0.160*	0.166	1.526
	(0.256)	(0.093)	(0.128)	(2.007)
9 years	0.820***	0.168	0.174	1.629
	(0.287)	(0.105)	(0.143)	(2.266)
10 years	0.942***	0.197*	0.188	1.753
	(0.320)	(0.117)	(0.161)	(2.512)
11 years	1.013***	0.203	0.234	2.004
	(0.352)	(0.129)	(0.176)	(2.769)
12 years	1.138***	0.270*	0.276	2.086
	(0.384)	(0.141)	(0.193)	(3.051)
13 years	1.265***	0.298	0.266	2.100
	(0.420)	(0.182)	(0.211)	(3.337)
14 years	1.195***	0.251	0.292	1.931
	(0.452)	(0.169)	(0.231)	(3.580)
N	3262309	3583834	3029986	6986
R^2	0.303	0.321	0.341	0.713

Notes: Standard errors clustered by principal-school (district in column 4) shown in parentheses. The dependent variables is listed in the column header. Models include fixed effects for principal, school, and year. Columns 1–3 also include fixed effects for grade. Covariates include student characteristics (columns 1–3) and time-varying school characteristics.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table I.A3: The Returns to Principal Experience Including Prior Test Scores

	Prior-year Score			Prior-school Score		
	Math (1)	ELA (2)	Sci (3)	Math (4)	ELA (5)	Sci (6)
Total Principal Experience						
0 years (base)						
1 year	0.017**	0.005	0.009	0.023***	0.003	-0.001
	(0.007)	(0.004)	(0.006)	(0.009)	(0.005)	(0.008)
2 years	0.021**	0.009	0.025***	0.030**	0.008	0.007
	(0.010)	(0.006)	(0.009)	(0.013)	(0.008)	(0.012)
3 years	0.035**	0.013	0.030**	0.051***	0.010	0.004
	(0.014)	(0.009)	(0.013)	(0.018)	(0.011)	(0.018)
4–6 years	0.047**	0.012	0.041**	0.064**	0.012	0.009
	(0.021)	(0.013)	(0.019)	(0.025)	(0.016)	(0.025)
7–9 years	0.060**	0.012	0.057**	0.075**	0.011	0.007
	(0.030)	(0.020)	(0.028)	(0.036)	(0.024)	(0.034)
10–14 years	0.098**	0.031	0.064	0.135***	0.026	-0.025
	(0.040)	(0.028)	(0.039)	(0.051)	(0.031)	(0.050)
N	2320609	2612118	2106567	1477933	1728090	1295608
R^2	0.579	0.625	0.600	0.549	0.607	0.573
Joint Test of Year FE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Notes: Standard errors clustered by principal-school shown in parentheses. The dependent variables are standardized student achievement scores. The omitted category is principals who have zero years of experience. Models include fixed effects for principal, school, year, and grade. Covariates include student characteristics and time-varying school characteristics. "Joint Test of Year FE" shows the p-value from an F-test that the estimated year fixed effects are jointly zero.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table I.A4: The Returns to Principal Experience Using Narrow Experience Bins

	Math			ELA			Sci		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Total Principal Experience									
0 years (base)									
1 year	0.021***		0.014	0.013***		0.003	0.004		-0.004
	(0.007)		(0.012)	(0.005)		(0.007)	(0.007)		(0.011)
2 years	0.031**		0.001	0.020**		-0.008	0.014		-0.015
	(0.012)		(0.019)	(0.009)		(0.012)	(0.011)		(0.015)
3 years	0.053***		0.012	0.035***		-0.000	0.026*		-0.008
	(0.018)		(0.026)	(0.012)		(0.017)	(0.016)		(0.021)
4–5 years	0.082***		0.040	0.048***		0.013	0.031		-0.010
	(0.024)		(0.032)	(0.017)		(0.021)	(0.022)		(0.028)
6–7 years	0.093***		0.019	0.064***		0.005	0.037		-0.019
	(0.032)		(0.044)	(0.023)		(0.028)	(0.030)		(0.035)
8–9 years	0.099**		0.019	0.081***		0.011	0.026		-0.035
	(0.042)		(0.055)	(0.029)		(0.035)	(0.038)		(0.044)
10–11 years	0.136***		0.037	0.096***		0.012	0.032		-0.033
	(0.052)		(0.065)	(0.037)		(0.047)	(0.051)		(0.061)
12–14 years	0.172***		0.075	0.153***		0.059	0.068		-0.041
	(0.067)		(0.088)	(0.053)		(0.073)	(0.065)		(0.074)
School-Specific Experience									
0 years (base)									
1 year		0.024***	0.013		0.017***	0.015**		0.010	0.012
		(0.007)	(0.011)		(0.005)	(0.007)		(0.007)	(0.010)
2 years		0.043***	0.043**		0.033***	0.039***		0.028**	0.039***
		(0.012)	(0.017)		(0.008)	(0.011)		(0.011)	(0.014)
3 years		0.069***	0.060**		0.052***	0.052***		0.043***	0.048**
		(0.017)	(0.025)		(0.012)	(0.016)		(0.016)	(0.020)
4–5 years		0.096***	0.066**		0.066***	0.055***		0.053**	0.059**
		(0.024)	(0.032)		(0.017)	(0.021)		(0.023)	(0.028)
6–7 years		0.129***	0.116***		0.097***	0.092***		0.070**	0.083**
		(0.032)	(0.044)		(0.023)	(0.028)		(0.031)	(0.036)
8–9 years		0.136***	0.123**		0.121***	0.111***		0.063	0.088*
		(0.043)	(0.057)		(0.029)	(0.035)		(0.041)	(0.046)
10–11 years		0.191***	0.160**		0.151***	0.139***		0.070	0.091
		(0.057)	(0.071)		(0.037)	(0.048)		(0.054)	(0.066)
12–14 years		0.175**	0.110		0.192***	0.135*		0.185**	0.212**
		(0.072)	(0.099)		(0.045)	(0.076)		(0.076)	(0.086)
N	3034743	3034743	3034743	3328312	3328312	3328312	2819174	2819174	2819174
R^2	0.301	0.301	0.301	0.319	0.319	0.319	0.339	0.339	0.339

Notes: Standard errors clustered by principal-school shown in parentheses. The dependent variables are standardized student achievement scores. The omitted category is principals who have zero years of experience. Models include fixed effects for principal, school, year, and grade. Covariates include student characteristics and time-varying school characteristics.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table I.A5: Year Fixed Effects Estimates from Achievement Models

	IVM			Cen. Growth		
	Math (1)	ELA (2)	Sci (3)	Math (4)	ELA (5)	Sci (6)
Year Fixed Effects						
2007 (base)						
2008	-0.012	-0.009	-0.020**	-0.011	-0.011	-0.017*
	(0.009)	(0.007)	(0.009)	(0.009)	(0.007)	(0.009)
2009	-0.018	-0.017*	-0.035***	-0.017	-0.020**	-0.028**
	(0.013)	(0.010)	(0.012)	(0.012)	(0.010)	(0.012)
2010	-0.088***	-0.053***	-0.072***	-0.083***	-0.055***	-0.061***
	(0.017)	(0.013)	(0.015)	(0.015)	(0.012)	(0.015)
2011	-0.122***	-0.082***	-0.109***	-0.113***	-0.083***	-0.095***
	(0.020)	(0.015)	(0.018)	(0.018)	(0.014)	(0.018)
2012	-0.108***	-0.080***	-0.104***	-0.097***	-0.078***	-0.087***
	(0.023)	(0.017)	(0.020)	(0.020)	(0.015)	(0.020)
2013	-0.104***	-0.064***	-0.095***	-0.090***	-0.062***	-0.074***
	(0.026)	(0.020)	(0.023)	(0.023)	(0.017)	(0.022)
2014	-0.107***	-0.057***	-0.091***	-0.089***	-0.053***	-0.066***
	(0.029)	(0.022)	(0.026)	(0.025)	(0.019)	(0.025)
2015	-0.097***	-0.036	-0.068**	-0.075***	-0.030	-0.041
	(0.033)	(0.024)	(0.028)	(0.028)	(0.021)	(0.027)
2016	-0.102***	-0.058**	-0.090***	-0.077**	-0.052**	-0.059*
	(0.038)	(0.029)	(0.033)	(0.030)	(0.024)	(0.030)
2017	-0.101**	-0.034	-0.078**	-0.071**	-0.027	-0.043
	(0.040)	(0.029)	(0.034)	(0.032)	(0.024)	(0.030)
Joint Test of Year FE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Notes: Standard errors clustered by principal-school shown in parentheses. These estimates correspond to the models shown in Table I.2.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table I.A6: IVM Results Using Modified Year Bins

	Math (1)	ELA (2)	Sci (3)
Total Principal Experience			
0 years (base)			
1 year	0.019***	0.006*	0.009*
	(0.005)	(0.003)	(0.005)
2 years	0.026***	0.006	0.024***
	(0.008)	(0.005)	(0.008)
3 years	0.046***	0.014**	0.041***
	(0.010)	(0.007)	(0.011)
4–6 years	0.073***	0.018*	0.054***
	(0.014)	(0.010)	(0.015)
7–9 years	0.080***	0.021	0.071***
	(0.019)	(0.014)	(0.021)
10–14 years	0.118***	0.032*	0.091***
	(0.027)	(0.019)	(0.030)
Modified Year Bins			
2008–2009	-0.017*		
	(0.010)		
2010	-0.092***		
	(0.014)		
2011	-0.128***		
	(0.016)		
2012–2013	-0.115***		
	(0.018)		
2014–2015	-0.115***		
	(0.022)		
2016–2017	-0.119***		
	(0.027)		
2008–2009		-0.014*	
		(0.007)	
2010		-0.054***	
		(0.010)	
2011–2012		-0.083***	
		(0.012)	
2013–2014		-0.065***	
		(0.014)	
2015		-0.041**	
		(0.016)	
2016		-0.065***	
		(0.021)	
2017		-0.041**	
		(0.019)	
2008			-0.020**
			(0.009)
2009			-0.035***
			(0.012)
2010			-0.070***
			(0.015)
2011–2012			-0.107***
			(0.018)
2013–2014			-0.097***
			(0.021)
2015			-0.075***
			(0.024)
2016–2017			-0.087***
			(0.028)
N	3034743	3328312	2819174
R^2	0.301	0.319	0.339

Notes: Standard errors clustered by principal-school shown in parentheses. The dependent variables are standardized student achievement scores. The omitted category is principals who have zero years of experience. Models include fixed effects for principal, school, year, and grade. Covariates include student characteristics and time-varying school characteristics.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table I.A7: The Returns to School-Specific Experience by Length of Stay in School

	Math				ELA				Sci			
Length of Spell (x) =	2	3	5	10	2	3	5	10	2	3	5	10
School-Specific Experience												
0 years (base)												
1 year	0.034***	0.026***	0.033***	0.024***	0.013*	0.008	0.015***	0.011**	0.008	0.012	0.015**	0.006
	(0.011)	(0.008)	(0.007)	(0.008)	(0.007)	(0.005)	(0.005)	(0.005)	(0.010)	(0.008)	(0.006)	(0.007)
2 years	0.042***	0.037***	0.040***	0.032**	0.017***	0.015*	0.016**	0.013	0.030***	0.030**	0.028***	0.015
	(0.010)	(0.014)	(0.012)	(0.013)	(0.006)	(0.008)	(0.007)	(0.008)	(0.009)	(0.012)	(0.010)	(0.011)
3 years	0.065***	0.060***	0.062***	0.050**	0.028***	0.029***	0.023**	0.018	0.044***	0.041***	0.048***	0.021
	(0.015)	(0.014)	(0.017)	(0.020)	(0.009)	(0.009)	(0.011)	(0.012)	(0.013)	(0.013)	(0.016)	(0.017)
4–6 years	0.097***	0.087***	0.087***	0.070***	0.032***	0.032**	0.026*	0.012	0.057***	0.052***	0.057***	0.016
	(0.021)	(0.020)	(0.022)	(0.027)	(0.012)	(0.013)	(0.014)	(0.016)	(0.019)	(0.019)	(0.020)	(0.024)
7–9 years	0.129***	0.112***	0.112***	0.081**	0.050***	0.048**	0.040**	0.017	0.073***	0.064**	0.070**	0.008
	(0.032)	(0.029)	(0.031)	(0.038)	(0.019)	(0.019)	(0.020)	(0.026)	(0.028)	(0.027)	(0.029)	(0.038)
10–14 years	0.182***	0.156***	0.158***	0.178***	0.068**	0.065**	0.058**	0.090**	0.086**	0.073*	0.078*	0.044
	(0.048)	(0.044)	(0.044)	(0.064)	(0.027)	(0.027)	(0.027)	(0.040)	(0.040)	(0.040)	(0.041)	(0.072)
Interactions												
1 year x Spell >= x	-0.008	-0.001	-0.021**	-0.011	-0.002	0.006	-0.005	0.012	0.004	0.001	-0.008	0.006
	(0.011)	(0.009)	(0.009)	(0.021)	(0.007)	(0.006)	(0.006)	(0.022)	(0.010)	(0.008)	(0.008)	(0.029)
2 years x Spell >= x		0.004	0.001	-0.007		0.004	-0.001	-0.007		-0.002	0.002	0.012
		(0.013)	(0.011)	(0.022)		(0.008)	(0.007)	(0.025)		(0.011)	(0.009)	(0.043)
3 years x Spell >= x			-0.005	-0.018			0.000	0.014			-0.001	0.030
			(0.013)	(0.023)			(0.008)	(0.028)			(0.011)	(0.045)
4–6 years x Spell >= x				0.008				0.048				0.069
				(0.030)				(0.031)				(0.051)
7–9 years x Spell >= x				0.033				0.048				0.033
				(0.036)				(0.033)				(0.061)
N	3072943	2907300	2539881	1830425	3394732	3219404	2823009	2031196	2862682	2712853	2378006	1737399
R^2	0.299	0.299	0.300	0.304	0.319	0.318	0.317	0.320	0.339	0.340	0.343	0.350

Notes: Standard errors clustered by principal-school shown in parentheses. The dependent variables are standardized student achievement scores. Models include fixed effects for principal, school, year, and grade. Covariates include student characteristics and time-varying school characteristics. The interaction group is defined by an indicator for whether a principal stays in their school for at least x years, where x is defined at the top of the column. Right-censored principal-school spells are dropped from the model if the highest observed year is less than the length of the spell. For example, if $x = 10$, I only keep school-principal spells that ended prior to 2017 (the last year of the data stream) or at least 10 years long by 2017.
* p < 0.1, ** p < 0.05, *** p < 0.01.

Table I.A8: The Returns to School-Specific Experience by Whether Principal is Observed in Multiple Schools

	Math (1)	ELA (2)	Sci (3)
School-Specific Experience			
0 years (base)			
1 year	0.021***	0.009**	0.006
	(0.007)	(0.004)	(0.006)
2 years	0.034***	0.013*	0.025***
	(0.011)	(0.007)	(0.010)
3 years	0.060***	0.028***	0.041***
	(0.015)	(0.010)	(0.014)
4–6 years	0.090***	0.035**	0.057***
	(0.021)	(0.014)	(0.020)
7–9 years	0.113***	0.055***	0.074**
	(0.031)	(0.020)	(0.029)
10–14 years	0.153***	0.069**	0.086**
	(0.047)	(0.029)	(0.041)
Interactions			
1 year x Multiple Schools	-0.000	0.000	0.011
	(0.008)	(0.005)	(0.008)
2 years x Multiple Schools	0.004	0.007	0.007
	(0.011)	(0.007)	(0.009)
3 years x Multiple Schools	-0.011	-0.006	0.001
	(0.014)	(0.009)	(0.013)
4–6 years x Multiple Schools	-0.028*	-0.017	-0.015
	(0.016)	(0.011)	(0.016)
7–9 years x Multiple Schools	-0.027	-0.041**	-0.034
	(0.028)	(0.018)	(0.029)
10–14 years x Multiple Schools	0.010	-0.016	-0.053
	(0.037)	(0.030)	(0.051)
N	3262311	3583835	3029986
R^2	0.303	0.321	0.341

Notes: Standard errors clustered by principal-school shown in parentheses. The dependent variables are standardized student achievement scores. Models include fixed effects for principal, school, year, and grade. Covariates include student characteristics and time-varying school characteristics. The interaction group is defined by an indicator for whether the principal is observed in two or more schools across the study period.
* p < 0.1, ** p < 0.05, *** p < 0.01.

Table I.A9: The Returns to Principal Experience for Hired and Inherited Teachers (Main Effect + Interactions)

	All Turnover (1)	Transfer (2)	Exit (3)	Position Change (4)
School-Specific Experience				
0 years (base)				
1 year	-0.008*	-0.007**	-0.002	0.000
	(0.004)	(0.003)	(0.003)	(0.001)
2 years	-0.014**	-0.016***	0.000	0.000
	(0.006)	(0.005)	(0.004)	(0.001)
3 years	-0.025***	-0.022***	-0.007	-0.000
	(0.009)	(0.007)	(0.006)	(0.002)
4–6 years	-0.030**	-0.025**	-0.010	0.000
	(0.013)	(0.010)	(0.008)	(0.003)
7–9 years	-0.032	-0.034**	-0.003	0.001
	(0.021)	(0.016)	(0.014)	(0.004)
10–14 years	-0.024	-0.035	0.006	0.001
	(0.031)	(0.023)	(0.022)	(0.007)
Interactions				
1 year x Prin Hired Teacher	-0.015***	-0.015***	-0.008*	0.004**
	(0.006)	(0.005)	(0.004)	(0.002)
2 years x Prin Hired Teacher	-0.016***	-0.012**	-0.010**	0.000
	(0.006)	(0.005)	(0.005)	(0.002)
3 years x Prin Hired Teacher	-0.013**	-0.011**	-0.011**	0.003
	(0.006)	(0.005)	(0.005)	(0.002)
4–6 years x Prin Hired Teacher	-0.025***	-0.021***	-0.015***	0.000
	(0.006)	(0.005)	(0.004)	(0.002)
7–9 years x Prin Hired Teacher	-0.021**	-0.017***	-0.015**	-0.001
	(0.008)	(0.006)	(0.006)	(0.002)
10–14 years x Prin Hired Teacher	-0.061***	-0.048***	-0.032	0.002
	(0.019)	(0.011)	(0.020)	(0.005)
N	350843	320877	318333	296924
R^2	0.078	0.089	0.070	0.026
Joint Test of Hired Int	0.001	0.000	0.036	0.161

Notes: Standard errors clustered by principal-school shown in parentheses. The dependent variable is a binary indicator for whether the teacher left their position in the current year, with the type of turnover listed in the column header. Models include fixed effects for principal, school, and year. Models include teacher and school characteristics, an indicator for whether the principal left the school at the end of the year, and an indicator for being a "hired" teacher. "Inherited" teachers are those who have more school-specific experience than the principal. "hired" teachers are those with the same or less school-specific experience than the principal.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table I.A10: The Returns to Principal Experience for Turnover of Effective and Ineffective Teachers (Main Effect + Interactions)

	All Turnover (1)	Transfer (2)	Exit (3)	Position Change (4)
School-Specific Experience				
0 years (base)				
1 year	-0.002	-0.004	0.003	-0.001
	(0.005)	(0.004)	(0.003)	(0.001)
2 years	-0.007	-0.011*	0.005	-0.002
	(0.008)	(0.006)	(0.005)	(0.002)
3 years	-0.013	-0.013	-0.000	-0.002
	(0.010)	(0.009)	(0.007)	(0.003)
4–6 years	-0.024*	-0.023*	-0.002	-0.005
	(0.015)	(0.012)	(0.009)	(0.003)
7–9 years	-0.023	-0.026	0.004	-0.007
	(0.025)	(0.021)	(0.015)	(0.006)
10–14 years	-0.057*	-0.056**	-0.000	-0.014
	(0.033)	(0.028)	(0.024)	(0.009)
Interactions				
1 year x Low VA Teacher	-0.011	-0.010	-0.006	0.001
	(0.007)	(0.006)	(0.005)	(0.002)
2 years x Low VA Teacher	-0.011	-0.007	-0.008	0.001
	(0.008)	(0.007)	(0.005)	(0.002)
3 years x Low VA Teacher	-0.010	-0.014*	0.000	0.002
	(0.009)	(0.007)	(0.006)	(0.002)
4–6 years x Low VA Teacher	-0.003	-0.004	-0.005	0.004*
	(0.008)	(0.007)	(0.005)	(0.002)
7–9 years x Low VA Teacher	-0.007	-0.010	-0.001	0.003
	(0.013)	(0.010)	(0.009)	(0.003)
10–14 years x Low VA Teacher	0.017	-0.005	-0.001	0.026**
	(0.030)	(0.025)	(0.022)	(0.012)
1 year x High VA Teacher	-0.019***	-0.014**	-0.009**	0.002
	(0.007)	(0.006)	(0.005)	(0.002)
2 years x High VA Teacher	-0.014*	-0.008	-0.005	-0.003
	(0.007)	(0.006)	(0.005)	(0.002)
3 years x High VA Teacher	-0.026***	-0.019***	-0.009	-0.002
	(0.008)	(0.007)	(0.005)	(0.003)
4–6 years x High VA Teacher	-0.019**	-0.008	-0.009*	-0.003
	(0.007)	(0.006)	(0.005)	(0.003)
7–9 years x High VA Teacher	-0.019*	-0.011	-0.004	-0.007*
	(0.012)	(0.010)	(0.008)	(0.004)
10–14 years x High VA Teacher	-0.006	-0.018	0.006	0.008
	(0.023)	(0.016)	(0.017)	(0.010)
N	164585	154140	147697	140613
R^2	0.089	0.105	0.075	0.039
Joint Test of Low VA Int	0.628	0.572	0.678	0.203
Joint Test of High VA Int	0.043	0.166	0.404	0.207

Notes: Standard errors clustered by principal-school shown in parentheses. The dependent variable is a binary indicator for whether the teacher left their position in the current year, with the type of turnover listed in the column header. Models include fixed effects for principal, school, and year. Models include teacher and school characteristics, an indicator for whether the principal left the school at the end of the year, and indicators for being high and low value-added. Value-added categories correspond to the top 20%, middle 60%, and bottom 20% of the statewide distribution.
* p < 0.1, ** p < 0.05, *** p < 0.01.

Table I.A11: Principal Experience and Teacher Quality

	(1)	(2)	(3)
Total Principal Experience			
0 years (base)			
1 year	-0.008		-0.017
	(0.007)		(0.015)
2 years	0.005		-0.012
	(0.011)		(0.018)
3 years	0.013		-0.002
	(0.016)		(0.024)
4–6 years	0.003		-0.034
	(0.022)		(0.030)
7–9 years	-0.013		-0.069*
	(0.031)		(0.040)
10–14 years	-0.019		-0.066
	(0.044)		(0.055)
School-Specific Experience			
0 years (base)			
1 year		-0.001	0.012
		(0.007)	(0.014)
2 years		0.016	0.025
		(0.011)	(0.018)
3 years		0.025	0.025
		(0.016)	(0.024)
4–6 years		0.030	0.055*
		(0.023)	(0.031)
7–9 years		0.034	0.087**
		(0.033)	(0.043)
10–14 years		0.009	0.056
		(0.062)	(0.075)
N	164585	164585	164585
R^2	0.186	0.186	0.186

Notes: Standard errors clustered by principal-school shown in parentheses. The dependent variable is a teacher's value-added score in the given year. Models include fixed effects for principal, school, and year. Models include teacher and school characteristics, and an indicator for whether the principal left the school at the end of the year.
* p < 0.1, ** p < 0.05, *** p < 0.01.

Table I.A12: Heterogeneity in the Returns to Experience by School Poverty (Main Effect + Interactions)

	Math (SD)		ELA (SD)		Sci (SD)	
	(1)	(2)	(3)	(4)	(5)	(6)
School-Specific Experience						
0 years (base)						
1 year	0.011	0.009	0.005	0.003	0.006	0.002
	(0.007)	(0.008)	(0.004)	(0.005)	(0.006)	(0.007)
2 years	0.026**	0.023*	0.011	0.006	0.022**	0.011
	(0.010)	(0.012)	(0.007)	(0.008)	(0.009)	(0.011)
3 years	0.041***	0.035**	0.019*	0.012	0.033**	0.016
	(0.015)	(0.018)	(0.010)	(0.011)	(0.014)	(0.017)
4–6 years	0.062***	0.053**	0.022	0.011	0.045**	0.020
	(0.020)	(0.024)	(0.014)	(0.016)	(0.020)	(0.024)
7–9 years	0.093***	0.079**	0.047**	0.032	0.057**	0.020
	(0.029)	(0.036)	(0.021)	(0.024)	(0.028)	(0.034)
10–14 years	0.140***	0.120**	0.074**	0.058	0.062	0.007
	(0.047)	(0.055)	(0.032)	(0.036)	(0.043)	(0.051)
Interactions						
1 year x Low-Poverty	0.011	-0.007	-0.002	-0.024*	-0.002	-0.010
	(0.013)	(0.018)	(0.009)	(0.013)	(0.013)	(0.017)
2 years x Low-Poverty	0.020	-0.018	0.028**	-0.021	0.018	0.003
	(0.017)	(0.030)	(0.012)	(0.023)	(0.015)	(0.029)
3 years x Low-Poverty	0.046**	-0.007	0.031*	-0.039	0.023	0.002
	(0.021)	(0.043)	(0.016)	(0.034)	(0.023)	(0.044)
4–6 years x Low-Poverty	0.095***	0.012	0.065***	-0.049	0.049**	0.012
	(0.020)	(0.058)	(0.016)	(0.045)	(0.022)	(0.060)
7–9 years x Low-Poverty	0.119**	-0.025	0.048	-0.144*	0.074	0.003
	(0.053)	(0.090)	(0.036)	(0.079)	(0.050)	(0.106)
10–14 years x Low-Poverty	0.119**	-0.067	0.073**	-0.172*	0.083	0.033
	(0.048)	(0.124)	(0.035)	(0.104)	(0.054)	(0.140)
1 year x High-Poverty	0.020*	0.033**	0.016**	0.033***	0.012	0.030**
	(0.010)	(0.014)	(0.007)	(0.009)	(0.009)	(0.012)
2 years x High-Poverty	0.006	0.035	0.003	0.041***	0.009	0.050**
	(0.014)	(0.023)	(0.009)	(0.015)	(0.012)	(0.020)
3 years x High-Poverty	0.020	0.066**	0.013	0.071***	0.024	0.086***
	(0.016)	(0.033)	(0.011)	(0.021)	(0.015)	(0.029)
4–6 years x High-Poverty	-0.007	0.070	0.003	0.092***	0.008	0.102**
	(0.019)	(0.045)	(0.014)	(0.030)	(0.019)	(0.041)
7–9 years x High-Poverty	-0.063**	0.045	-0.017	0.118***	0.031	0.164***
	(0.028)	(0.065)	(0.019)	(0.045)	(0.036)	(0.063)
10–14 years x High-Poverty	-0.171***	0.002	-0.112*	0.065	0.058	0.254***
	(0.053)	(0.098)	(0.058)	(0.074)	(0.056)	(0.086)
Year Fixed Effects	✓		✓		✓	
Year x FRPL Fixed Effects		✓		✓		✓
N	3034717	3034717	3328229	3328229	2819110	2819110
R^2	0.301	0.302	0.319	0.319	0.339	0.339
Joint Test of Low-Pov Int	0.000	0.480	0.001	0.099	0.298	0.932
Joint Test of High-Pov Int	0.001	0.007	0.012	0.002	0.426	0.075

Notes: Standard errors clustered by principal-school shown in parentheses. In each column, I show the estimated main effect for school-specific principal experience and interactions by school poverty level. Models include fixed effects for principal, school, and grade. Covariates include student characteristics and time-varying school characteristics. High, medium, and low-poverty refer to the percentage of students in the school who qualify for free/reduced price lunch: 0-30%, 30-70%, 70-100%. These categories are time-invariant and absorbed by the school fixed effect.
* p < 0.1, ** p < 0.05, *** p < 0.01.

Table I.A13: Heterogeneity in the Returns to Experience by School Poverty (Separate Models)

	Math			ELA			Sci		
	Low-Pov (1)	Med-Pov (2)	High-Pov (3)	Low-Pov (4)	Med-Pov (5)	High-Pov (6)	Low-Pov (7)	Med-Pov (8)	High-Pov (9)
School-Specific Experience									
0 years (base)									
1 year	-0.005	0.008	0.037***	-0.025**	0.002	0.031***	-0.015	0.001	0.026***
	(0.016)	(0.007)	(0.012)	(0.013)	(0.005)	(0.007)	(0.015)	(0.007)	(0.010)
2 years	-0.002	0.020*	0.048**	-0.019	0.005	0.040***	0.004	0.010	0.050***
	(0.026)	(0.012)	(0.020)	(0.022)	(0.008)	(0.012)	(0.026)	(0.011)	(0.016)
3 years	0.016	0.032*	0.089***	-0.034	0.010	0.071***	0.007	0.016	0.086***
	(0.037)	(0.017)	(0.028)	(0.031)	(0.011)	(0.017)	(0.040)	(0.017)	(0.024)
4–6 years	0.048	0.049**	0.101***	-0.047	0.009	0.086***	0.013	0.020	0.097***
	(0.051)	(0.024)	(0.037)	(0.043)	(0.016)	(0.024)	(0.055)	(0.024)	(0.034)
7–9 years	0.043	0.073**	0.095*	-0.114	0.029	0.123***	0.019	0.021	0.147***
	(0.085)	(0.035)	(0.055)	(0.076)	(0.024)	(0.037)	(0.098)	(0.034)	(0.053)
10–14 years	0.043	0.108**	0.102	-0.119	0.052	0.098	0.042	0.005	0.222***
	(0.110)	(0.054)	(0.084)	(0.101)	(0.036)	(0.063)	(0.121)	(0.052)	(0.070)
N	392955	1907131	734631	435035	2135345	757849	364507	1749871	704732
R^2	0.291	0.248	0.215	0.299	0.257	0.224	0.279	0.249	0.245

Notes: Standard errors clustered by principal-school shown in parentheses. In each column, I estimate a model for the returns to school-specific experience on the sample defined by the column header (high, medium, or low-poverty). Models include fixed effects for principal, school, year, and grade. Covariates include student characteristics and time-varying school characteristics. High, medium, and low-poverty refer to the percentage of students in the school who qualify for free/reduced price lunch: 0-30%, 30-70%, 70-100%.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table I.A14: Heterogeneity in the Returns to Experience by School Level (Main Effect + Interactions)

	Math (SD)		ELA (SD)		Sci (SD)	
	(1)	(2)	(3)	(4)	(5)	(6)
School-Specific Experience						
0 years (base)						
1 year	0.022***	0.022***	0.011**	0.010*	0.013**	0.015**
	(0.007)	(0.008)	(0.005)	(0.005)	(0.006)	(0.007)
2 years	0.049***	0.049***	0.025***	0.021**	0.041***	0.043***
	(0.011)	(0.013)	(0.008)	(0.009)	(0.010)	(0.011)
3 years	0.078***	0.076***	0.049***	0.044***	0.064***	0.066***
	(0.015)	(0.018)	(0.011)	(0.012)	(0.014)	(0.016)
4–6 years	0.104***	0.104***	0.061***	0.054***	0.074***	0.077***
	(0.021)	(0.025)	(0.015)	(0.018)	(0.020)	(0.022)
7–9 years	0.121***	0.121***	0.085***	0.072***	0.095***	0.102***
	(0.032)	(0.036)	(0.022)	(0.026)	(0.031)	(0.033)
10–14 years	0.159***	0.159***	0.108***	0.087**	0.130***	0.140***
	(0.049)	(0.054)	(0.034)	(0.039)	(0.045)	(0.047)
Interactions						
1 year x Middle School	0.003	-0.009	0.004	0.006	-0.008	-0.012
	(0.010)	(0.013)	(0.007)	(0.008)	(0.009)	(0.012)
2 years x Middle School	-0.014	-0.038*	-0.005	-0.003	-0.020*	-0.027
	(0.012)	(0.020)	(0.008)	(0.014)	(0.010)	(0.019)
3 years x Middle School	-0.018	-0.056**	-0.017	-0.015	-0.031**	-0.039
	(0.015)	(0.028)	(0.010)	(0.020)	(0.014)	(0.028)
4–6 years x Middle School	-0.010	-0.069*	-0.022*	-0.017	-0.017	-0.030
	(0.017)	(0.039)	(0.012)	(0.029)	(0.017)	(0.040)
7–9 years x Middle School	0.012	-0.083	-0.041**	-0.031	-0.036	-0.060
	(0.032)	(0.056)	(0.019)	(0.044)	(0.035)	(0.063)
10–14 years x Middle School	0.012	-0.104	-0.023	-0.005	-0.099*	-0.124
	(0.063)	(0.091)	(0.039)	(0.065)	(0.052)	(0.086)
1 year x High School	-0.021*	-0.001	-0.010	-0.008	-0.011	-0.013
	(0.012)	(0.016)	(0.007)	(0.010)	(0.012)	(0.017)
2 years x High School	-0.053***	-0.010	-0.013	-0.009	-0.033**	-0.036
	(0.016)	(0.026)	(0.009)	(0.017)	(0.014)	(0.026)
3 years x High School	-0.079***	-0.020	-0.042***	-0.034	-0.049***	-0.053
	(0.018)	(0.036)	(0.011)	(0.024)	(0.018)	(0.038)
4–6 years x High School	-0.105***	-0.013	-0.056***	-0.041	-0.056***	-0.061
	(0.022)	(0.051)	(0.012)	(0.034)	(0.020)	(0.052)
7–9 years x High School	-0.115***	0.027	-0.053***	-0.025	-0.056	-0.066
	(0.031)	(0.074)	(0.018)	(0.052)	(0.034)	(0.077)
10–14 years x High School	-0.159**	0.049	-0.080	-0.036	-0.067	-0.080
	(0.073)	(0.113)	(0.067)	(0.075)	(0.069)	(0.111)
Year Fixed Effects	✓		✓		✓	
Year x Level Fixed Effects		✓		✓		✓
N	2884580	2884580	3157297	3157297	2684642	2684642
R^2	0.299	0.300	0.316	0.316	0.338	0.338
Joint Test of Middle Sch Int	0.623	0.324	0.147	0.527	0.199	0.494
Joint Test of High Sch Int	0.000	0.584	0.000	0.177	0.112	0.800

Notes: Standard errors clustered by principal-school shown in parentheses. In each column, I show the estimated main effect for school-specific principal experience and interactions by school level. Models include fixed effects for principal, school, and grade. Covariates include student characteristics and time-varying school characteristics. School level is time-invariant and absorbed by the school fixed effect.
* p < 0.1, ** p < 0.05, *** p < 0.01.

Table I.A15: Heterogeneity in the Returns to Experience by School Locale (Main Effect + Interactions)

	Math (SD)		ELA (SD)		Sci (SD)	
	(1)	(2)	(3)	(4)	(5)	(6)
School-Specific Experience						
0 years (base)						
1 year	0.021***	0.023***	0.009**	0.009*	0.008	0.016**
	(0.007)	(0.008)	(0.004)	(0.005)	(0.007)	(0.008)
2 years	0.035***	0.038***	0.013**	0.013*	0.027***	0.040***
	(0.011)	(0.013)	(0.007)	(0.008)	(0.009)	(0.011)
3 years	0.057***	0.060***	0.029***	0.029***	0.037***	0.057***
	(0.015)	(0.018)	(0.010)	(0.011)	(0.014)	(0.017)
4–6 years	0.090***	0.090***	0.037***	0.038**	0.045**	0.077***
	(0.020)	(0.025)	(0.014)	(0.015)	(0.019)	(0.023)
7–9 years	0.119***	0.119***	0.057***	0.060**	0.044	0.099***
	(0.031)	(0.037)	(0.022)	(0.025)	(0.029)	(0.035)
10–14 years	0.161***	0.162***	0.096***	0.107***	0.089*	0.159***
	(0.050)	(0.056)	(0.033)	(0.037)	(0.048)	(0.054)
Interactions						
1 year x Urban School	-0.006	-0.007	0.001	0.003	0.006	-0.009
	(0.009)	(0.012)	(0.006)	(0.008)	(0.009)	(0.012)
2 years x Urban School	-0.012	-0.009	-0.001	-0.000	-0.001	-0.026
	(0.012)	(0.019)	(0.008)	(0.013)	(0.011)	(0.018)
3 years x Urban School	-0.014	-0.007	-0.011	-0.010	0.005	-0.034
	(0.015)	(0.026)	(0.010)	(0.018)	(0.014)	(0.026)
4–6 years x Urban School	-0.051***	-0.032	-0.021*	-0.022	0.008	-0.053
	(0.017)	(0.037)	(0.012)	(0.025)	(0.018)	(0.037)
7–9 years x Urban School	-0.056**	-0.026	-0.013	-0.021	0.067**	-0.039
	(0.027)	(0.054)	(0.017)	(0.037)	(0.034)	(0.059)
10–14 years x Urban School	-0.061	-0.014	-0.062	-0.085	0.003	-0.133*
	(0.062)	(0.083)	(0.048)	(0.058)	(0.051)	(0.080)
1 year x Suburban School	-0.004	-0.016	-0.005	-0.004	-0.008	-0.021
	(0.011)	(0.015)	(0.007)	(0.008)	(0.010)	(0.013)
2 years x Suburban School	0.007	-0.019	0.019**	0.021	-0.003	-0.028
	(0.014)	(0.022)	(0.009)	(0.014)	(0.011)	(0.019)
3 years x Suburban School	0.003	-0.031	0.009	0.008	0.012	-0.025
	(0.017)	(0.029)	(0.011)	(0.019)	(0.016)	(0.028)
4–6 years x Suburban School	0.006	-0.035	0.006	0.007	0.019	-0.034
	(0.018)	(0.037)	(0.011)	(0.023)	(0.017)	(0.037)
7–9 years x Suburban School	-0.023	-0.078	-0.007	-0.004	0.022	-0.062
	(0.030)	(0.055)	(0.018)	(0.031)	(0.031)	(0.055)
10–14 years x Suburban School	-0.041	-0.134	-0.022	-0.016	-0.034	-0.145*
	(0.052)	(0.082)	(0.030)	(0.048)	(0.046)	(0.079)
Year Fixed Effects	✓		✓		✓	
Year x Locale Fixed Effects		✓		✓		✓
N	3011760	3011760	3304432	3304432	2798667	2798667
R^2	0.300	0.301	0.318	0.318	0.338	0.338
Joint Test of Urban Sch Int	0.101	0.902	0.477	0.477	0.424	0.387
Joint Test of Suburb Sch Int	0.872	0.664	0.368	0.272	0.646	0.371

Notes: Standard errors clustered by principal-school shown in parentheses. In each column, I show the estimated main effect for school-specific principal experience and interactions by school locale (urban, suburban, town/rural). Models include fixed effects for principal, school, and grade. Covariates include student characteristics and time-varying school characteristics. School locale is time-invariant and absorbed by the school fixed effect.
* p < 0.1, ** p < 0.05, *** p < 0.01.

CHAPTER II

Principal Quality and Student Attendance

Policymakers are increasingly concerned with decreasing student absenteeism. Recent reports find that approximately fifteen percent of students are "chronically absent" each year, which is typically defined as missing ten percent or more of the instructional days in a school year (Jordan et al., 2018). Increased attention to student absences is exemplified by recent reforms in the Every Student Succeeds Act (ESSA), which requires states to include chronic absenteeism rates as part of their report cards to the federal government. Additionally, as part of the ESSA reform that requires states to include in their accountability plan at least one indicator of school quality or student success that is not involve state test scores or graduation rates (commonly referred to as the "fifth indicator"[1]), 36 states have included an indicator related to student attendance (Jordan and Miller, 2017).

Undoubtedly, the federal requirement to track and report these outcomes explains the choice by so many states to focus on attendance, whereas other potential indicators (e.g., school climate or students' socio-emotional skills) require capacity and resources to design new measures and implement them across the state. Nevertheless, a large body of evidence suggests that reducing absenteeism may be a worthy target for school improvement efforts. Student absences are linked to a wide range of important outcomes, including lower academic achievement (Goodman, 2014; Gershenson et al., 2017; Aucejo and Romano, 2016), lower graduation rates (Allensworth and Easton, 2007; Schoeneberger, 2012), increased alcohol/drug use (Hallfors et al., 2002; Henry and Thornberry, 2010), and lower likelihood of future employment (Cattan et al., 2017). Beyond the direct effect on achievement, student attendance is a common proxy measure for character skills (e.g., conscientiousness), which are valued by employers in the labor market (Heckman and Kautz, 2013).

[1] The four other indicators are reading and math proficiency, high school graduation rate, English language proficiency, and student test score growth

Increased attention to student attendance and non-cognitive skills more broadly has spurred exploration of the extent to which schools affect these outcomes. Particularly given that school and educator accountability has primarily focused on standardized achievement tests, impacts on non-test outcomes are not well-understood. Recently, a small body of research has begun to investigate the relationship between teacher quality and students' non-test outcomes, including attendance (Jackson, 2012; Gershenson, 2016; Backes and Hansen, 2018; Liu and Loeb, 2017). These studies use value-added models (VAMs)—initially developed to examine student achievement—to isolate teachers' individual contributions to students' non-test outcomes. Two important findings emerge from these studies of non-test value-added models (N-VAMs). The first is that teacher quality matters for students' non-test outcomes, and their effects on non-test outcomes tend to be comparable, if not larger, than their effects on achievement. The second finding is that, on average, teachers who excel at increasing student achievement are not the same teachers who excel at increasing non-test outcomes. Together, these findings suggest that accountability systems that do not consider non-test outcomes fail to capture the contributions of teachers who benefit students in other ways.

While prior work establishes that teachers can contribute to increasing student attendance, there is reason to believe that principals may also be important. A large literature explores how principals affect student outcomes, though these studies primarily focus on achievement. A broadly accepted conclusion of this literature is that, in contrast to teachers, principal effects on student achievement are indirect (Hallinger and Heck, 1998; Witziers et al., 2003; Grissom and Loeb, 2011). The main channels through which principals can affect achievement are influencing school-level factors that in turn affect student learning. For instance, principals are the primary human capital managers for schools; their influence over hiring and retention affects the quality of the school's teaching staff (Grissom and Bartanen, 2019b; Cohen-Vogel, 2011; Jacob, 2011; Rockoff et al., 2010). Another important avenue is a principal's effect on school climate, which itself has direct effects on learning (Kraft et al., 2016). Both human capital management and school climate likely are also important channels through which principals can impact student attendance.

Principals may influence student attendance through other channels as well. For instance, research consistently highlights student disengagement as an important driver of student absenteeism (Black et al., 2014). While engagement is undoubtedly tied to classroom instruction and school culture, principals may also exercise influence through their management of extra-curricular offerings, such as after-school programs or athletic teams. Principals also control and coordinate school policies that target struggling students (e.g., Response to Intervention), which can help (or harm) student attendance. Similarly, principals can implement data systems (or leverage existing ones) to identify and target supports to students at risk of missing school, or establish procedures for mandatory interventions when students pass a certain absence threshold.

This study helps fill a gap in the literature by connecting principal quality to student attendance outcomes. Specifically, I estimate principal effects on student attendance in a value-added (VA) framework using 11 years of statewide data from Tennessee. Despite increased emphasis on the role of school leaders for improving student attendance, no studies to date have explicitly estimated the extent to which principals matter for attendance outcomes. Prior studies of principal VA have almost exclusively focused on achievement, consistently demonstrating that principals contribute meaningfully to variation in student test scores (Coelli and Green, 2012; Branch et al., 2012; Dhuey and Smith, 2014; Grissom et al., 2015a; Chiang et al., 2016; Dhuey and Smith, 2018). These studies also raise important concerns about modeling principal effects, which helps inform the empirical approach used in this study. I employ a two-way (principal and school) fixed effects model, as well as an estimator that relaxes the assumption that principal quality is fixed over time. The large number of principals observed across more than a decade of statewide data helps provide the necessary variation to isolate the individual contributions of principals to improving student attendance. To be specific, I answer the following research questions:

1. What effect, if any, do principals have on student attendance?
2. How does the magnitude of principal effects on attendance vary by school context?
3. To what extent are estimates of principals' effects on attendance correlated with other measures of principal effectiveness?

I make several contributions to the existing literature. First, to my knowledge, this is the first study to estimate principal VA to student attendance, and one of only a handful of studies that extend VAMs to non-test outcomes. Second, I make a methodological contribution by implementing an estimator for principal VA that has not been used in prior studies. Additionally, I leverage much more data than existing studies to estimate principal effects. While prior studies are limited by short panels, small numbers of schools, and limited grade bands (e.g., grade 12 only), I draw on 11 years of data from roughly 1800 schools serving grades K–12. As highlighted by others, successfully implementing models that can separate principal effects from other confounding factors requires large administrative datasets that have only recently become available. Third, I draw on data from Tennessee's educator evaluation system to examine the extent to which effectiveness measures (e.g., rubric-based scores from central office supervisors) are predictive of principals' contributions to increasing student attendance and test scores.

II.1 Literature Review

In this section, I review four areas of literature that are relevant to my study of principal effects on students absences. First, I motivate briefly student attendance as an important outcome for policymakers and researchers. Second, I explore different mechanisms through which principals may be able to improve student attendance. Third, I review the existing work that estimates teacher effects on student attendance. These studies are important as they provide a framework through which to approach estimating principal effects on attendance. Finally, I describe the handful of existing studies that estimate principal effects on student achievement. In particular, these studies highlight the modeling challenges that are specific to estimating principal value-added.

II.1.1 Why Care About Student Attendance?

A large body of evidence has found that student attendance is linked to both academic and non-academic outcomes. However, one of the challenges with ascertaining the effects of absences is that there are few exogenous sources of variation. In terms of observable characteristics, students who miss school regularly are more likely to be lower-income, have lower baseline achievement, and have behavior problems. Truant students also likely differ in terms of factors unavailable in administrative datasets, such as family hardship, loss of residence, and academic or social disengagement. Even designs that account for unobserved student heterogeneity, such as a student fixed effects design, do not address time-varying factors, which are likely important drivers of absenteeism. As a result, most studies examining absenteeism are correlational rather than causal.

Two exceptions are studies that exploit exogenous variation (snowfall and influenza cases) to isolate the effect of increased absences (Goodman, 2014; Aucejo and Romano, 2016). While these instrumental variables designs overcome some of the endogeneity concerns with absenteeism, their interpretations as local average treatment effects may have limited generalizability if the effects of weather- or illness-induced absences differs from the effect of absences induced by other sources, such as student disengagement. On the whole, there is relatively convincing evidence that missing school lowers student achievement (Gershenson et al., 2017; Goodman, 2014; Aucejo and Romano, 2016), with suggestive evidence of negative effects on other academic and non-academic outcomes (Allensworth and Easton, 2007; Schoeneberger, 2012; Hallfors et al., 2002; Henry and Thornberry, 2010).

II.1.2 The Role of Principals in Improving Student Attendance

The antecedents of student absenteeism are numerous; a recent review categorized them into four broad themes: (1) student-specific factors, (2) family-specific factors, (3) school-specific factors, and (4) community-specific factors (Black et al., 2014). While principals clearly cannot resolve fully all of these areas, there is ample evidence based on the existing literature that schools—and by extension, principals—can take steps to increase student attendance. I highlight some of these areas below.

II.1.2.1 Student Engagement and Teacher Quality

A major driver of attendance is students' interest and motivation in school. Students who enjoy learning and feel supported and safe at school are less likely to have attendance problems. This is particularly true for middle and high school students, who have greater agency over decisions to attend school regularly. Therefore, an important avenue through which principals can affect student attendance is by improving students' attachment to school. As with most school functions, teachers are the most important input. The bulk of student time is spent in a classroom, and thus teachers play a critical role in shaping students' experiences at school. Prior work finds that students are more likely to miss class when the teacher lacks instructional effectiveness or has poor classroom management (Monk and Ibrahim, 1984). More recent work demonstrates that teachers have large effects on student self-efficacy and behavior and happiness in class (Blazar and Kraft, 2017), which likely influences their attendance habits. Other studies demonstrate substantial variation in teachers' effects on student attendance (Gershenson, 2016; Liu and Loeb, 2017), though they do not establish the mechanisms that connect teacher quality to student attendance.

Given the evidence that teachers are instrumental in promoting student attendance, a primary channel through which principals affect student attendance is human capital management. In particular, principals can act strategically in terms of personnel management to improve student outcomes (Cohen-Vogel, 2011; Grissom and Bartanen, 2019b; Grissom et al., 2014; Rockoff et al., 2010). While prior studies have typically posited that principals seek to maximize teacher quality with respect to raising student test scores, it stands to reason that they may also value teachers that improve other dimensions of student performance, including attendance.

II.1.2.2 Family and Community Engagement

While chronic absenteeism is increasing in prominence as a policy issue, the importance of student attendance is not always communicated clearly to parents. Relatedly, parents of highly truant students often believe that their child's attendance records are average or above average compared to the child's peers (Rogers et al., 2017; Rogers and Feller, 2018) In that vein, studies have found that providing parents with information about their child's attendance or the importance of attendance more broadly can help improve school attendance rates (e.g., Epstein and Sheldon, 2002; Roderick et al., 1997; Rogers et al., 2017). In a randomized controlled trial in Philadelpha, Rogers and Feller (2018) found that providing a postcard encouraging parents to improve their child's attendance was effective in reducing absenteeism by 10% or more. Roderick et al. (1997) also find that the nature of the communication matters; one-way or punitive interactions may not be beneficial, particularly if parents feel that they are only contacted after attendance issues have become a major problem. Even communication with parents that does not explictly focus on attendance may be beneficial. For instance, Kraft and Rogers (2015) found that in a high school credit recovery program, a randomly-assigned intervention that delivered weekly individualized text messages to parents about their child's schoolwork decreased the probability of class absence by 2.5 percentage points. Similarly, Bergman (2015) found experimental evidence that providing parents with biweekly information about their child's missed assignment and grades resulted in 1.4 fewer missed classes during the semester.

Relatedly, prior work on student absenteeism suggests that promoting family and community involvement can improve student attendance (Epstein and Sheldon, 2002; Sheldon and Epstein, 2004). For instance, Sheldon and Epstein (2004) found that communication with families about attendance, celebrating good attendance with students and families, and providing community mentors to chronically absent students were all associated with decreased chronic absenteeism from one school year to the next. Epstein and Sheldon (2002) found that workshops for parents related to student attendance were associated with decreases in chronic absenteeism, even though school staff often rated these activities as relatively ineffective.

A consistent theme across different strands of research examining student attendance is that efforts to reduce absenteeism are most effective when schools are unified in their approach. While teachers and other staff members are critical to creating safe and engaging classrooms and building individual relationships with students, successful coordination at the school level requires effective leadership from the principal. Principals are responsible for making structural and programmatic decisions that ultimately facilitate the interactions between students, families, and schools that help improve attendance rates.

II.1.3 Teacher Quality and Student Attendance

While no studies have examined principal effects on student attendance, a handful of studies consider teachers' effects. Similar to how the broader teacher effects literature informs the estimation of principal effects, prior studies of teacher effects on attendance provide important insights for estimating principal effects.

Gershenson (2016) estimates teacher VA to attendance using five years of statewide data from North Carolina. Specifically, the estimates a model with teacher and school-by-year FE, which identifies teacher effects by comparing teachers who were in the same school during the same academic year. The dependent variable is annual student absences, which are standardized by grade and year. Gershenson notes that operationalizing absences in alternative ways (e.g., unstandardized, natural logs, indicators for chronically absent) does not produce meaningfully different estimates of teacher VA. He finds strong evidence that teachers have causal effects on student absences—the estimated standard deviation of the teacher effects is 0.07 s.d. (in terms of student-level absence s.d.). Relevant for the current study, he finds that conditional on controlling for prior-year absences, controlling for prior-year achievement in math and reading produces essentially identical estimates of teacher VA. Finally, Gershenson also examines correlations between teachers' attendance VA and achievement VA. Similar to Jackson (2016), he finds correlations that are close to zero, meaning that teachers who excel at increasing student test score growth are not necessarily those who improve student attendance.

Backes and Hansen (2018) use data from Miami-Dade County Public Schools to estimate teacher VA to five non-test student outcomes: absences, suspensions, GPA, classes failed, and grade repetition. While their ostensible primary research question is whether Teach For America (TFA) corps members are more effective at improving these outcomes than observably-similar non-TFA teachers, they also focus heavily on validating non-test VAMs (N-VAMs) for teachers. Where other studies that estimate N-VAMs use a straightforward fixed effects approach, Backes and Hansen employ the model proposed by Chetty et al. (2014) that accounts for drift in teacher quality (instead of assuming it is fixed over time). The authors find that the standard deviations of teacher effects on these non-test outcomes are often similar in magnitude—and sometimes larger—to those for achievement outcomes. For instance, in middle school, they find standard deviations of 0.16 and 0.15 for unexcused absences and suspension absences, respectively, compared to 0.12 for math and 0.08 for ELA. Teacher effects on unexcused absences are slightly larger in high schools (0.17 s.d.) and substantially smaller in elementary schools (0.11 s.d.). To examine the forecast unbiasedness of the N-VAMs, the authors implement the Chetty et al. (2014) teacher switching quasi-experiment. Although their estimates are imprecise due to the relatively small sample of teachers, they find strongest evidence of forecast unbiasedness for elementary and middle schools. Their high school estimates suggest that N-VAMs are not unbiased estimates of teachers' effects in this context. Additionally, they find that their VA estimates for unexcused absences in elementary school do not pass validity tests, which could raise questions about identifying teacher effects on attendance outcomes.

Ladd and Sorensen (2017) use data from North Carolina to estimate the returns to teacher experience for achievement and non-achievement student outcomes. These analyses are slightly different in that they do not directly estimate teacher value-added; rather, they exploit within-teacher variation in student outcomes to identify the the returns to experience. A necessary condition of finding returns to experience, then, is that teachers have causal effects on these outcomes. Their basic approach is to estimate models with teacher and student or school FE, where the teacher FE are treated as nuisance parameters. The authors control for covariates commonly included in value-added models, such as lagged test scores, student demographics, and classroom/school demographics. As Ladd and Sorensen note, non-test outcomes often have abnormal distributions, such that Poisson or negative binomial regression models should be preferred to linear regression. However, given the necessity to include high-dimensional fixed effects, use of these alternative models is infeasible. Instead, they construct binary outcomes and estimate linear probability models. For instance, their preferred attendance outcome is an indicator for whether a student has an absence rate above the 75th percentile. The authors find economically meaningful returns to teacher experience for both achievement and absence outcomes. Compared to a novice teacher, a teacher with 21+ years of experience reduces the proportion of high student absenteeism by 14.5 percentage points. This result is consistent with the conclusion that teachers have meaningful effects on attendance outcomes.

II.1.4 Estimating Principal Effects

Compared to the robust literature on teacher quality and estimating teacher value-added, there are few studies that investigate principal value-added. To be specific, I distinguish between principal value-added and a much larger body of work that examines correlational relationships (most often correlational) between principals and student test scores. Within the principal effects literature, there are two general types of studies. The first group of studies aim to estimate directly the variance of principal effects by exploiting year-to-year changes in achievement between successive cohorts of students (Branch et al., 2012; Coelli and Green, 2012). The intuition of this approach is that to the extent that principal quality matters for student outcomes, year-to-year changes in achievement will be greater in magnitude in cases where there was a principal transition between year t and year $t+1$. The second group of studies produce estimates of effectiveness for individual principals using value-added models (Grissom et al., 2015a; Dhuey and Smith, 2014, 2018; Chiang et al., 2016).

Branch et al. (2012) use a large sample of principals from Texas to estimate the variance of principal effects on student achievement. Specifically, they extend a methodology proposed by Rivkin et al. (2005) that leverages teacher transitions to estimate the variance of teacher effects. They find that the lower bound of the variance of principal effects is 0.05 student-level s.d. The authors note that while this direct estimation tends to produce variance estimates substantially smaller than the variance of principal value-added estimates, even 0.05 s.d. constitutes a meanginful impact given that principals affect all of the students in a school.

Coelli and Green (2012) use data from British Columbia to estimate the variance of principal effects on high school graduation rates and 12th grade exam scores. A strength of the study is that the B.C. education system explictly sought to rotate principals through different schools, which helps to provide the critical variation needed to separate principal and school effects. In addition to estimating a model that assumes principal effects are immediate and constant (i.e., principals do not vary in their effectivness over time and their effects manifest as soon as they enter the school), they allow for the possibility that principal effects are increasing as a function of tenure in the school. Put simply, a new principal may not have much capacity to affect student outcomes in their first years in the school, which would imply that the variance of principal effects would be close to zero for those years. As principals gain tenure, they have more influence over school factors that affect student outcomes (e.g., hiring and retention of high-quality teachers, creation of a positive school climate), and thus the impact of a high-quality principal is larger in magnitude.

Coelli and Green (2012) estimate that, using a model that treats principal effects as constant, a 1 s.d. increase in principal quality increases graduation rates by 1.8 percentage points (compared to a baseline mean of 82%). When accounting for the dynamic nature of principal effects, they find that principals become more consequential—a 1 s.d. increase in principal quality raises graduation rates by 2.6 percentage points. Due to the small sample of principals, however, their variance estimates are imprecise. The authors also note that, because the average observed tenure of principals in their data is three years, estimators that treat principal effects as constant will understate the importance of principals who remain in schools for a long period of time. They also find that, in terms of the variance attributable to principals, effects on English exam score are much larger than those for graduation rates. According to their calculations, if all students were in schools with principals in their sixth year at that school, differences in principal quality would explain roughly half of the student-level variation in English scores (compared to only 8% of the variation in graduation rates). This figure is almost certainly overstated, given repeated findings that teacher quality accounts for less than 20% of the student-level variation in achievement (e.g., Rivkin et al., 2005; Chetty et al., 2014).

Dhuey and Smith (2014) also use data from British Columbia to estimate principal value-added. Specifically, they examine student test score gains between grade 4 and grade 7, rather than gains between adjacent years. Due to data limitations (i.e., missing test scores, students changing schools), they drop more than half of their student sample. Additionally, because 5th and 6th grade achievement scores are unavailable, they must account for the effect of principals in these intermediary grades. Their approach is to define a modified school fixed effect, which is a unique combination of school, grade 5 principal, and grade 6 principal. Their value-added model includes both principal and the modified school fixed effects. While Dhuey and Smith (2014) do not report network sizes, the high dimensionality of this modified school effect is concerning. They find that variation in principal quality has substantial effects on students' achievement. In their two-way FE model, a principal 1 s.d.

above the mean increases math achievement by 0.41 s.d. and reading achievement by 0.30 s.d.

Grissom et al. (2015a) use data from Miami-Dade schools to explore different approaches to estimating principal effects. Specifically, they outline three general approaches: school effectiveness, relative within-school effectiveness, and school improvement. School effectiveness is essentially a school value-added model, and does not attempt to separate the contribution of the principal from the contribution of the school. Relative within-school effectiveness is similar to the approaches pursued by other studies that use both principal and school fixed effects. Their third approach, school improvement, attempts to address the possibility that a new principal may not be able to have an immediate and constant effect on student achievement. If, for instance, principals affect student learning through hiring and retention of effective teachers, we might expect that it takes several years for a new principal to affect student outcomes through this channel. The school improvement approach models principal value-added as the coefficient on a principal-specific linear time trend. The model also contains a principal-specific intercept, such that a principal's effectiveness is the average change in student growth relative to this intercept (i.e., average growth prior to the principal entering the school). In addition to examining the distribution of these VA estimates, Grissom et al. (2015a) also show their correlations with external evaluation measures, such as school accountability grades, principals' evaluation ratings, and school climate measures.

The authors find wide variation in the distribution of principal VA depending on model choice. For math, they find that the standard deviation of principal VA ranges from 0.18 (school effectiveness) to 0.06 (relative within-school effectiveness) or 0.05 (school improvement). In general, VA estimates from school FE versus school and principal FE have a fairly large positive correlation, while the principal-by-school time trends (school improvement) were uncorrelated with all other VA estimates. School effectiveness estimates were consistently predictive of external evaluation measures, even after controlling for school characteristics, and relative within-school effectiveness estimates were also predictive to a lesser extent. By contrast, school improvement estimates were uncorrelated with all of the external evaluation measures. The authors conclude that attention to model choice is important, particularly with respect to using principal VA estimates for evaluation purposes. They also note that the most conceptually appealing approaches (i.e., school improvement and relative within-school effectiveness) are the least predictive of external evaluation measures.

Chiang et al. (2016) evaluate the extent to which school VA (i.e., a school FE from a model that does not separate principal effects from school effects) is predictive of principal VA (i.e., principal FE from a model that also includes school FE). To be specific, the authors compare school and principal VA estimates from "close-to-independent" samples of students. Principal VA estimates are estimated separately by grade and come from a five-year period, while school VA estimates (also estimated separately by grade) come from a year not used to generate principal VA (the year prior to the first year of the principal VA sample). By comparing principal and school VA within each grade and using different years, only a small number of students who repeated grades will contribute to both VA estimates. Chiang et al. (2016) argue that this approach is advantageous because it prevents finding a mechanical relationship between principal and school VA that reflects common transitory shocks (i.e., components of principal or school effectiveness that are not persistent across years) and common measurement error. The authors find that while principals appear to have substantial effects on student achievement (adjusted standard deviations of 0.14 and 0.11 in math and reading, respectively), school VA is a very poor predictor of principal VA.

Dhuey and Smith (2018) use 12 years of statewide data from North Carolina to estimate principal VA. Specifically, they estimate VA using both school and principal FE, finding that the standard deviation of principal effects is 0.17 for math and 0.12 for reading. They also find that principal education is a weak predictor of value added. The primary contribution of the study is that is utilizes a larger dataset than previous studies, which should improve the quality of the VA estimates.

II.2 Data

This study analyzes longitudinal administrative data from Tennessee covering the 2006–07 through 2016–17 school years, provided by the Tennessee Department of Education (TDOE) via the Tennessee Education Research Alliance at Vanderbilt University. The Tennessee data contain detailed information about students' enrollment and attendance, including the dates of enrollment and withdrawal at each school, the dates of absences from school, and a flag for whether the absence was *excused* or *unexcused*. Additionally, I can access student demographics, including race/ethnicity, gender, free/reduced-price lunch eligibility, gifted status, and special education status. Finally, I can access a student's full test score history, which includes end-of-year achievement scores in math, reading, and science for grades 3–8, and end-of-course (EOC) exams for high school students.[2]

In addition to the student files, I access staff files that contain detailed job history and demographic information. Importantly, these files allow me to identify school principals in each year. I also draw on demographic characteristics include gender, race, and educational attainment. In total, I observe 3,877 principals working in 1,700 schools from 2006–07 through 2016–2017.

II.2.1 Operationalizing Student Attendance

As mentioned above, the student files contain individual entries for every recorded student absence. In Tennessee, a student who misses more than 50% of the school day is recorded as absent. Entries also indicate whether the absence was excused or unexcused—the criteria for these designations are determined at the district level. Figure II.1 shows the distribution of total absences by grade level. Roughly one-third of Tennessee students have 10 or more absences each year. Thirteen percent have 18 or more absences, which is the typical threshold for chronic absenteeism. However, there is a substantial increase in chronic absenteeism after students enter high school. For example, 24 percent of 12th grade students are chronically

[2] EOC exam requirements vary by year. In 2016–17, students took exams for Algebra I, Geometry, Algebra II, English I, English II, English III, Chemistry, and Biology.

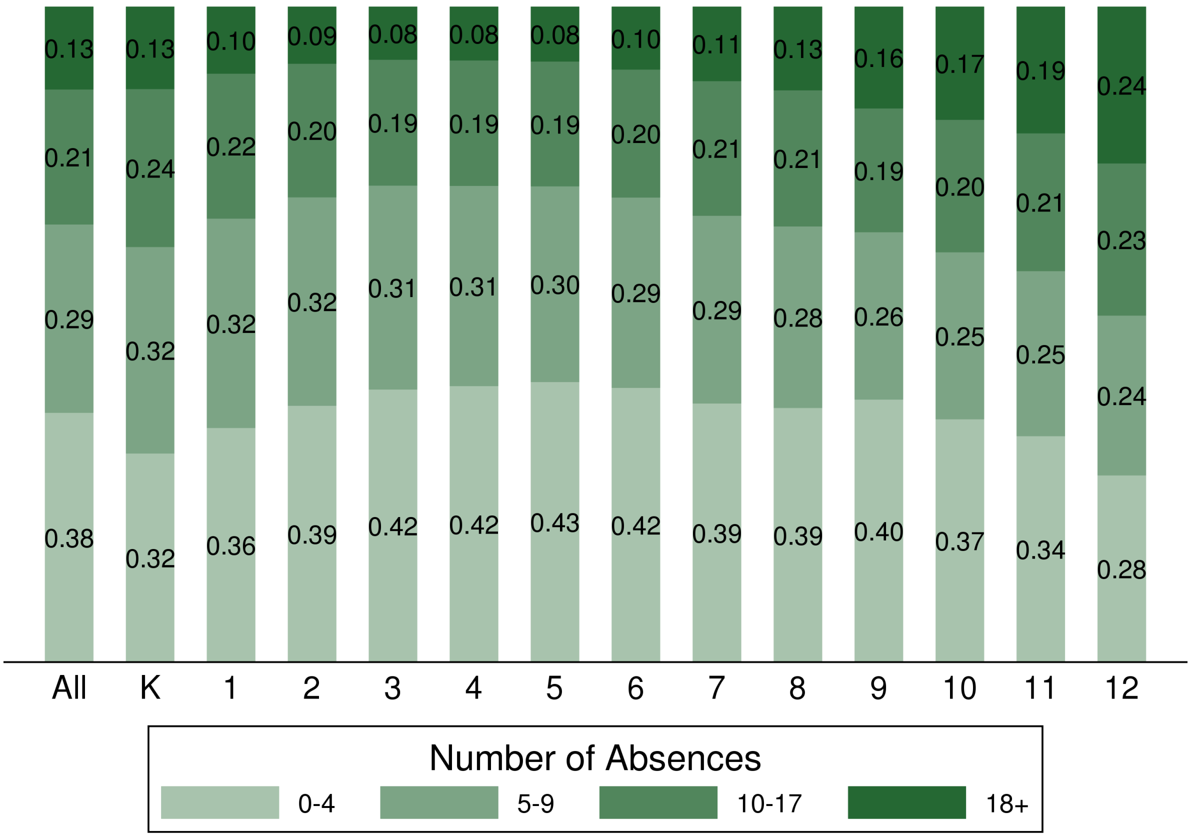


Figure II.1: Student Absenteeism by Grade

absent, compared to only 13 percent of 8th grade students and 16 percent of 9th grade students. Kindergarten students also have high absenteeism rates relative to their older elementary school peers.

I construct four outcome measures of student attendance. First, I compute each student's total absence rate, which is the number of absences divided by the total number of school days for which the student was enrolled in the school. Additionally, I compute excused and unexcused absence rates in the same fashion. While prior studies of attendance VA for teachers have modeled the discrete number of absences, I utilize rates to implicitly account for differing school year lengths across districts and for students who are not enrolled at a single school for the entire year.[3] To make the principal effect estimates for absences comparable to those for achievement, I standardize each absence rate within grade and year.[4] Finally, I also create an indicator for chronic absenteeism, which simply takes a value of 1 if a student's absence rate exceeds 10%, which is the standard threshold in prior work (Jordan et al., 2018).

II.3 Methods

II.3.1 Estimating Principal Value-Added

There are two main methods to estimate principal effects. The first is to estimate directly the variance of principal effects, which indicates how much principals matter in terms of affecting student outcomes, but does not produce effectiveness estimates for individual principals. The second method, and the one I follow in this study, is to use value-added modeling to estimate each principal's effect on student outcomes. Grissom et al. (2015a) provide a useful framework for modeling principal effects, which relates student outcomes to principal performance:

$$A_{ijs} = f(X_{ijs}, S(P_{js}, O_s)) \tag{II.1}$$

where a student's attendance or achievement is a function of their own characteristics (including family inputs), X, and the effectiveness of the school, S. School effectiveness has two aspects: the performance of the principal, P, and the effect of other school factors, O, that the principal cannot control. The primary goal of this analysis is to generate unbiased and precise estimates of P, which depends critically on the relationship between P and O. Below, I discuss four specific considerations.

[3] Students who do not remain in one school/classroom for the entire year are often dropped from VA analyses, because it is unclear how to separate the effects of multiple teachers/schools on end-of-year exams. However, with absences, there is no such problem, since absence rates are specific to the period of time that a student is enrolled in the school. Additionally, dropping high-mobility students constitutes an important external validity threat, given that this sub-population of students tend to be the most likely to be chronically absent.

[4] Alternative ways of operationalizing absences (e.g., logarithmic transformations, using the raw count data) are very highly correlated with my preferred principal effect estimates.

II.3.1.1 What is the correlation between P and O?

A fundamental issue with any value-added approach is accounting for non-random assignment. For instance, a major challenge in the teacher effects literature is how to account for the fact that some teachers are consistently assigned above-average or below-average students. If VAMs fail to fully account for this sorting, they will produce biased estimates—some teachers will receive have consistently lower VA than their true effectiveness (and vice-versa). In the case of principals, the concern is that some principals consistently work in "better" or "worse" schools (higher or lower O), such that we might conflate their effectiveness with the effect of school factors they cannot control. However, this sorting only becomes problematic if there is positive or negative sorting—if there is a correlation between P and O. Put more simply, if more effective principals (higher P) systematically work in more or less effective schools (higher or lower O), we must account for this non-random sorting to obtain unbiased estimates of P.

II.3.1.2 Are P and O fixed parameters, or do they vary over time?

VAMs often leverage longitudinal data to employ fixed effects approaches that treat effectiveness as fixed over time. In estimating principal effects, assuming that P and O are fixed allows for a more simplified modeling approach. In reality, these parameters may drift from year to year. For instance, principals may improve as they gain experience, or school effectiveness may decline due to deteriorating facilities or increasing neighborhood violence. While a more flexible model that allows for dynamic effectiveness may be conceptually appealing, it can come at the cost of imprecise estimates, with no guarantee that the added flexibility decreases bias.

II.3.1.3 Does the importance of P relative to O change over time?

A more specific concern raised in prior studies of principal effects is that a principal's ability to impact student outcomes is a function of how long they have been in the school (Coelli and Green, 2012; Grissom et al., 2015a). For instance, if principals primarily affect student outcomes through strategic hiring and retention of teachers, a new-to-school principal may have little opportunity to shape outcomes, as they largely inherit the teaching staff from the prior principal. Instead, the principal's influence grows with each additional year they remain in the school.

II.3.1.4 To what extent can O be captured by controlling for school characteristics typically present in longitudinal administrative datasets?

As highlighted above, a primary challenge of estimating principal VA is to successfully isolate P from O. In an ideal world, one could directly observe the various inputs that comprise O and control for them when estimating P. To that end, VAMs typically include school-level controls, such as averages of student demographics. However, to the extent that these controls fail to fully capture O, more sophisticated approaches are required to isolate P. Prior studies have almost uniformly addressed this problem by estimating models that include school fixed effects. This approach is sufficient insofar as the unobserved factors in O are fixed over time. Models that include school fixed effects will be biased if there are time-varying unobserved school factors that are correlated with P. For instance, school fixed effects would be insufficient in eliminating bias if gentrifying neighborhoods (O increasing over time) systematically replace less effective principals with more effective principals, and these school-level changes are not captured by characteristics available in administrative datasets.

II.3.2 Approach 1: Principal and School Fixed Effects

My first approach directly addresses the problem of nonrandom sorting of principals to schools by explicitly modeling both principal and school effects. This approach has become standard in the principal effects literature. Specifically, I estimate principal effects via ordinary least squares using the following model:

$$Y_{isjt} = \alpha Y_{i,t-1} + \beta X_{it} + \gamma S_{st} + \tau_t + \pi_g + \delta_j + \theta_s + \epsilon_{isjt} \tag{II.2}$$

where i, s, j, and t index students, schools, principals, and years, respectively. Y, a student's standardized attendance rate or achievement score, is a linear function of their prior-year outcomes; a vector of student characteristics, X; a vector of school characteristics, S; year fixed effects, τ; grade fixed effects, π; principal fixed effects, δ; school fixed effects, θ; and a random error term, ϵ. For attendance, the prior-year outcomes include cubic functions of a student's prior-year attendance rates (excused and unexcused). Not controlling for prior-year test scores allows me to use students from all grades, rather than the subset who were in grades 3–8 in the prior year or who took an EOC exam. Including these additional grades increases the network sizes in the two-way FE model and also allows me to produce estimates for roughly 100 additional principals.[5] For achievement models, I control for prior test scores in math, reading, and science, as well as prior-year attendance rates.[6] Student characteristics include sex, race/ethnicity, free/reduced-price lunch qualification, special education classification, gifted classification, whether the student is repeating the grade, and whether the student has any enrollment spells in another school in the current school year. School characteristics are school-level averages of the student characteristics.

The principal fixed effect, δ_j, is the parameter of interest; it captures the extent to which the actual attendance rates of students of principal j are higher or lower than what would be predicted by students' prior attendance and achievement, their individual characteristics, their grade, the school year, and the characteristics of their school. The model accounts for school quality in two ways: the time-varying averages of student demographics (described above) and a school fixed effect, θ_s.

It is important to note that because a majority of students in year t had the same

[5]Principal effects from attendance models that control for prior test scores are consistently correlated above 0.90 with those from the main specification.

[6]Tennessee students in grades 3–8 take yearly achievement tests in math, reading, and science. High school students take end-of-course exams in Algebra I and II, English I, II, and III, Biology, and Chemistry. For students in grades 3–8, prior scores are from the prior-year tests. For high school students, prior scores are, for each subject, the student's most recent score. For Algebra I, English I, and Biology exams, the prior score is the student's 8th grade score in math, reading/ELA, or science. For Algebra II, English II or III, and Chemistry, the prior score is the most recent EOC exam in the same subject (Algebra I is the prior score for Algebra II). If a previous EOC score is unavailable, the 8th grade score is used.

principal in year $t-1$, including prior-year outcomes (attendance or achievement) is violation of strict exogeneity and potentially biases the principal effect estimates. The degree of this bias likely depends on the nature of how principals affect student outcomes. Controlling for the prior-year outcome is effectively investigating whether a principal causes *continued* improvement in the outcome, as opposed to a one-time increase. For student achievement, examining continued improvement may be perfectly reasonable. In the case of attendance, however, the inclusion of prior-year outcomes may be more problematic. If, for instance, a new principal enters a school and implements a new program that reduces absenteeism, we might expect to observe a bump in attendance in that year. If attendance rates are stable in the next year, we would like to still conclude that the principal was effective at reducing absenteeism. However, if we condition on prior-year attendance, we will conclude that the principal had no effect in year 2.

One potential solution to this issue is to remove the prior-year outcome from the model entirely. However, this heightens the risk of bias from nonrandom student sorting between schools, particularly if students are responding to changes in school quality that are not captured by school fixed effects or demographic controls. I propose a different solution, which is to include in the model the student's most recent prior outcome from their prior school. For example, consider an 8th grade student who has been in her current school since 6th grade but attended a different school in 5th grade. Instead of controlling for her 7th grade attendance, which is endogenous to both the current principal and school effect (assuming that the principal is not in her first year at the school), I control for her 5th grade attendance, which is not affected by the current school or principal. While this approach allows me to avoid the endogeneity concern, it does come at the cost of excluding a non-trivial number of students. Specifically, I can only estimate this model for students that have prior-year outcomes in a different school, which systematically excludes students in lower grades and students in the earlier years of the data.

Another important note is that models with multiple sets of high dimensional fixed effects require sufficient mobility between units to identify each unit's effect. In the case of estimating teacher value-added with a long panel, having many teachers per school combined with relatively frequent transfer rates makes this concern somewhat trivial (e.g., Mansfield, 2015). Because the typical school has only one principal per year, however, separating principal and school effects is more difficult.

Consider the case of school X, which has a single principal across the data stream. Our goal is to ascertain that principal's effectiveness—in other words, we want to know the extent to which School X's performance is higher or lower than what it would have been under a principal of average quality. With only a single principal, we have no point of reference; we cannot determine whether the school's performance is a result of the principal or some other unobserved factor. Now imagine that school X has two principals across the data stream: principal A and principal B. Our goal is the same—we want to know principal A's (or B's) effectiveness relative to a principal of average quality. In this case, we can compare principal A and principal B to produce a local measure of effectiveness. Note that this provides only partial information with respect to our original goal. We can determine that principal A is more effective than principal B (or vice-versa), but we still do not know principal A's effectiveness relative to the average principal.

Now suppose that principal A, after leaving school X, moves to school Y, and that school Y was previously run by principal C. This situation allows us to obtain two estimates of principal A's effectiveness—one estimate relative to principal B and one relative to principal C. If we make the assumption that principal A's true effectiveness is the same in school X and school Y, then we can also compare principal B to principal C via transitivity. This example illustrates the basic mechanics of the two-way fixed effects model. By exploiting principal movement, we can construct connected networks of principals and schools. Ideally, we would want to produce a singular network that connects all principals. A singular network allows us to easily compare a given principal's effectiveness to the average principal.

In reality, patterns of principal sorting and high rates of attrition from the principalship lead to a large number of disconnected networks that often contain only a few principals. Small networks present two challenges for value-added estimation. First, as explained in the simplified example above, principals can only be compared to other principals in the same network. With a many small networks, most principals cannot be directly compared. If the average difference in true effectiveness among principals in the same network is smaller than among the full sample of principals (i.e., if principals of similar quality tend to end up in the same network of schools), principal VA estimates will be biased. Consider a group of highly effective principals in a single school district. Patterns of principal turnover make these principals more likely to be in the same network than principals from another district.[7] In the two-way FE model, these principals' VA estimates are based of off their performance relative to one another, rather than the full sample of principals. By construction, their average VA is zero, even though the average of their true effectiveness is positive. Second, disconnected networks may lead to an underestimation of the variance of principal effects. Again, if principals tend to be in networks with principals of similar quality, the average deviation in quality between two networked principals will be smaller than the average deviation of two principals chosen at random. In this case, the tendency of schools to consistently have high- or low-quality principals will be captured by the school fixed effect, and principal quality will appear less variable.

Table II.1 shows the distribution of networks formed by Tennessee principals based on the analytic sample for estimating attendance VA (Appendix Table II.A1 shows the networks for math VA). The table categorizes networks by size, according to the number of schools in the network. The second column shows the number of networks of a given size. The final three columns show the mean number of principals, the total number of principals, and the percentage of principals in a network of each size. Similar to prior studies, the modal principal is in a single-school network, meaning that their individual effect is determined by their performance relative to their predecessors and/or successors. On average, principals in

[7] Grissom and Bartanen (2019a), for instance, show that across-district principal transfers are rare in Tennessee.

Table II.1: Distribution of Principal Networks from Two-Way Fixed Effects Model

Network Size (# of Schools)	# of Networks	Mean # of Principals	Total Principals	% of Principals
1	560	2.4	1335	39.8
2	133	3.9	525	15.7
3	50	5.6	282	8.4
4	24	8.2	197	5.9
5	11	9.4	103	3.1
6	4	12.0	48	1.4
7	7	15.0	105	3.1
8	4	18.5	74	2.2
9	4	18.8	75	2.2
10	1	21.0	21	0.6
11	1	25.0	25	0.7
12	1	15.0	15	0.4
15	1	27.0	27	0.8
41	1	80.0	80	2.4
46	1	91.0	91	2.7
74	1	154.0	154	4.6
98	1	196.0	196	5.8
All	805	4.2	3353	100.0

Notes: Networks refer to the mobility groups of principals and schools from a two-way FE model for attendance outcomes.

a single-school network are compared to 1.4 other principals, and they comprise 40 percent of the total number of principals in the analytic sample. However, there are a fair number of principals who are in large networks. For example, there are four networks that have more than 50 principals each, which account for roughly 15 percent of principals. These networks are formed by schools in large urban school districts, which tend to have higher rates of intra-district transfer among principals. Overall, the average network has 4.2 principals, and the average principal is in a network with 22 other principals.

To summarize, the major advantage of the two-way FE model is that, in theory, it does not conflate the effects of school and principals. The disadvantages of this approach are that it requires sufficient mobility of principals across schools to identify the effects. Even using a 10-year statewide panel, I find that the the majority of principals' value-added estimates are identified from relative comparisons with only a handful of other principals. Additionally, I cannot produce estimates for a handful of principals who were the only principal in their school across the data stream. The implication is that δ_j can only be interpreted as the effectiveness of principal j relative to the average effectiveness of principals in the same network.

There are two main identifying assumptions needed for equation II.2 to produce consistent estimates of δ_j. The first is that, conditional on S_{st}, θ_s is additively separable and fixed over time. In other words, once controlling for school characteristics, unobserved school factors that affect student attendance are the same in each year. This implies that estimates of principal quality (δ_j) are not conflated with the effects of school quality, and that there are no complementarities between principals and schools. The second assumption is that principal effects are time-invariant and immediate. In other words, a principal's effect on student outcomes in her first year is the same as her effect in any other year.

II.3.3 Approach 2: Drift-Adjusted Value-Added

A shortcoming of the simple two-way FE approach is that it is inflexible with respect to changes in principal effectiveness over time. However, prior research suggests two main reasons for why principal effectiveness changes over time. First, principals tend to become more effective as they gain experience (Grissom et al., 2018; Clark et al., 2009). Second, the potential for a principal to affect school performance may change across her tenure in a school. For instance, a new-to-school principal inherits many of the teachers hired under the old principal. To the extent that human capital management is an important avenue through which principals affect student outcomes (Jacob, 2011; Branch et al., 2012; Grissom and Bartanen, 2019b), it may take several years for the effect of a high-quality principal to manifest itself through improved student outcomes.

My preferred principal value-added model allows for changes in performance by producing an effect estimate for each principal-year observation. Specifically, I estimate a modified version of the estimator developed by Chetty et al. (2014) to estimate teacher value-added. The estimator has three steps to produce value-added for principal j in school s in year t: (1) residualize students' attendance (or achievement) outcomes on a vector of observable characteristics (the same student- and school-level controls from equation 1); (2) estimate the best linear predictor of mean attendance residuals for all students in school s with principal j in year t based on mean attendance residuals for principal j in prior or future years; (3) use the coefficients of the best linear predictor to predict principal value-added in year t.

One major difference between my approach and the approached used by Chetty et al. (2014) is that I include both school and principal fixed effects in the residualization step. Specifically, Chetty et al. (2014) residualize test scores on observable characteristics using within-teacher variation (i.e., with teacher fixed effects). When computing the residuals, they add back in the teacher fixed effects. I perform a similar process with principal effects, with the exception that the residualization includes both principal and school fixed effects. When computing the residuals, I add back in the principal fixed effects, but not the school fixed effects. This approach accounts for the possibility that the vector of observable characteristics does not fully control for between-school heterogeneity that is correlated with principal quality and student outcomes. However, as with approach 1, employing both principal and school fixed effects means that the value-added estimates are only comparable for principals within the same connected network.

Beyond producing a distinct effect estimate for each year, there are two important differences between the estimator developed by Chetty et al. (2014) and the fixed effects model in approach 1. First, the value-added estimates in approach 2 are leave-year-out measures, meaning that principal-school estimates in year t do not incorporate student outcomes from year t. Second, whereas the principal estimates from the fixed effects models in approach 1 includes school-level shocks and student errors (which is the motivation for the Empirical Bayes approach), approach 2 inherently produces shrunken estimates. To the extent that student residuals in a given year are higher or lower due to transitory shocks and/or student-level measurement error (as opposed to the true principal effect), this variation is uncorrelated (in expectation) with residuals in past or future years. Using these past and future residuals to predict contemporaneous residuals, then, will produce an estimate that is shrunken towards the sample mean.

Another important component of approach 2 is that it allows for "drift" in performance over time. Specifically, Chetty et al. (2014) exploit a stationarity assumption[8] that allows, for instance, the correlation of residuals between year t and year $t-1$ to be different from the correlation between year t and year $t-2$. As such, principal performance can vary over time for reasons other than idiosyncratic shocks. This is distinct from approach 1, which estimates an average effect across each principal's career (more accurately, for the years over the data stream).

Table II.2 shows the autocorrelation vectors for principal VA across each attendance and achievement outcome (Table II.A2 shows the corresponding autocovariances). The autocorrelations for achievement outcomes are substantially larger than those for teacher VA in Chetty et al. (2014). In other words, compared to teachers, principals' mean test-score residuals from prior and future years are much more predictive of current-year test-score residuals. For instance, math residuals from year $t+1$ or $t-1$ are correlated with math residuals from year t at 0.75, whereas Chetty et al. (2014) estimated correlations of 0.43 and 0.48 for teachers in elementary and middle school, respectively. This is not particularly surprising, given that the number of students used to produce these residuals is much larger for principals than for teachers (i.e., all students in the school versus a single classroom or handful of classrooms for teachers). An additional reason for the larger correlations is that, in contrast to teacher VA, principal VA draws on data from the same students in different

[8] Specifically, the assumption requires that mean principal quality does not vary over time and that the correlation of principal quality, school shocks, and student errors across years only depends on the amount of time between those years.

Table II.2: Autocorrelation Vectors for Principal Value-Added by Outcome

Lag	N	Att	Abs (U)	Abs (E)	Chr Abs	Math	Read	Sci
1	10571	0.67	0.72	0.62	0.69	0.75	0.64	0.72
2	7834	0.62	0.63	0.53	0.63	0.68	0.55	0.64
3	5667	0.58	0.58	0.49	0.61	0.64	0.49	0.59
4	3944	0.57	0.57	0.53	0.59	0.62	0.45	0.56
5	2642	0.59	0.57	0.53	0.61	0.63	0.42	0.59
6	1681	0.59	0.58	0.48	0.61	0.67	0.40	0.60
7	976	0.59	0.61	0.45	0.61	0.66	0.43	0.62

Notes: N refers to sample size for attendance models (achievement models have slightly smaller sample sizes).

years. In the case of teachers, VA in year t is predicted using test score residuals from different *cohorts* of students. As discussed above, however, since the majority of students remain in the same school in successive years, the residuals from prior and future years will contain many of the same students as the current year. This will increase the correlations among residuals in comparison to residuals from different cohorts of students.

There are similar patterns of "drift" across each of the outcomes. Beyond the first lag, the correlation with year t residuals decreases. However, the apparent lower bound on this drift is, again, much larger for principals than for teachers. Additionally, the drift pattern flattens out much more quickly—beyond the third lag, there is essentially no further drift except for excused absences and reading achievement. An additional note about these correlations is that the attrition rate is quite high, which is evident from the decreases sample size for larger lags. Compared to teachers, a much smaller proportion of principals remain in the data (i.e., as a principal in any school) over time. This makes the estimated correlations at larger lags substantially less precise. As such, when estimating principal VA I impose a drift limit of 5 (i.e., the covariances for lags larger than 5 are assumed to be the same as the covariance for 5). However, changing the drift limit does affect substantially the VA estimates, since the magnitude of drift beyond the first lag is fairly low. Figure II.A3 shows the estimated effect of principal value-added on current-year student outcomes. Under the stationarity assumption, this regression should have a slope of 1. Indeed, across each outcome, the estimates slopes are close to 1 (except for ELA, I cannot reject the null hypothesis that the slope is equal to 1). In Appendix II.6, I implement a test for forecast bias in the VA estimates using twice-lagged test scores, similar to that used in Chetty et al. (2014) and Rothstein (2010). I find that the drift-adjusted approach produces minimally biased estimates of principal VA.

II.4 Results

II.4.1 Do Principals Affect Student Attendance?

Table II.3 shows the distribution of principal value-added estimates from approaches 1 and 2—the two-way FE and drift-adjusted models—for attendance (total attendance, unexcused absences, excused absences, chronic absenteeism) and achievement (math, reading, science) outcomes. For attendance outcomes using the fixed effects approach, I can generate estimates for 3,353 principals—86 percent of the principals who worked in Tennessee K–12 public schools during this time.[9] The number of principal estimates for achievement outcomes is smaller, since some schools (e.g., grades K–2 only) do not have end-of-year or end-of-course exams. The drift-adjusted approach yields attendance estimates for 14,811 principal-by-year observations between 2008 and 2017 (80% of the total principal-by-year observations in Tennessee).[10] The standard deviations of the estimates represent how much principals vary in their effects on student outcomes; a larger standard deviation indicates that principal quality is more consequential for the particular outcome. Except for chronic absenteeism, which is expressed as a probability (since it can only take a value of 0 or 1), the units for attendance and achievement outcomes are student-level standard deviations. In other

[9]Most of the 14% of excluded principals are those who are in a network of a single principal (herself or himself). In this case, the principal and school FE are perfectly collinear and I cannot produce an estimate of the principal effect. A handful of principals are also excluded due to missing data.

[10]Here, the percentage of principals with non-missing estimates is lower because of the leave-year-out approach used to calculate drift-adjusted VA. To compute a drift-adjusted estimate, I effectively require that a principal had at least two principal-by-year observations.

Table II.3: Distribution of Value-Added Estimates

				Percentile of Estimates				
	N	SD	IQR	10th	25th	50th	75th	90th
Fixed Effects								
Attendance	3353	0.283	0.171	-0.274	-0.090	-0.002	0.080	0.231
Absences (U)	3353	0.311	0.190	-0.308	-0.096	-0.000	0.094	0.287
Absences (E)	3353	0.325	0.225	-0.327	-0.113	0.004	0.112	0.311
Chronic Absenteeism	3353	0.100	0.056	-0.093	-0.028	-0.000	0.028	0.081
Math	3074	0.280	0.223	-0.287	-0.118	-0.004	0.104	0.283
Reading	3064	0.228	0.125	-0.193	-0.068	-0.002	0.057	0.166
Science	3049	0.282	0.182	-0.244	-0.094	0.000	0.088	0.252
Drift-Adjusted								
Attendance	14811	0.225	0.141	-0.210	-0.076	-0.003	0.065	0.176
Absences (U)	14811	0.253	0.156	-0.252	-0.079	0.000	0.077	0.220
Absences (E)	14811	0.230	0.170	-0.236	-0.087	0.002	0.083	0.229
Chronic Absenteeism	14811	0.072	0.044	-0.069	-0.023	-0.000	0.021	0.062
Math	12790	0.205	0.153	-0.199	-0.078	-0.002	0.075	0.199
Reading	12756	0.097	0.068	-0.090	-0.036	-0.001	0.032	0.083
Science	12717	0.173	0.125	-0.151	-0.057	0.003	0.068	0.168

Notes: Absence estimates are multiplied by -1 to facilitate comparison with attendance and achievement estimates. Except for chronic absenteeism, all units refer to student-level standard deviations. Chronic absenteeism remains on a 0 to 1 scale. Sample sizes for fixed effects are at the principal level, whereas drift-adjusted are at the principal-by-year level.

words, they represent a student's rank in the distribution within each grade and year. This allows for a direct comparison of the magnitude of principal effects on attendance versus achievement outcomes.

Studies of principal and teacher effects typically use the standard deviation of the VA estimates as the preferred measure of variability. In other words, these studies interpret the magnitude of effects according to the standard deviation of the estimates. This approach makes sense if the estimates are normally distributed. However, as shown in Figure III.A11, the distribution of the drift-adjusted estimates is non-normal. Specifically, the presence of a small number of outliers inflates the standard deviation. To provide a more substantive

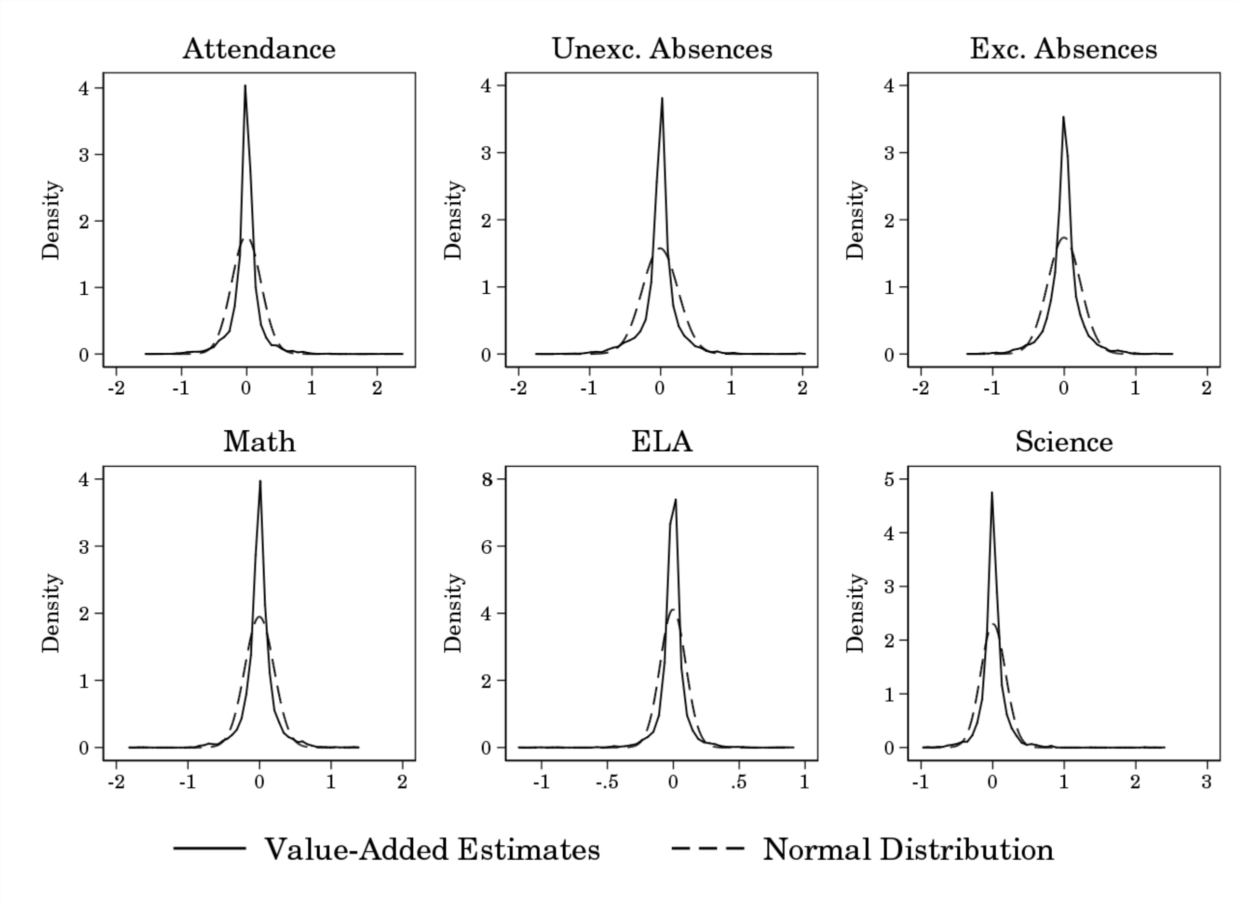


Figure II.2: Distribution of Drift-Adjusted Value-Added Estimates

interpretation of the magnitude of principal effects on student outcomes, I compute the interquartile range (i.e., the difference between the 25th and 75th percentile in the distribution), which is my preferred measure of the variability given these non-normal distributions.

I focus my discussion on the estimates from the drift-adjusted approach.[11] Comparing across the estimates, I find that principals' effects on attendance outcomes are comparable to—if not larger than—their effects on achievement outcomes. For instance, the IQR of principal VA to attendance rates is 0.17 student-level standard deviations, which is larger than reading VA (0.07 SD) and science VA (0.13 SD). Substantively, moving from the 25th to the 75th percentile in the distribution of principal VA increases student attendance by

[11] As shown in Table II.A3, the FE and drift-adjusted estimates are highly correlated, which is to be expected given that the drift-adjusted estimates are derived from the fixed effects estimates. However, the fixed effects estimates are unshrunken, which inflates the standard deviation and IQR.

0.14 standard deviation, which corresponds to 2.2 additional instructional days on a 180-day calendar. In terms of both SD and IQR, principal effects on attendance and math are roughly equal in magnitude.

In addition to overall attendance, I also construct estimates for three other attendance-related measures: unexcused absence rates, excused absence rates, and chronic absenteeism. For each of these measures, I multiply the estimate by -1 such that *increases* in VA correspond to *decreases* in absenteeism. I find that principals' effects on attendance operate through effects on both excused and unexcused absences.[12] Unlike the other outcomes in Table II.3, chronic absenteeism estimates are expressed in terms of the probability that a student has an attendance rate below 90%, rather than standard deviations. I find that moving from the 25th to 75th percentile lowers the probability that a student will be chronically absent in the current year by 4.4 percentage points, or roughly 30% of the base rate among Tennessee students. Overall, Table II.3 demonstrates that principals have substantive effects on student attendance, which operate through effects on both excused and unexcused absences.

II.4.2 Does the Importance of Principal Quality for Attendance Vary by School Context?

My second research question examines whether the magnitude of principal effects vary by three categories of school context: school level, school locale, and school poverty (as measured by the percentage of students who quality for free/reduced-price lunch). Table II.4 shows the IQR of the drift-adjusted VA estimates within each of the levels of these categories.[13] One important note is that for attendance VA, instead of reporting the VA estimates produced from standardized attendance/absence rates, I use the estimates for the *unstandardized* rates. As previously shown in Figure II.1, attendance rates vary substantially by grade, and are systematically lower in early elementary and high school grades. While using a standardized

[12]Note that because the attendance outcomes are standardized rates, they are not directly comparable with one another in terms of magnitude. A standard deviation increase in total attendance rate represents more absences than a standard deviation decrease in unexcused or excused absence rate, for instance.

[13]Table II.A4 shows the standard deviations.

Table II.4: Interquartile Range of Drift-Adjusted Value-Added Estimates by School Characteristics

	Att%	Abs% (U)	Abs% (E)	Chr Abs	Math	Read	Sci
School Level							
Elementary	1.0	1.0	0.7	0.045	0.156	0.070	0.128
Middle	1.1	0.9	0.8	0.040	0.142	0.069	0.121
High	1.2	1.1	0.8	0.043	0.152	0.062	0.125
School Locale							
Urban	1.4	1.5	1.0	0.061	0.204	0.090	0.162
Suburban	1.0	0.8	0.7	0.044	0.143	0.063	0.132
Town/Rural	0.9	0.8	0.7	0.038	0.132	0.059	0.107
School Poverty							
0–30% FRPL	0.8	0.7	0.6	0.033	0.152	0.073	0.126
30–80% FRPL	1.0	0.9	0.7	0.040	0.139	0.062	0.115
80–100% FRPL	1.6	1.7	1.1	0.069	0.226	0.095	0.170

Notes: Attendance and absence outcomes are expressed as rates that range from 0 to 100%.

measure is helpful for comparing estimates of principal effects on attendance and achievement (i.e., comparing how much principals matter in terms of changing students' *rank* in the distribution within their given grade and year), these standardized measures will, for instance, understate the importance of principal quality for high school attendance. To ensure the comparisons between school types are not unduly influenced by differences in average attendance rates by grade, I examine the unstandardized outcomes.[14] Specifically, the units are unadjusted attendance/absence rates that range from 0 to 100%.

In terms of school level, the IQR of principal effects on attendance rates are similar. For example, moving from the 25th to 75th percentile in principal quality for high schools increases the average student's attendance rate by 1.2 percentage points, which is 20% larger

[14] While the standardization changes the scale of the principal effects, the rank correlations are very high between models that use standardized versus unstandardized attendance outcomes.

than the magnitude for elementary school principals. However, the IQR for effects on chronic absenteeism is slightly *larger* in elementary schools (0.045) than high schools (0.043). This is likely due to the fact that elementary schools have more students near the chronic absenteeism cutoff (10% absence rate), such that marginal increases in attendance are more likely to move a student under this cutoff. For achievement outcomes, the magnitude of principal effects is again quite comparable across school levels.

Across locale and school poverty, a clear pattern emerges: principal effects in urban and high-poverty schools are larger across both attendance and achievement outcomes. For instances the IQR in urban schools 1.4 percentage points, compared to 1.0 and 0.9 in suburban and town/rural schools. This difference is largely driven by unexcused absences, where the magnitude in urban schools (1.5 percentage points) is nearly twice as large as suburban and town/rural schools (0.8 percentage points). I find similarly sized differences for student achievement in math, reading, and science. The largest differences in the magnitudes of principal effects on attendance are between high-poverty and low-poverty schools. For instance, the IQR for attendance rates is twice as large in high-poverty schools (1.6 percentage points) as low-poverty schools (0.8 percentage points). Moving from the 25th to 75th percentile in principal quality lowers the probability of chronic absenteeism by nearly 7 percentage points in high poverty schools.

II.4.3 Comparing Attendance Effects to Other Measures of Principal Quality

My final research questions examines whether principals who excel at increasing student attendance can be identified using other measures of principal effectiveness. Specifically, I compare principals' attendance VA to their achievement VA, as well as high-stakes rubric-based observation scores from their supervisors. Table II.5 compares principals' estimated VA to attendance and achievement outcomes. Given the findings in Table II.3 that principals have substantial effects on student attendance, a natural question is whether principals who increase student achievement also improve attendance outcomes. To answer this question,

Table II.5: Spearman Correlations Among Value-Added Estimates

	Att	Abs (U)	Abs (E)	Chr Abs	Math	Read	Sci
Attendance	1.00						
Absences (U)	0.62	1.00					
Absences (E)	0.40	-0.19	1.00				
Chronic Absenteeism	0.82	0.54	0.45	1.00			
Math	0.11	0.08	0.02	0.07	1.00		
Reading	0.06	0.05	0.02	0.06	0.47	1.00	
Science	-0.05	-0.04	-0.01	-0.03	0.36	0.45	1.00

I compute Spearman rank correlations among the VA estimates.[15] The first column of Table II.5 compares principals' attendance VA to other outcomes. There are mechanical correlations between attendance and absence VA estimates, given that a student's attendance rate is determined by the sum of their excused and unexcused absences.

Comparing principals' effects on attendance to achievement scores, there are small positive correlations with math and reading VA for each model. For instance, the rank correlation between attendance and math (reading) VA is 0.11 (0.06). By contrast, there is a small *negative* correlation between a principal's impact on attendance versus science achievement (-0.05). In substantive terms, these findings demonstrate that principals who improve student attendance are not necessarily the same principals whose students have the greatest achievement growth. To the extent that student attendance is an important educational outcome even beyond its relationship with achievement, the results in Table II.5 suggest that focusing exclusively on student achievement to identify principal quality will fail to capture the contributions of principals to improving students' non-cognitive or character skills, such as attendance.

Table II.6 examines correlations between drift-adjusted principal VA estimates and prin-

[15]Pearson correlations produce very similar results to those shown in Table II.5.

Table II.6: Predicting Principal Ratings from Supervisors Using Value-Added Estimates

	Att (1)	Abs (U) (2)	Abs (E) (3)	Chr Abs (4)	Math (5)	Reading (6)	Science (7)
Principal Value-Added (std)	0.006	0.023	-0.022	0.008	0.037	0.044**	0.080***
	(0.015)	(0.022)	(0.026)	(0.019)	(0.023)	(0.018)	(0.020)
Network FE	✓	✓	✓	✓	✓	✓	✓
District-by-Year FE	✓	✓	✓	✓	✓	✓	✓
N	7326	7326	7326	7326	5886	5855	5847

Notes: Standard errors clustered by district in parentheses. Dependent variable is a principal's standardized average observation score. Principal VA estimates are estimated using the drift-adjusted approach and subsequently standardized. Column headers denote the outcome used to calculate principal value-added. Sample sizes refer to principal-by-year observations, which begin in 2011–12.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

cipals' rubric-based ratings from supervisors. Specifically, I regress standardized supervisor ratings on principal value-added (standardized), fixed effects for each connected network used to construct principal VA, and fixed effects for district-by-year. Columns 1–4 show that there is no correlation between attendance VA and supervisor ratings. In other words, these rubric-based ratings do not contain information about principals' contributions to increasing student attendance. Columns 5–7 show the results for achievement VA in math, reading, and science. Here, there is a positive correlation between supervisor ratings and value-added. However, the magnitude of these relationships is quite smaller. For instance, a 1 SD increase in principal VA to math achievement is associated with only a 0.037 SD increase in supervisor ratings. Reading and science VA ($\beta = 0.044$ and $\beta = 0.08$) have only slightly stronger correlations with supervisor ratings.

II.5 Discussion and Conclusions

Student attendance is increasingly recognized as an important measure of school success, which has spurred research that examines the extent to which schools affect attendance outcomes. To date, studies have almost exclusively focused on teachers, and we have convincing evidence that teachers play an important role in decreasing student absenteeism. However, no studies have considered the effect of principals, despite strong conceptual reasons to believe that principals can influence absenteeism. This study makes several contributions to the existing literature. First, to my knowledge this is the only study that estimates principal effects on attendance. Second, I employ an estimator for value-added that improves upon existing studies of principal effects. Third, I shed light on the multidimensional nature of principal quality by comparing principals' effects on attendance to their effects on achievement.

The central finding of this paper is that principals have substantive effects on student attendance. Moving from the 25th to 75th percentile in principal quality raises student attendance by 0.14 standard deviations, which corresponds to 2.2 additional instructional days. Further, principals have even larger effects in high-poverty and urban schools, which have the highest rates of student absenteeism. From the perspective of policymakers and district leaders, this suggest that intervening with principals could be an effective means to address high rates of chronic absenteeism. That said, an important limitation is that this analysis is unable to identify the pathways through which some principals are effectively lowering absence rates. Future work should aim to identify these mechanisms, which could provide useful guidance about specific ways to target development opportunities.

Another important finding is that, similar to studies of attendance effects for teachers, principals who improve student attendance are not necessarily those who increase test scores. This highlights the multidimensional nature of principal quality and suggests that, insofar as attendance is an outcome worthy of attention, accountability systems designed around identifying principals who increase test scores will fail to identify principals who are improving attendance—and by extension, principals who improve other non-test-score outcomes. Indeed, I find that principals' rubric-based scores from supervisors, which comprise half of their summative evaluation rating, are not predictive of their contributions to improving student attendance.

II.6 Appendix

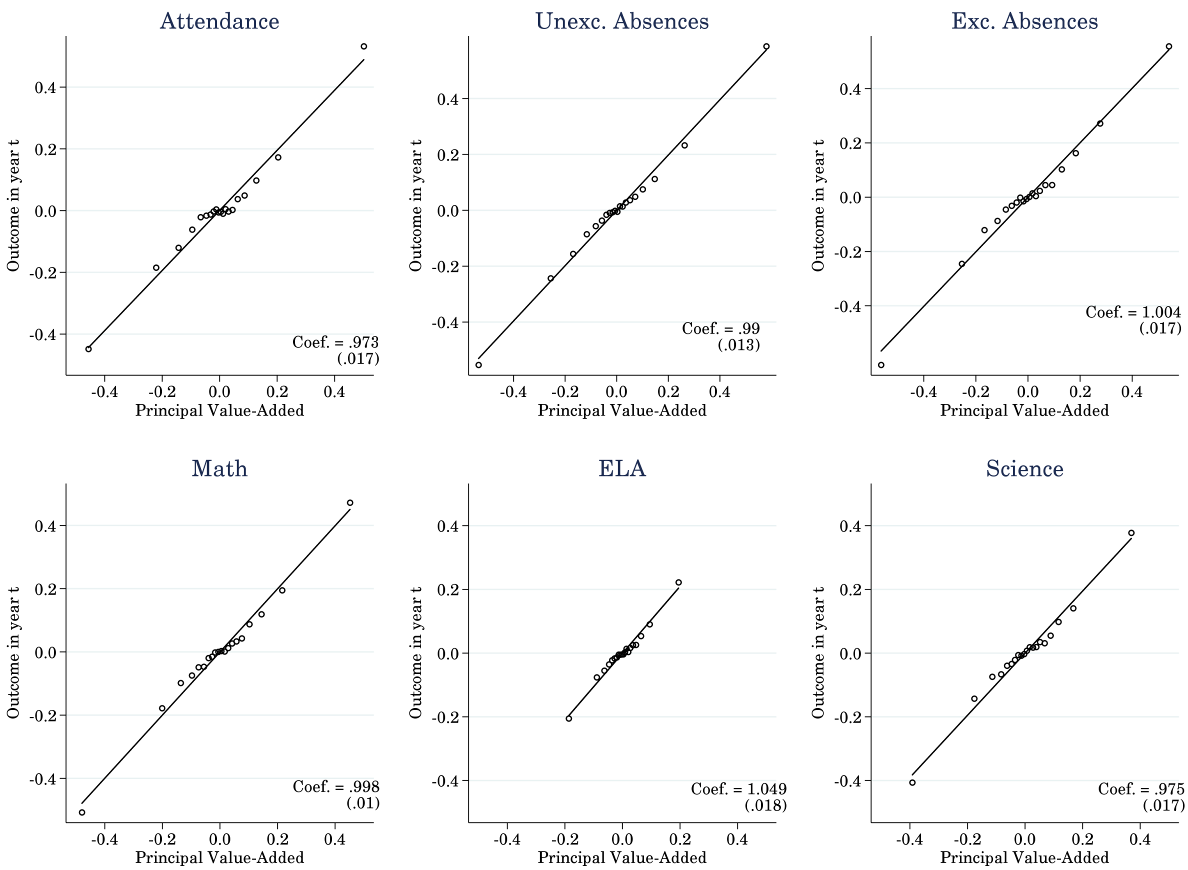


Figure II.A3: Effect of Principal Value-Added on Current-Year Outcomes

Table II.A1: Distribution of Principal Networks (Math Achievement)

Network Size (# of Schools)	# of Networks	Mean # of Principals	Total Principals	% of Principals
1	527	2.4	1251	40.7
2	133	3.9	517	16.8
3	54	5.7	307	10.0
4	19	8.2	156	5.1
5	9	9.8	88	2.9
6	3	15.0	45	1.5
7	8	14.9	119	3.9
8	4	19.0	76	2.5
9	2	14.0	28	0.9
10	1	21.0	21	0.7
11	1	25.0	25	0.8
14	1	26.0	26	0.8
30	1	64.0	64	2.1
35	1	65.0	65	2.1
43	2	83.5	167	5.4
65	1	119.0	119	3.9
All	767	4.0	3074	100.0

Notes: Networks refer to the mobility groups of principals and schools from a two-way FE model for math achievement.

Table II.A2: Autocovariance Vectors for Principal Value-Added by Outcome

Lag	N	Att	Abs (U)	Abs (E)	Chr Abs	Math	Read	Sci
1	10571	0.043	0.057	0.061	0.005	0.040	0.008	0.028
2	7834	0.041	0.051	0.053	0.005	0.038	0.007	0.026
3	5667	0.041	0.048	0.049	0.005	0.037	0.007	0.025
4	3944	0.041	0.048	0.053	0.005	0.036	0.006	0.025
5	2642	0.046	0.050	0.052	0.006	0.038	0.007	0.027
6	1681	0.051	0.056	0.050	0.006	0.041	0.007	0.029
7	976	0.055	0.061	0.052	0.006	0.037	0.009	0.031

Notes: N refers to sample size for attendance models (achievement models have slightly smaller sample sizes).

Table II.A3: Correlation Between Value-Added Estimates from Fixed Effects and Drift-Adjusted Models

	Spearman	Pearson
Attendance	0.94	0.89
Absences (U)	0.96	0.91
Absences (E)	0.92	0.90
Chronic Absenteeism	0.96	0.92
Math	0.94	0.92
Reading	0.85	0.89
Science	0.92	0.90

Table II.A4: Standard Deviation of Drift-Adjusted Value-Added Estimates by School Characteristics

	Att%	Abs% (U)	Abs% (E)	Chr Abs	Math	Read	Sci
School Level							
Elementary	1.7	1.5	1.0	0.077	0.207	0.110	0.190
Middle	1.4	1.4	1.1	0.063	0.192	0.078	0.150
High	1.9	1.8	1.1	0.066	0.212	0.076	0.149
School Locale							
Urban	2.1	2.0	1.3	0.086	0.257	0.113	0.190
Suburban	1.7	1.8	1.1	0.068	0.175	0.100	0.219
Town/Rural	1.3	1.2	0.8	0.064	0.174	0.084	0.144
School Poverty							
0–30% FRPL	1.6	1.5	1.0	0.058	0.176	0.082	0.157
30–80% FRPL	1.6	1.5	0.9	0.066	0.187	0.089	0.163
80–100% FRPL	2.0	1.9	1.3	0.093	0.263	0.126	0.207

Notes: Attendance and absence outcomes are expressed as rates that range from 0 to 100%.

Table II.A5: Replication of Table II.A4 Using Standardized Attendance Outcomes

	Att	Abs (U)	Abs (E)	Chr Abs	Math	Read	Sci
School Level							
Elementary	0.244	0.272	0.232	0.077	0.207	0.110	0.190
Middle	0.194	0.228	0.237	0.063	0.192	0.078	0.150
High	0.188	0.216	0.208	0.066	0.212	0.076	0.149
School Locale							
Urban	0.282	0.315	0.288	0.086	0.257	0.113	0.190
Suburban	0.231	0.284	0.236	0.068	0.175	0.100	0.219
Town/Rural	0.181	0.199	0.184	0.064	0.174	0.084	0.144
School Poverty							
0–30% FRPL	0.207	0.242	0.203	0.058	0.176	0.082	0.157
30–80% FRPL	0.213	0.238	0.206	0.066	0.187	0.089	0.163
80–100% FRPL	0.268	0.303	0.301	0.093	0.263	0.126	0.207

Testing for Forecast Bias in Principal Value-Added Estimates

One major shortcoming in the principal effects literature is that few studies have explicitly tested the extent to which the estimates produced by principal VA models are valid or reliable measures of principal performance. This stands in stark contrast to the teacher effects literature, where a large number of studies have attempted to validate teachers' value-added estimates. Among the most convincing of such studies is Chetty et al. (2014), who propose sorting tests using students' twice-lagged test scores and information about family income from tax data, as well as a quasi-experiment that leverages plausibly exogenous variation in grade-level teacher turnover. Other studies have used random assignment of students and teachers to classrooms to test the validity of VAMs. Unfortunately, in the case of principals the options are much more limited.

I test for bias in drift-adjusted VA using a sorting test similar to that employed in Chetty et al. (2014). Specifically, Chetty et al. (2014) show that one can estimate forecast bias in VA estimates by regressing predicted test scores (based on observable characteristics excluded from the VA model) on the VA estimates. The intuition of this approach is that, if VA estimates are forecast unbiased (i.e., the control vector fully accounts for nonrandom sorting), these omitted characteristics will not be correlated with their principal's VA in the current year. Among the characteristics used to test for bias is twice-lagged test score (Rothstein, 2010). I also draw on students' prior outcomes that are omitted from the VA estimation. Here, however, the procedure is slightly more complex given the aforementioned issue of endogeneity between principal fixed effects and prior-year outcomes. Specifically, my VA models control for a *prior-school* outcome, which is often not the most recent prior score. Thus, for the sorting test I draw on a student's attendance/achievement outcome

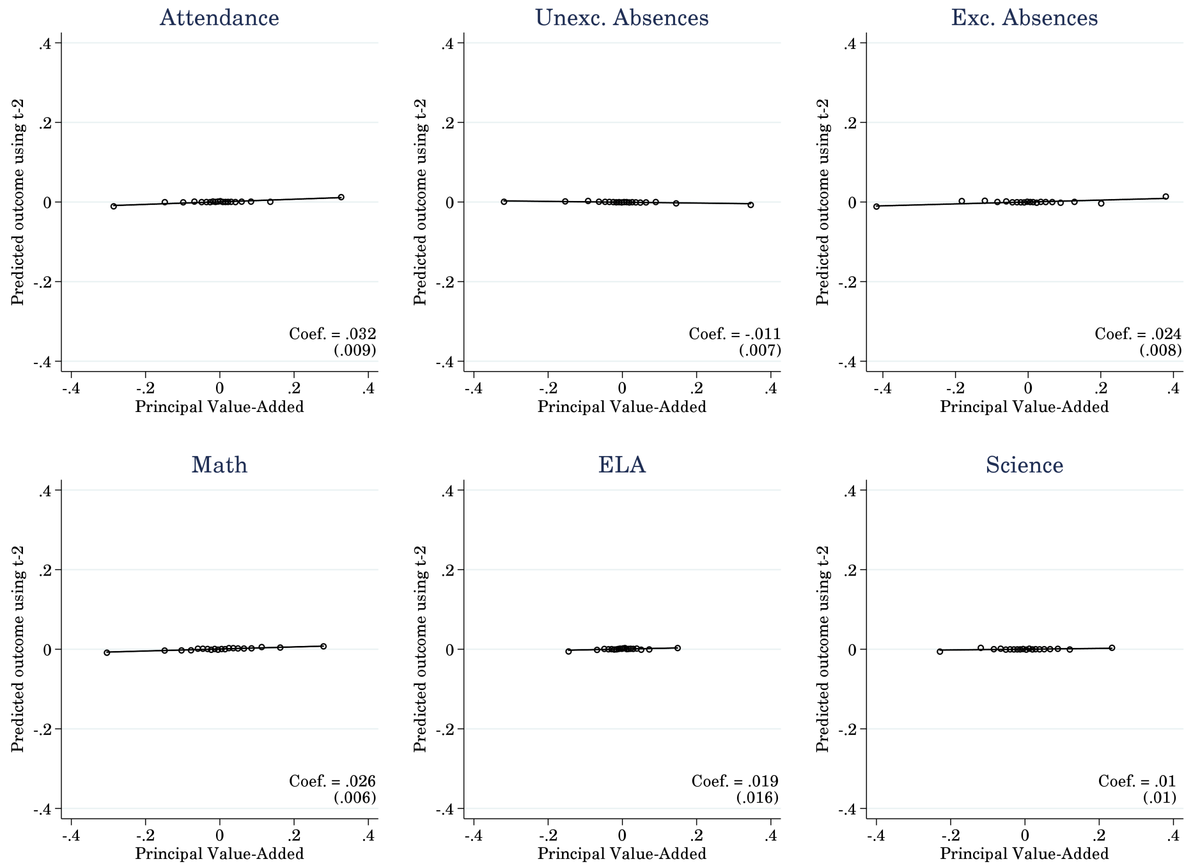


Figure II.A4: Effect of Principal Value-Added on Prior-School Outcomes

from the year prior to the year of the score used to estimate principal VA. For instance, if an 8th grade student's most recent prior-school outcome is from 5th grade, I use their 4th grade outcome to estimate forecast bias.[16]

The results of the test for forecast bias are shown for attendance and achievement outcomes in Figure II.A4. Each plot shows the coefficient from regressing the predicted score (based on twice-lagged prior-school outcomes) on the principal VA estimates. Across each outcome, the estimated slope is close to zero, indicating little to no forecast bias. For instance, the attendance results (top left) imply that a 1 SD increase in principal VA (roughly 0.22 on the VA scale) is associated with a 0.007 SD increase in predicted student attendance. By comparison, the equivalent test by Chetty et al. (2014) for teacher VA in math and reading yielded a coefficient of 0.022, which is similar in magnitude to all of the estimates in Figure II.A4.

[16]Perhaps a more straightforward approach is to restrict the sample to students who made a structural move in the current year (i.e., 9th grade students who moved from a middle school to a high school). In this case, the prior-school outcome would simply be the prior-year outcome, and I could then use the twice-lagged outcome to test for bias. However, this greatly limits the sample and I find very similar results compared to when using all available students.

CHAPTER III

The Impacts of Principal Turnover

Principal turnover rates in the United States are high, surpassing even the rates of teacher turnover that have motivated so much policy attention and research (see Grissom et al., 2015b; Guarino et al., 2006; Barnes et al., 2007; Carroll, 2007). According to recent National Center for Education Statistics (NCES) estimates, principal turnover rates are approximately 18% nationally (Goldring and Taie, 2018). The importance of school leadership and the numerous pathways through which school principals impact students and teachers (e.g., Boyd et al., 2011; Coelli and Green, 2012; Grissom, 2011; Grissom et al., 2015a; Kraft et al., 2016) raise substantial concern about the frequency of principal turnover and the potential for it to negatively impact schools. For example, principals can drive school improvement through instructional leadership, building a learning climate at the school, supporting teacher improvement, and implementing strategies to recruit and retain effective teachers (Grissom et al., 2013; Sebastian and Allensworth, 2012; Robinson et al., 2008; Grissom and Bartanen, 2019b; Simon and Johnson, 2015), processes with important relational components that take time to nurture and that may be disrupted in a transition to a new school leader.

The likely disruptive effects of principal turnover might be offset or magnified depending on whether turnover leads to a gain or loss in principal quality—that is, whether, on average, turnover means the arrival of a more effective or less effective replacement. New principals typically have fewer years of school leadership experience than the principals they replace (Grissom et al., 2019), which, given evidence that principals become more effective with experience (Clark et al., 2009; Grissom et al., 2018), suggests that replacement exacts costs on effectiveness. On the other hand, Grissom and Bartanen (2019a) find that turnover among Tennessee principals is concentrated among the state's lowest performers, suggesting that replacement may improve principal quality in some schools.

The question of how these potential costs and benefits balance have motivated several studies of how school outcomes, including student achievement, change when a principal turns over (e.g., Béteille et al., 2012; Miller, 2013). These studies have documented that student test scores fall and teacher turnover increases in the years following a principal transition. However, prior studies have not overcome a key challenge to interpreting these correlations as causal—namely, that school performance typically begins falling prior to a turnover event (Miller, 2013), making it difficult to disentangle the effects of the turnover event from the effects of factors that led to the turnover event, and to account for the potential "mean reversion" that may follow a principal transition. If principals leave in response to declines in school performance or other time-varying factors, the school fixed effects strategies pursued in prior work will be insufficient for estimating unbiased turnover impacts. Importantly, these prior estimates could be biased upward or downward from the true causal effect of principal turnover, raising the possibility that existing work provides incorrect guidance for school districts seeking to improve management and allocation of school leaders.

This study extends our understanding of the consequences of principal turnover by producing plausibly causal estimates of the effects of principal turnover on a variety of school outcomes. We employ a difference-in-differences design with a matched comparison group to resolve the endogeneity between principal transitions and school performance. Specifically, for each "treatment" school experiencing a principal turnover event between year t and $t+1$, we construct a "comparison group" of schools with identical trajectories of achievement and personnel turnover over the prior five years who do not experience a turnover event between year t and $t+1$. Outcomes in these comparison schools provide a reasonable counterfactual for what would have happened in treatment schools in the absence of the principal turnover event, permitting us to isolate the average effect of a principal transition.

Beyond the explicit identification of a counterfactual group for producing arguably causal estimates, our study moves beyond simple dichotomous measures of principal turnover (i.e., whether the principal remains in the school in the following year) to consider whether effects vary by different types of leadership transitions (e.g., transfers to a different school, promotions to central office). This innovation is important because not all principal turnover events are created equal. For example, Grissom and Bartanen (2019a) find that low-performing principals are substantially more likely to be demoted into non-principal school-level positions or to exit education altogether, while high performers are more likely to move into central administration. Relatedly, Walsh and Dotter (2019) find positive effects on student achievement of dismissals of presumably under-performing principals. Different kinds of principal transitions thus may have different net impacts on school outcomes.

Using longitudinal administrative data from Missouri and Tennessee, we aim to answer three main research questions. First, what is the effect of principal turnover on school performance, especially student achievement, up to five years after the transition? Second, to what extent do these effects vary by the type of principal turnover (transfers, exits, promotions, demotions)? Finally, what mechanisms explain the connection between principal turnover and school performance? Our answers to this last question extend the literature significantly beyond investigating the role of teacher retention in linking principal turnover to student achievement (Béteille et al., 2012; Miller, 2013) to examine changes in teacher quality, school climate, and principal quality in explaining principal turnover effects.

III.1 Connecting Principal Turnover to School Performance

Following Ronfeldt et al. (2013), who examine how teacher turnover harms student achievement, we hypothesize two broad mechanisms through which principal turnover may affect school performance: disruptive effects and replacement effects.[1] Disruptive effects arise if principal transitions undermine important channels through which principals affect school outcomes. Research suggests many such possibilities. As one example, principals influence the organizational climate and culture of the school via setting clear goals, establishing expectations, engaging with the community, and creating structures to facilitate teachers' work,

[1] In the case of teachers, Ronfeldt et al. (2013) refer to these as "compositional effects."

which in turn can focus and enhance instructional practices and increase student achievement (e.g., Sebastian and Allensworth, 2012; Robinson et al., 2008; Leithwood and Jantzi, 2000; Marks and Printy, 2003). Gaining an understanding of school needs and establishing relationships with teachers and other members of the school community likely are fundamental to this chain; it may take new leaders considerable time to build the required knowledge and relationships that exiting principals possess. As another example, principals can drive student outcomes via effective human capital management, including strategic hiring and retention of effective teachers (Grissom and Bartanen, 2019b; Master, 2014). To the degree that principal transitions create instability and uncertainty about future leadership, such transitions may make it more difficult for a school to attract and retain high-performing teachers.

Disruptive effects of principal turnover likely are negative. As Ronfeldt et al. (2013) note, however, the overall impact of turnover depends on the combination of these negative disruptive effects with any *replacement* effects. Replacement effects are those associated with acquiring a new principal who is more or less effective than the outgoing principal. If districts tend to replace outgoing principals with even less effective new principals, replacement effects will be negative. In contrast, if principal turnover tends to result in higher-quality leadership in the principal's office, replacement effects will be positive, and potentially even positive enough to outweigh any disruptive effects. Importantly, it may take several years for positive replacement effects to outweigh disruptive effects, making principal turnover harmful in the short-term but beneficial in the medium- to long-term.

Several prior studies have attempted to test the overall effects of principal turnover on school, teacher, and student outcomes. However, measuring the causal effect of changing principals is challenging because principal turnover is not an exogenous event. As Miller (2013) demonstrates in a study of principal turnover in North Carolina, school performance is on a downward trajectory, on average, for several years prior to when the principal leaves the school. This pattern raises the possibility that declines in school performance may drive turnover, making it challenging to isolate the effect of the transition. The rebounding to the school's "steady state" one or two years after the principal transition that Miller (2013) finds may have occurred even if the school had kept the same principal. Other confounding factors may also be at play. For example, in the case of struggling school, district leaders may choose to change principals in conjunction with other strategies to improve school performance, such as targeted professional development for teachers or funding for new instructional materials. Alternatively, the downturn in school performance may be driven by short-term factors that the principal cannot control, such as neighborhood violence or a spate of retirements among effective teachers. In either of these situations, a naïve comparison of pre-transition and post-transition outcomes fails to isolate the effect of a principal switch.

To our knowledge, only one existing study has employed an empirical strategy that isolates the effect of principal turnover from these alternative explanations. Walsh and Dotter (2019) use a difference-in-differences design to estimate the effect of a principal replacement policy in Washington, DC Public Schools, which was part of a broader set of accountability reforms in the district. They find that schools whose principals were replaced under the policy had 0.09 SD higher school-wide student achievement after the third year of the new principal compared to schools that kept the same principal.[2] Our study complements this one by examining effects of a broad range of turnover types (as opposed to a specific replacement policy) across two states. Distinguishing among different types of principal turnover (e.g., exits, promotions to central office), which other studies generally have not done, is important because transitions of different types may represent different processes with different consequences for school performance. For example, exits and demotions to non-principal positions may be more likely among low-performers who are removed or counseled out of the principal role, while moves to central office may be more likely for more effective principals

[2] In a study of data from Miami-Dade County Public Schools, Béteille et al. (2012) find that having a new principal is associated with a decline in math scores of 0.007 SD, and that this relationship is driven by new principals who are novices, i.e., who have no prior principal experience. As the authors note, the interpretation of this estimate as the causal effect of the transition rests on the assumption that schools are not experiencing a downward achievement trajectory prior to principal turnover, which the study does not explicitly examine.

(Farley-Ripple et al., 2012; Grissom and Bartanen, 2019a). Replacement effects may vary in these cases. To the extent that some kinds of transitions are anticipated and planned for more than others, disruptive effects may vary as well.

Building on existing research, our study makes three main contributions to our understanding of the effects of principal turnover. First, we employ an identification strategy that more plausibly isolates the impact of principal turnover from the circumstances that may have led to the transition. Second, we move beyond a binary turnover measure to examine whether specific types of principal transitions are more or less harmful to school performance. Finally, we investigate a range of potential mechanisms—both compositional and disruptive—through which the effects of principal turnover may operate.

III.2 Data and Measures

This study uses longitudinal administrative data from two states: Missouri and Tennessee. As of 2016, Missouri has 65,000 public school teachers working in 2,400 schools and 565 districts, while Tennessee has 72,000 public school teachers working in 1,800 schools and 147 districts. Missouri data were obtained via a data request to the Department of Elementary and Secondary Education, while the Tennessee data were accessed through the Tennessee Education Research Alliance (TERA) at Vanderbilt University with approval from the Tennessee Department of Education. Missouri personnel records were available from 1991 to 2016, while Tennessee records spanned 2002 to 2016. However, the analysis years begin in 2001 and 2007, respectively, due to the unavailability of student achievement data in earlier years.[3] We also draw on school information contained in the Common Core of Data (CCD), a repository of school-by-year information maintained by the National Center for

[3]Specifically, we leverage publicly available school achievement data from Missouri that provide mean scale scores at the school-by-grade-by-year level. These files were directly downloaded from the website for the Missouri Department of Elementary and Secondary Education. To create a school-level measure for math and reading achievement, we first standardize these grade-level scale scores across the state within subject, grade, and year. We then collapse these grade-level scores to the school level, weighting by the enrollment in each grade. Finally, we standardize these school-level averages within each year. In Tennessee, we construct a parallel measure using the student-level scale scores. We first average these scale scores within each school-by-grade-by-year cell, then follow the same procedure used for the Missouri data.

Education Statistics. We use information about student enrollment size, proportion of Black students, proportion of Hispanic students, and the proportion of students eligible for free- or reduced-price lunch, which serves as a measure of school poverty.

III.2.1 Measuring Principal Turnover

Both datasets contain job classification and location information that allows identification of principals and what school they worked in each year. From these longitudinal administrative data files, we create a binary principal turnover variable, which takes a value of 1 if a principal in a school in year t is not the principal in the school in year $t+1$, and zero otherwise. Figure III.1 shows the distribution of the number of principal transitions experienced by a school across the study period. In both states, almost all schools changed principals at least once and many had multiple transitions, a complication that we address in the modeling section.

We also examine the effects of different types of principal turnover: transfers, exits, promotions, and demotions. Transfers are principals in year t who move to a principal position in a different school in year $t+1$. In Missouri, roughly half of transfers are to a different district, compared to only 10% in Tennessee.[4] Exits are principals in year t who are not working in the state's K–12 public education system in year $t+1$. These exits include principals who retire, principals who are fired by the district (or whose contracts are not renewed), and principals who chose to leave public education (e.g., to work in a private school or a different industry); we cannot differentiate among these exit types. Promotions are principals in year t who move "upward" to a central office position in year $t+1$, such as superintendent or principal supervisor. Demotions are moves "downward" from a principal position in year t to a school-based, non-principal position in year $t+1$. The majority of demotions in both states are principals moving to assistant principal positions.

Table III.1 shows yearly principal turnover rates in each state for the analysis period. Principal turnover rates are slightly higher in Missouri, on average, at 20% each year, com-

[4] Missouri has a much larger number of school districts and schools, despite having a smaller population than Tennessee.

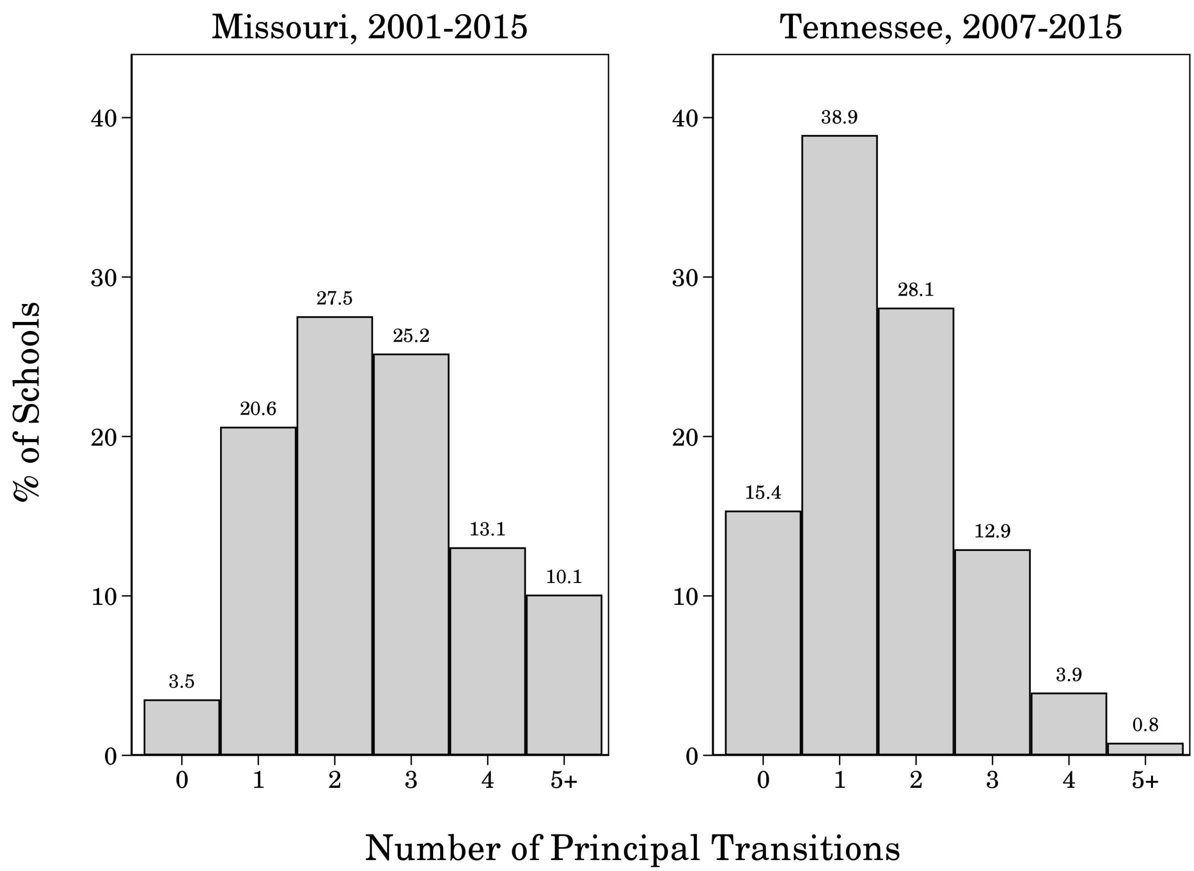


Figure III.1: Number of Principal Transitions Experienced Within Schools

Notes: Number of principal transitions refers to the number of times that a given school changed principals during the study period (shown in the plot header).

pared to 18% in Tennessee. The breakdown of turnover into types is also fairly similar between the states. Among the four types of principal turnover we categorize (transfers, exits, promotions, and demotions), exits are the most common, followed by transfers, which are slightly more frequent in Missouri. While overall rates of position changes are similar between Missouri and Tennessee, demotions are relatively more common in Tennessee, with more principal promotions to central office in Missouri.

III.2.2 Outcome Measures

Our main outcome of interest is student achievement in math and reading. Specifically, we examine school-average achievement scores on statewide exams in grades 3 to 8 and end-of-course exams for high school students. We do not have student-level scores for Missouri, so we convert school mean scale scores to school-level standard deviations (standardized within year). We construct a parallel outcome in Tennessee by aggregating student-level scale scores to the school level. Thus, our achievement outcomes reflect school performance relative to the distribution of schools in each state.

In addition to achievement outcomes, we estimate the effect of principal turnover on teacher turnover. Specifically, we operationalize teacher turnover in two ways: the proportion of teachers in the school in year $t-1$ who are still in the school in year t and the number of teachers who are new to the new school in year t. These two measures are highly correlated. We provide both to complement traditional analysis of turnover rates with an estimate (number of new-to-school teachers) that is concrete and helpful in interpreting magnitude.

We also draw on measures of teacher quality and school climate to probe the potential mechanisms connecting principal turnover to school performance. These measures are only available in the Tennessee data. Specifically, we construct a yearly measure of a school's average teacher quality by estimating teacher value-added (VA) in math and reading using the drift-adjusted approach proposed by Chetty et al. (2014).[5] We average the teacher VA

[5]The estimation steps are as follows. First, we residualize student test scores (separately by subject) on a vector of prior-year test scores, student characteristics (race/ethnicity, gender, FRPL eligibility, gifted

Table III.1: Principal Turnover by Year in Missouri and Tennessee

	All Turnover	Transfers	Exits	Promotions	Demotions
Missouri					
2001	0.198	0.060	0.072	0.042	0.024
2002	0.197	0.054	0.077	0.040	0.026
2003	0.203	0.060	0.076	0.037	0.033
2004	0.189	0.046	0.081	0.042	0.022
2005	0.196	0.055	0.071	0.041	0.029
2006	0.207	0.058	0.074	0.051	0.026
2007	0.202	0.056	0.071	0.045	0.033
2008	0.225	0.061	0.101	0.042	0.023
2009	0.187	0.047	0.081	0.035	0.028
2010	0.202	0.055	0.087	0.039	0.031
2011	0.187	0.051	0.076	0.042	0.021
2012	0.223	0.063	0.090	0.042	0.030
2013	0.217	0.058	0.092	0.045	0.024
2014	0.176	0.047	0.064	0.039	0.027
2015	0.214	0.061	0.075	0.049	0.030
All	0.202	0.056	0.079	0.039	0.025
Tennessee					
2007	0.165	0.041	0.065	0.032	0.030
2008	0.191	0.044	0.081	0.031	0.037
2009	0.174	0.042	0.073	0.029	0.030
2010	0.159	0.041	0.066	0.018	0.037
2011	0.185	0.044	0.078	0.034	0.030
2012	0.196	0.048	0.083	0.035	0.033
2013	0.178	0.040	0.076	0.031	0.033
2014	0.185	0.038	0.076	0.035	0.040
2015	0.179	0.039	0.078	0.032	0.033
All	0.179	0.042	0.075	0.031	0.033

Notes: Transfers are principals who move to a different school in the next year but remain a principal. Exits are principals who leave the state education system. Promotions are principals who move to a central office position in the next year. Demotions are principals who move to non-principal, school-based position, such as assistant principal or teacher.

estimates for each school-by-year observation and subsequently standarize them.

Our measure of school climate comes from teachers' responses on a yearly statewide survey in Tennessee, which was first administered in 2011–12. Part of the survey includes items that assess teachers' perceptions of school climate. Examples of items include, "The staff at this school like being here; I would describe us as a satisfied group", and "There is an atmosphere of trust and mutual respect within this school." Using factor analysis, we reduce the full set of responses to a single teacher-by-year score, then average these scores within each school-by-year cell. Details on the factor analysis, including the full set of items, is available in Appendix III.9. Importantly, the survey window runs from early March to the middle of April, which is roughly one month prior to the last day of school. This timing means that many teachers likely complete the survey prior to knowing whether their principal will remain in the school for the next year.

III.3 Methods

III.3.1 Estimating the Effects of Principal Turnover

Our primary analysis seeks to estimate the causal effect of principal turnover on school performance. In the language of potential outcomes, we wish to compare the actual performance of schools that changed their principal in a given year to *what would have happened* in those schools had they kept their principal for at least another year. Given that we cannot observe the true counterfactual for these "treated" schools, we aim to construct a plausible counterfactual using schools that did not change principals in the given year. To be specific, we estimate the following non-parametric difference-in-differences model:

$$Y_{ijkt} = \alpha_1 T_i + \sum_{k=-5,k\neq 0}^{5} D_k \delta_k + \sum_{k=-5,k\neq 0}^{5} (D_k * T_i)\,\beta_k + \gamma S_{it} + \tau_{jt} + \mu_i + \epsilon_{ijkt} \tag{III.1}$$

status, special education status, lagged absences, grade repetition, and whether the student changed schools at least once during the year), school- and grade-level averages of these student characteristics, grade-by-year fixed effects, and teacher fixed effects. After computing the student residuals, we add back the teacher fixed effects and estimate the best linear predictor of a teacher's average student residuals in the current year based on their residuals from prior and future years. The coefficients from this best linear predictor are then used to predict a teacher's value-added in the current year.

where i, j, and t index schools, districts, and school years, respectively, and k is years relative to a principal transition. The parameters of interest are β_k, which are the coefficients on the indicators for years relative to a principal transition (D_k) interacted with the treatment indicator T_i. The omitted category is $k = 0$, which is the final year of the departing principal. Also included in the model are time-varying school characteristics (enrollment size and the proportions of black, Hispanic, and FRPL-eligible students), school fixed effects, and district-by-year fixed effects. School fixed effects account for unobserved, time-invariant school characteristics that are related to both school performance and the likelihood that a principal leaves their position, such as the quality of school facilities or the characteristics of the neighborhood. Inclusion of district-by-year fixed effects is also important, as they account for secular trends and shocks at the district and state level, such as superintendent turnover, changes to human resources policies, or the implementation of a high-stakes evaluation system. Their inclusion also restricts the identifying variation to districts that have multiple schools, which includes 94% and 99% of school-by-year observations in Missouri and Tennessee, respectively. We cluster standard errors at the school level in all models.

For β_k to be an unbiased estimator of the causal effect of principal turnover, we must assume that performance in "treated" schools (i.e., schools that changed principals after year t) would have followed the same trajectory as "comparison" schools (i.e., schools that did not change principals after year t). However, we know from prior work that this assumption is unlikely to hold. Specifically, schools that change principals after year t tend to be on a downward trajectory in terms of student achievement (Miller, 2013). Thus, even conditional on school fixed effects and the other controls in the model, there likely remains substantial endogeneity bias. In more concrete terms, the decline in school performance in treatment schools means that parallel trends do not hold between treatment and comparison schools, which undermines the assumption that the comparison schools constitute a plausible counterfactual for the treatment schools. We address this challenge by constructing a matched comparison group that does not change principals after year t but experiences the same downward trend in student achievement, has similar demographic characteristics, and has a similar history of teacher and principal turnover. Prior work demonstrates that combining matching with a difference-in-differences approach can more successfully mitigate bias than either method on its own (e.g., Mueser et al., 2007).

III.3.2 Constructing a Comparison Group

For schools that change principals, we construct a comparison group by matching to schools from the same state that did not experience a principal transition in the given year. For example, schools in Tennessee that changed principals in 2012 are matched to schools in Tennessee that did not change principals in 2012. The purpose of our matching strategy is to construct a comparison group that meets the parallel trends assumption for our difference-in-differences model. The details of our matching strategy are as follows. First, we estimate a logistic regression model (separately for each state and year) that predicts the probability of having principal transition in the current year as a function of (1) current and lagged (up to five years) school achievement levels in math and reading, (2) current and lagged proportion of new-to-school teachers, (3) binary indicators for principal transitions in each of the prior five years, (4) current and lagged principal experience, and (5) current school demographics. Using the estimated propensity scores from this model, we employ a kernel matching algorithm to construct a comparison group of schools. Additionally, we restrict our matching to the area of common support, which drops a handful of treatment schools in each year (see Table III.A4). While the excluded schools are few in number, they are those with the highest estimated propensity scores (i.e., schools with the lowest achievement levels and highest teacher/principal turnover). Figure III.A11 demonstrates good overlap between treatment and comparison schools after matching. Additionally, Table III.A5 shows that the treatment and comparison groups are similar in terms of observable characteristics.

The key identification assumption of our difference-in-differences model is that treatment schools would have followed the trajectory of comparison schools in the absence of treatment.

Specifically, given the use of school fixed effects, district-by-year fixed effects, and our matching strategy that employs up to five lags in the outcome variable, any potential confounders would need to be (a) time-varying at the school level, (b) correlated with principal turnover in the current year, (c) not fully explained by current and lagged achievement and teacher turnover, and (d) correlated with future school performance. This strategy rules out, for example, bias from principal turnover due to falling school performance or a principal transition induced by superintendent turnover. It may not rule out bias from a change in school inputs that are coincident to the principal transition and potentially improve school performance, such as a district appropriating additional funds for after-school programs or building improvements. Nonetheless, this approach represents a substantial step toward isolating the causal effect of principal turnover.

III.3.3 Modeling Multiple Events

An additional challenge in estimating the effect of a principal transition is that schools can experience multiple principal transitions in a short span. For instance, in both states the mean number of years between principal transitions is 4.2. In Missouri, the modal school changed principals twice across the study period (2001–2015). Due to the shorter time span in Tennessee, the modal school changed principals only once (see Figure III.1). Still, many schools experienced multiple "treatments."

The problem of multiple events has not received much attention in the literature, and there is no commonly accepted method for event studies with multiple events (Sandler and Sandler, 2014). In the study most relevant to ours, Miller (2013) notes that many of the schools in her data experienced multiple principal transitions. Her approach was effectively to treat the principal transition as the unit of observation and arrange the data to have one physical observation per event per time period (i.e., "stacking" the data). For example, if a school changed principals in 2010 and again 2012, both "treatments" are modeled by including the school twice. In this hypothetical school, outcomes in 2013 represent both three years after a transition (in reference to 2010) and one year after a transition (in reference to 2012). Using evidence from Monte Carlo simulations, Sandler and Sandler (2014) suggest a different approach. Specifically, they propose to modify the standard event study approach, which includes a full set of mutually exclusive dummy variables for event time, to allow multiple indicators to be "turned on" at one time. For instance, in the aforementioned example of a school in 2013 with principal transitions in 2010 and 2012, each school-by-year observation is included only once in the model, but the indicators for one year since turnover and three years since turnover are both set to one.

We estimate models that employ both of these approaches. For our difference-in-differences model, we follow the "stacking" approach used by Miller. Here, it is necessary to treat the principal transition as the unit of observation because of our matching strategy that constructs a comparison group for each specific turnover event. Appendix III.8 demonstrates via simulation that our stacking approach yields reasonable estimates of the impact of turnover. However, we also use the approach suggested in Sandler and Sandler (2014) to estimate event study models that show descriptively the dynamics of school performance before and after a principal transition.

III.3.4 Examining Different Types of Principal Turnover

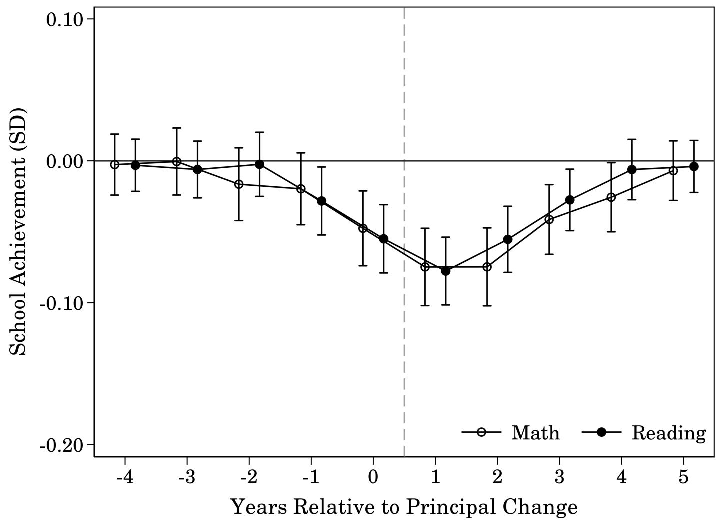
In addition to the average principal transition, we also seek to estimate the effect of different *types* of principal turnover on school performance. Here, we again estimate a difference-in-differences model, but we also construct a new comparison group for each type of turnover. Re-matching is important because different types of principal transitions likely reflect different circumstances in a school in the years leading up to turnover. For example, relative to the average principal transition, the pre-transition decline in performance is substantially steeper for demotions, where requires us to construct a comparison group that is more heavily weighted by lower-performing schools. We show the matching details for each type of principal turnover in Appendix III.7.

III.4 Results

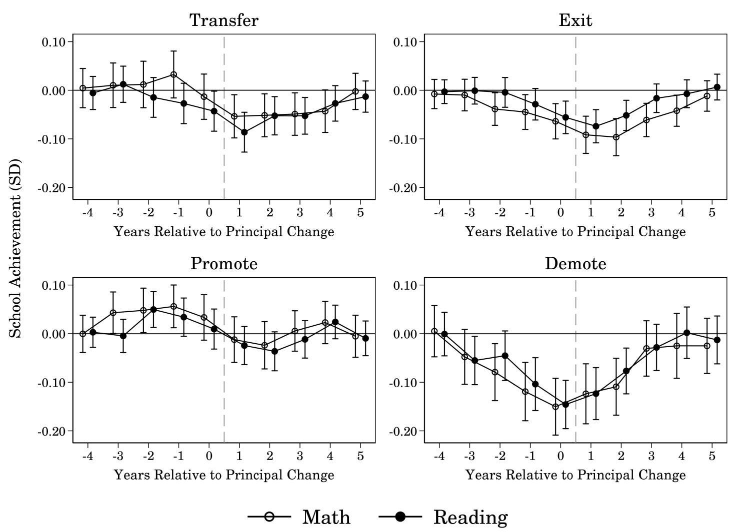
III.4.1 The Dynamics of School Achievement and Principal Turnover

We begin by showing the dynamics of school performance before and after a principal transition, which serves to replicate prior work and build the motivation for our identification strategy. In Figure III.2, panel A shows school-level math and reading achievement for the five years before and after any type of principal turnover. Panel B disaggregates turnover into four types: transfers, exits, promotions, and demotions. In each plot, the unit of observation is school-by-year and overlapping turnover events are captured by allowing multiple indicators to be "turned on" (e.g., a given school-by-year could be two years before their next turnover event and one year after their previous turnover event). Panel A replicates the finding from Miller (2013) that school achievement declines in the years prior to a principal transition but rebounds fairly quickly. Specifically, math and reading achievement begin to drop two years prior to a transition and continue to decline until the second year after the transition. By the fourth year after a turnover event, achievement returns to pre-transition levels, on average.

Figure III.2 panel B demonstrates substantial heterogeneity in the patterns of school achievement by the *type* of principal turnover. For instance, the pre-turnover dip shown in panel A is largest for demotions, where school performance declines by roughly 0.15 SD in the four years prior to the principal transition. By contrast, schools where the principal receives promotion to central office are on a slight upward trajectory prior to the transition. Exits, which constitute the greatest share of principal turnover, follow a pattern similar to the average transition, while transfers have perhaps a small pre-transition dip. Except in the case of demotions, school achievement continues to decline in the first year of the new principal, before rising in the second or third year after the transition. In the case of demotions, however, achievement is lowest in the final year of the departing principal and increases steadily after the transition. For all types of principal turnover, achievement returns to "normal" (i.e., the school's average achievement in the years not surrounding a principal transition) by the



(a) All Turnover



(b) Turnover Types

Figure III.2: School Achievement Before and After Principal Turnover

Notes: Panel A shows the estimated coefficients from regression models that predict school achievement in math/reading as a function of prior and future principal turnover events. Panel B disaggregates the binary turnover indicators into indicators for each type of turnover. The coefficients by type are estimated from a single regression model but shown in separate plots for clarity. Regressions in both panels include school fixed effects, district-by-year fixed effects, and time-varying school demographic characteristics. Schools from both states are pooled in a single regression to increase precision. Error bars show 95% confidence intervals.

fifth year after a principal turnover event, on average. Importantly, turnover type is likely a proxy for the circumstances in the school leading up to the transition or the quality of the departing and replacement principals. For instance, exits may reflect a declining school climate prior to the transition, whereas demotions may reflect district efforts to replace an ineffective leader with (presumably) a more effective successor.

Figure III.2 illustrates two important points. First, as documented in prior work, there is a fairly substantial pre-turnover dip prior to a principal turnover event, which presents a challenge for causal inference. Second, the dynamics of principal turnover—a pre-transition decline followed by a rebound under the new principal—vary by the type of principal turnover. These descriptive results by pathway suggest that estimates of the effect of the average principal turnover event may mask substantial heterogeneity in the effect by turnover type. The next part of our analysis addresses both of these points.

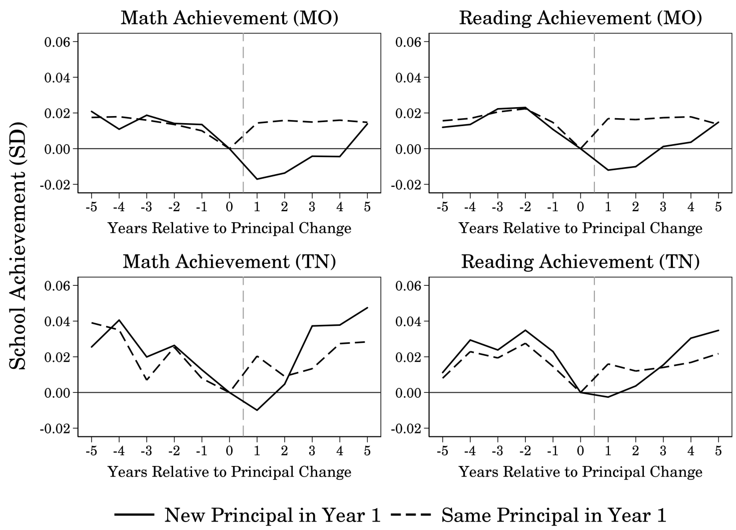
III.4.2 The Causal Effect of Principal Turnover

Next, we estimate the causal effect of principal turnover using a difference-in-differences model with a matched comparison group. We begin by examining a binary turnover outcome. Figure III.3 demonstrates our approach graphically. We plot marginal predictions from our difference-in-differences model for school achievement in math and reading (and teacher retention in Panel B) in our "treatment" schools (i.e., those who change principals between year 0 and year 1) and "comparison" schools. This plot shows that, after matching, these groups follow the same downward trajectory in school performance in the pre-treatment period, consistent with the idea that we have constructed a plausible counterfactual. The difference in outcomes between the treatment and comparison groups in years 1 through 5 isolates the effect of changing principals. Panel A shows that, in both Missouri and Tennessee, schools that changed principals had lower math and reading achievement in the first and second years after the transition. In both states, schools that keep the same principal between year 0 and 1 experience an immediate rebound in performance in year 1. Schools that change principals also rebound, but only after one or more additional years of depressed performance. In Panel B, we find that there is no obvious trend in teacher retention prior to a principal turnover event. However, there is a clear drop in teacher retention for schools that change principals between year 0 and year 1 that is not experienced by schools that keep their principal.

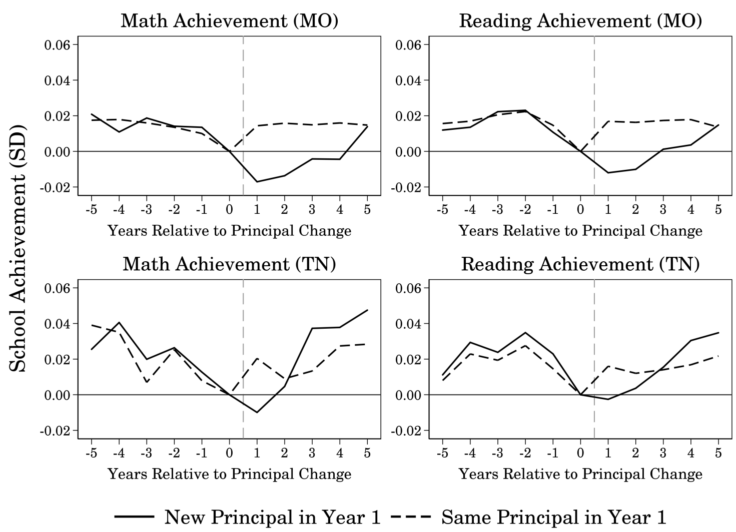
Figure III.4 plots the effect estimates from our difference-in-differences models (these results are also shown in Table III.A1). We estimate separate models by state for four outcomes: school achievement in math and reading, teacher retention rate (i.e., the proportion of last year's teachers who are still in the school in the current year), and the number of new-to-school teachers. Retention and new-to-school teachers are closely linked, but showing the count variable facilitates interpretation of the magnitude of the effect of principal turnover on teacher turnover. In each model, the omitted category is the final year of the departing principal. Statistically significant estimates in the pre-transition period (time $<$ 0) would suggest a violation of the parallel trends assumption, which would undermine the credibility of the post-transition coefficients (time $>$ 0) as causal estimates of the effect of principal turnover. Across the models, the pre-transition estimates are almost uniformly small in magnitude and not statistically significant, indicating that our matching procedure was successful.

The results in Figure III.4 show that principal turnover negatively affects math and reading achievement in both Missouri and Tennessee. The effect is largest in the first year after the transition. In Missouri (Tennessee), changing principals lowers math achievement by -0.031 (-0.030) school-level standard deviations in the first post-transition year. The first-year effects on reading achievement are -0.029 and -0.019 SD in Missouri and Tennessee, respectively. These estimates correspond to roughly 7% of the typical amount of variation in a school's achievement scores (i.e., the within-school standard deviation) over the study period in both states—similar to the magnitudes found by Miller (2013).[6] We can also

[6]Specifically, we compute the the within-school standard deviation in math/reading achievement across the study period, then compare the magnitude of the coefficients from Table III.A1 to the within-school



(a) All Turnover



(b) Turnover Types

Figure III.3: The Effect of Principal Turnover in Missouri and Tennessee

Notes: Each plot shows the estimated margins for treatment and comparison schools relative to year of a principal change (year 0). The model coefficients are shown in Table III.A1.

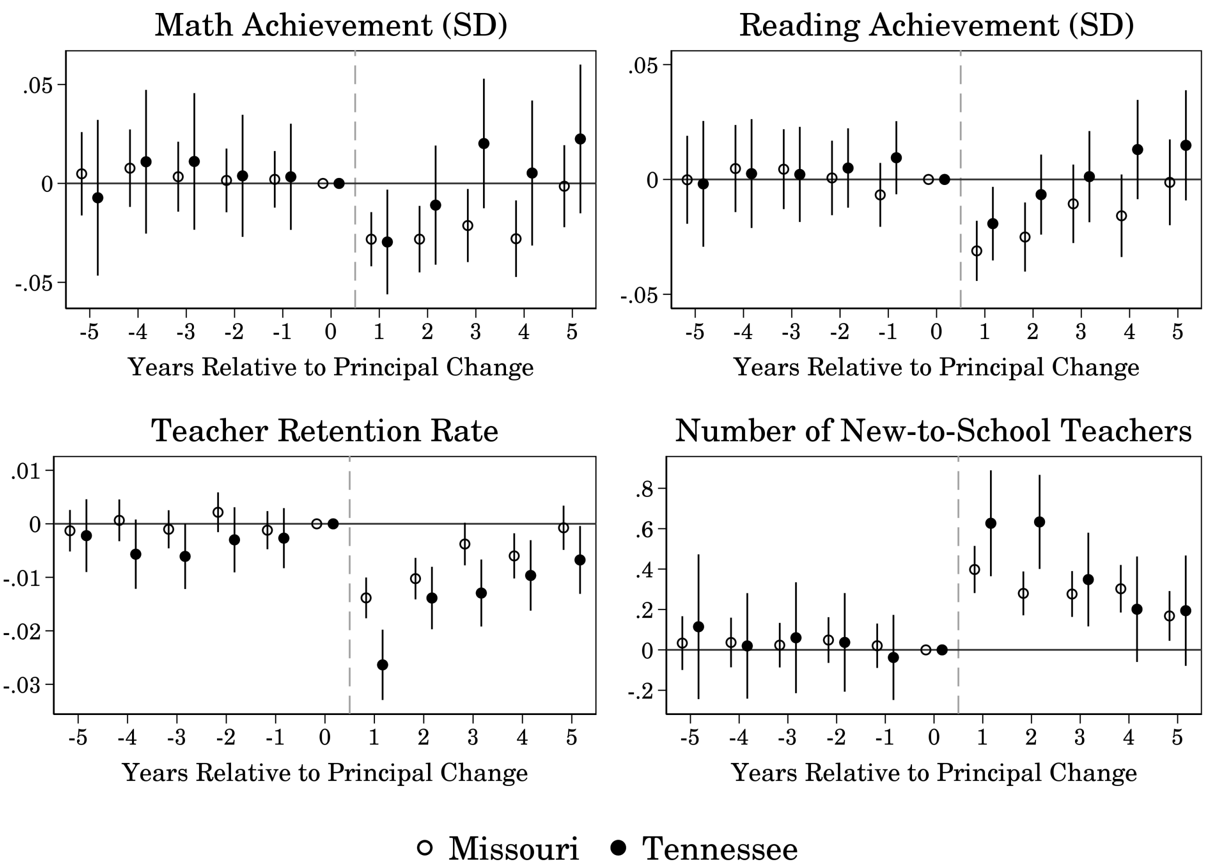


Figure III.4: The Effect of Principal Turnover in Missouri and Tennessee

Notes: Each plot shows the focal coefficients (Treat x Time) from difference-in-differences models shown in Table III.A1. Error bars show 95% confidence intervals.

roughly translate these effects to student-level standard deviations in Tennessee: a principal change lowers the average student's math (reading) score by 0.012 (0.007) SD.[7]

How do the first-year effects of principal turnover compare to the effects of teacher turnover? Estimating models at the school-year-grade level, Ronfeldt et al. (2013) find that 100% turnover lowers achievement in math and reading by 0.082 and 0.049 SD for all students in the grade, relative to no teacher turnover. Although the negative effects of principal turnover estimated here are smaller for the individual student, they are felt by all students in the school. For instance, the average school in Missouri (Tennessee) has a total enrollment of 408 (620) students compared to 100 (111) in the average grade level. Even small effects on individual students, then, produce a policy-relevant total impact when aggregated across an entire school.

After the first year, the estimated effects of principal turnover in Missouri and Tennessee differ somewhat. In Missouri, there is a consistent negative effect on math achievement up to four years after a principal transition. For reading, there is a significant negative effect in the second year, but while the third and fourth year estimates are negative, they are not statistically significant at conventional levels. In Tennessee, the negative effect of principal turnover is only significant in the first year. However, it is important to note that the composition of turnover in terms of types is different between these two states. In particular, a larger share of principal transitions in Tennessee are demotions, whereas promotions and transfers are more frequent in Missouri. To the extent that the effect of principal turnover may vary by type (which we address in the next section), we would expect the effect of the average turnover event in each state to differ.

Principal turnover also decreases teacher retention and, as a consequence, the number of teachers who are new to the school in the years after the principal transition. For example,

---

standard deviation of the median school in each state. These median within-school standard deviations for math achievement in Missouri and Tennessee are 0.44 and 0.41, respectively (0.39 and 0.25 for reading).

[7] We make this translation by multiplying the first-year coefficients by the ratio of the standard deviation of school-average scores to the standard deviation of individual student scores. These ratios are 0.40 and 0.38 for math and reading, respectively.

columns 3 and 7 in Table III.A1 show that the proportion of teachers retained from the previous year decreases by 1.4 percentage points and 2.6 percentage points in Missouri and Tennessee, respectively. These negative effects on retention continue for several years after a principal transition. In Tennessee, for example, schools that changed principals (between year t and $t+1$) have lower teacher retention than comparison schools even five years after the transition. How large are these negative effects on teacher retention? In columns 4 and 8, we show the effects of principal turnover on the *number* of new-to-school teachers. Overall, the magnitude of these effects are modest. For instance, the first-year impact on retention translates to roughly half of an additional new teacher. Summing across the five years after a transition, the average principal turnover event increases the number of new-to-school teachers by 1.1 in Missouri and 1.9 in Tennessee. Comparing Missouri and Tennessee, one interesting note is that while the effects on achievement were smaller in Tennessee, the effects on teacher turnover are larger. Again, these differences could in part reflect differences in the composition of turnover types between the two states.[8]

III.4.3 Do Different Types of Principal Turnover Have Different Effects?

The previous section establishes that the average principal turnover event has a negative effect on school performance. Specifically, there is a modest negative average effect on school achievement in the short-term and a longer-lasting, though still modest, negative effect on teacher retention. Next, we consider whether these effects vary by the type of principal turnover event. Specifically, we categorize principal turnover into four types: transfers, exits, promotions, and demotions. For each type, we employ our matching procedure to obtain comparison schools that match treatment schools on observable characteristics in the pre-transition period (see Appendix III.7). To increase precision, we estimate models that

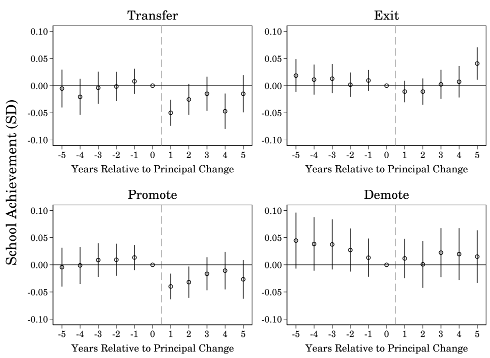
[8] We also considered whether *student* mobility may increase in response to principal turnover. For instance, families might respond to a principal transition by moving their child to a school with a more experienced leader. Using our student-level data from Tennessee, however, we found no evidence that within-year or end-of-year transfers were responsive to principal turnover. Both the pre- and post-transition trends were flat in treatment and comparison schools and difference-in-differences estimates were precise zeroes. Results available on request.

pool the data from both states.

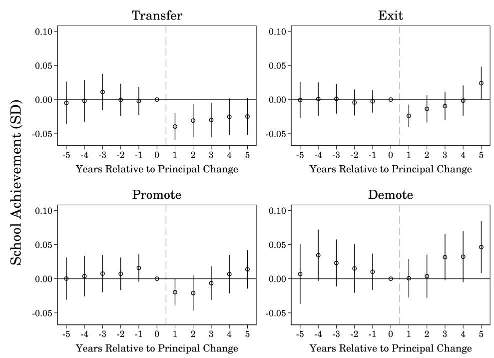
We show the results for each outcome in a series of coefficient plots displayed in Figure III.5, with the full results in Tables III.A2 and III.A3. Panels A and B show the results for math and reading achievement. In both subjects, the first-year effect of principal turnover is negative for transfers, exits (though only statistically significant for reading), and promotions. Principal transfers lower school achievement in math and reading by 0.050 SD and 0.040 SD in the first year after the transition. For exits and promotions, the first-year effects for math are -0.011 SD (ns) and -0.040 SD, respectively. For reading, these effects are -0.024 and -0.020, respectively. By contrast, there is no evidence of an effect for demotions—the coefficient for year 1 is close to zero in both subjects and not statistically significant.

Beyond the first year, there is some evidence of sustained negative effects for transfers and promotions. For instance, reading achievement is significantly lower in schools that experience a principal transfer even five years after the transition, though we note that the year 4 and 5 estimates are on the margins of statistical significance ($p < 0.10$). The effects on math achievement, though less consistent in terms of magnitude and significance, follow a similar pattern. For promotions, the negative effects in math and reading achievement persist until at least the third year after the transition. By contrast, the negative effects of a principal exit appears to dissipate quickly, and there is some suggestive evidence of a *positive* effect in both subjects in the fifth year after the transition. Five years after an exit, math achievement is 0.041 SD higher ($p < 0.01$) and reading achievement is 0.024 SD higher ($p < 0.10$). For demotions, there is some suggestive evidence of an upward trend in achievement after the transition, particularly for reading. The coefficients for years 3–5 after a demotion are positive in both subjects but have large standard errors. Nevertheless, we find a positive effect of demotions on reading achievement (0.046 SD, $p < 0.05$) in the fifth year after the transition.

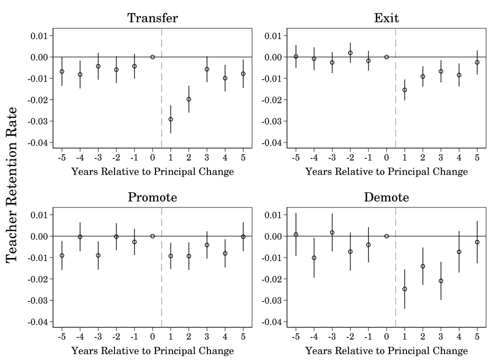
Panels C and D in Figure III.5 show the effects of different types of principal turnover on teacher retention rates and the number of new-to-school teachers (full results shown in Table



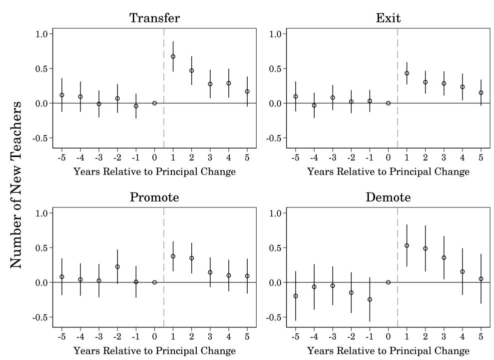
(a) Math Achievement



(b) Reading Achievement



(c) Teacher Retention



(d) New-to-School Teachers

Figure III.5: Results by Turnover Type

Notes: Each plot shows the focal coefficients (Treat x Time) from a difference-in-differences model for the listed turnover type. Full model results are shown in Tables III.A2 and III.A3. Matching details are shown in Appendix A. Error bars show 95% confidence intervals.

III.A3). Across all types of principal turnover, there are clear negative effects on teacher retention, which (by construction) increases the number of new-to-school teachers in the post-transition years. Similar to the achievement results, these negative effects on teacher retention are largest for transfers. However, one important difference between the results for achievement and teacher retention is that while demotions do not lower achievement, they do decrease teacher retention. Furthermore, in terms of the number of new-to-school teachers who enter the school following a principal transition, the increase is the *largest* for demotions. While these results may seem incompatible given the documented connection between teacher turnover and student achievement, one possibility is that there is no change in the quality of the school's teaching staff, either because the departing teachers are systematically less effective or their replacements (newly hired teachers) are equally effective.

III.4.4 How Does Principal Turnover Harm School Performance?

The remainder of our analysis focuses on exploring the potential explanations for the negative effects of principal turnover that we estimated in the previous sections. We begin by examining two potential mechanisms: teacher retention and principal experience. While we previously conceptualized teacher retention as an indicator of school performance, one explanation for our achievements results is that the negative effects of principal turnover are driven by the subsequent spike in teacher turnover (Ronfeldt et al., 2013). We also might expect that principal turnover harms school performance through the loss of an experienced leader. Principal experience is positively linked to student growth (Clark et al., 2009), and a large percentage of new-to-school principals have little or no prior principal experience (Grissom et al., 2019). To this point, across the study period in Missouri and Tennessee, approximately 60% of new-to-school principals had no prior experience as a principal. To probe the extent to which these proximal outcomes may explain the achievement effects, we include them as covariates in the achievement models and compare the estimated coefficients for the impact of principal turnover to the baseline model. Attenuation in the estimated effects of principal turnover when controlling for teacher retention or principal experience would provide suggestive evidence that these mechanisms explain the connection between principal turnover and student achievement.

Columns 1 and 4 in Table III.2 show the baseline results (pooled across both states) for math and reading achievement in columns 1 and 4. Columns 2 and 5 control for the teacher retention rate, while columns 3 and 6 control for years of prior experience of the principal who is currently leading the school. Comparing columns 1 and 4 to columns 2 and 5, we find that controlling for the teacher retention rate slightly attenuates the magnitude of the negative effects of principal turnover. There is a clear positive relationship between teacher retention and school achievement, but the drop in retention induced by a principal leaving the school explains very little of the negative effect of principal turnover on school achievement. The reason is simply that the negative effects of principal turnover on teacher retention, while estimated precisely, are not very large. A back-of-the-envelope calculation multiplying the first-year impact on teacher retention from Table III.A1 (-0.014 in Missouri and -0.026 in Tennessee) by the teacher retention coefficient (0.227) for math in column 2 (i.e., the predicted SD increase in school-level math achievement for a 100 percentage point increase in teacher retention) implies that the negative impact of principal turnover on teacher retention lowers math achievement by 0.003 SD and 0.006 SD in Missouri and Tennessee, respectively. These approximations correspond closely to the actual change in the first-year effect (0.004 SD) when we compare column 1 (baseline model) and column 2 (controlling for teacher retention) in Table III.2, which pools both states.[9]

Columns 3 and 6 control for the experience level (i.e., prior years of principal experience) of the current principal. The coefficients at the bottom of the table show clear positive returns to having an experienced principal.[10] Furthermore, we find that conditioning on

[9]Performing this analysis of potential mechanisms separately by state produces the same patterns shown in Table III.2.

[10]The relationship between principal experience and achievement can be identified using within-school variation even when including indicators for event time because schools in both the treatment and comparison groups have multiple principals in the pre- and post-transition periods, which produces variation in principal experience that is not correlated with event time.

Table III.2: What Mechanisms Explain the Effect of Principal Turnover?

	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
Time = 1 x Treat	-0.031***	-0.027***	0.002	-0.026***	-0.021***	0.001
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Time = 2 x Treat	-0.022***	-0.020***	-0.000	-0.021***	-0.018***	-0.002
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Time = 3 x Treat	-0.007	-0.006	0.009	-0.011*	-0.010	0.000
	(0.008)	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)
Time = 4 x Treat	-0.012	-0.011	-0.003	-0.007	-0.005	0.001
	(0.009)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)
Time = 5 x Treat	0.005	0.005	0.011	0.004	0.006	0.006
	(0.009)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)
Teacher Retention		0.227***			0.263***	
		(0.036)			(0.032)	
Principal Prior Experience						
1 year			0.016*			0.020***
			(0.009)			(0.007)
2 year			0.036***			0.037***
			(0.010)			(0.008)
3 year			0.053***			0.041***
			(0.011)			(0.009)
4 year			0.050***			0.056***
			(0.012)			(0.011)
5+ years			0.075***			0.066***
			(0.011)			(0.009)
R^2	0.818	0.819	0.825	0.865	0.866	0.873

Notes: Standard errors clustered by school shown in parentheses. Models include school characteristics, school fixed effects, and district-by-year fixed effects. Schools from Missouri and Tennessee are included in a single regression to increase precision. * p < 0.1, ** p < 0.05, *** p < 0.01.

principal experience makes the estimated effect of principal turnover zero. This finding is consistent with a conclusion that the loss of leadership experience drives turnover's negative effects.

If experience is a proxy for effectiveness, and a turnover event on average represents a net loss in effectiveness, then the effect of turnover will vary by the effectiveness of the outgoing principal. We examine this possibility in Table III.3. Specifically, we draw on Tennessee data to document within-school changes in principal quality before and after instances of principal turnover. Here, our models are descriptive in the sense that we do not attempt to estimate the causal effect of principal transitions on principal quality in the school. Instead, we use these models to provide support for our main results. Each model predicts the given measure as a function of indicators for time before or after a principal turnover event, similar to the specification used in Figure III.2. However, we cannot include as many leads and lags because of the fewer years of available data in Tennessee and the fact that some of our measures are only available beginning in 2011–12. We include indicators for the two years before and after a principal transition, along with school and district-by-year fixed effects. As with our main models, we show results for both an aggregated principal turnover measure and for each type of transition. In all cases, the reference group is years outside of the four-year window around a principal transition (i.e., periods of leadership stability in the school).

Our measure of principal quality is rubric-based ratings from supervisors, which all principals in Tennessee receive as part of the state's educator evaluation system first implemented in 2011–12. Prior work establishes that these ratings are internally consistent and correlated with student growth in math and reading (Grissom et al., 2018). Column 1 shows that, on average, a school's principal receives lower ratings (-0.17 SD) in the final year before a transition, which is consistent with the achievement results showing a "pre-transition" dip in performance. After a transition, principal ratings further decline to -0.37 SD before partially recovering to -0.16 SD in the second year after the transition. However, column 2 shows

Table III.3: Principals' Ratings from Supervisors Before and After Principal Turnover

	(1)	(2)
All Principal Turnover		
One Year Before	0.012	
	(0.027)	
Current Year	-0.167***	
	(0.035)	
One Year After	-0.371***	
	(0.028)	
Two Years After	-0.156***	
	(0.025)	
Principal Transfer		
One Year Before		0.143**
		(0.062)
Current Year		0.062
		(0.063)
One Year After		-0.362***
		(0.051)
Two Years After		-0.153***
		(0.045)
Principal Exit		
One Year Before		-0.011
		(0.040)
Current Year		-0.215***
		(0.059)
One Year After		-0.363***
		(0.040)
Two Years After		-0.167***
		(0.034)
Principal Promotion		
One Year Before		0.144***
		(0.051)
Current Year		0.038
		(0.066)
One Year After		-0.448***
		(0.053)
Two Years After		-0.182***
		(0.052)
Principal Demotion		
One Year Before		-0.242***
		(0.057)
Current Year		-0.632***
		(0.075)
One Year After		-0.267***
		(0.071)
Two Years After		-0.105
		(0.064)
N	7113	7113
R^2	0.758	0.766

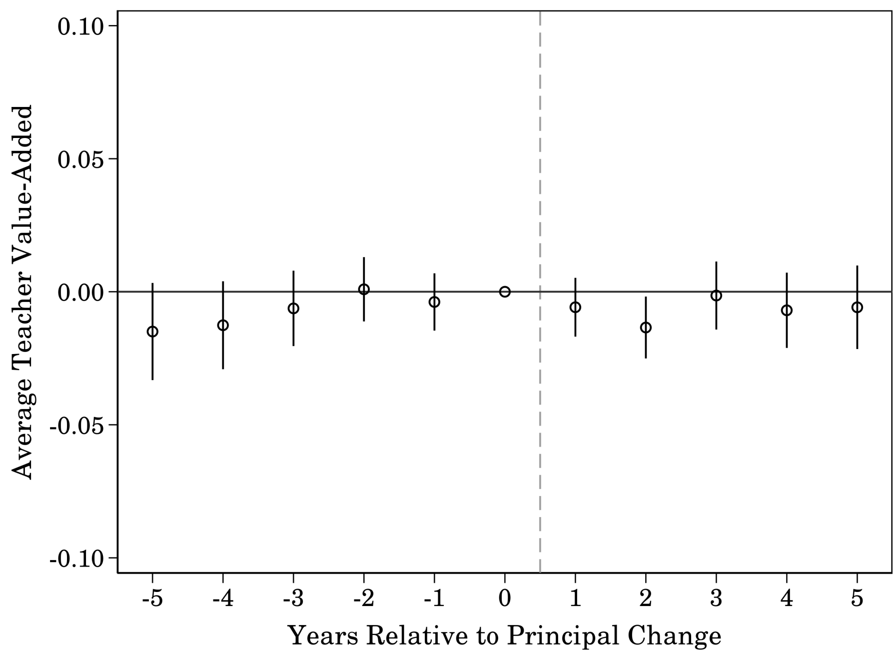
Notes: Standard errors clustered by school shown in parentheses. Models include school fixed effects, district-by-year fixed effects, and school demographic controls. "Current year" is the year of the principal turnover event (i.e., the principal leaves between year t and year $t+1$.). Supervisor ratings are calculated as the average across all rubric items and standardized within year.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

substantial heterogeneity by turnover type. For instance, in the case of transfers and promotions, the school's principal is rated slightly higher than average (relative to that school) in the years preceding the transition, while the new principal receives substantially lower scores in their first year in the school (-0.36 SD for transfers and -0.45 SD for promotions). In contrast, principal demotions are preceded by a precipitous drop in supervisor ratings in the two years prior to the transition (-0.24 and -0.63 SD). While the new principal scores much higher than their predecessor's final year, their rating is still below average in the first year after the transition (-0.27 SD). Finally, principals who exit the education system receive below-average ratings in their final year (-0.22 SD), with the new principal scoring below average in the first (-0.36 SD) and second years after the transition (-0.17 SD).

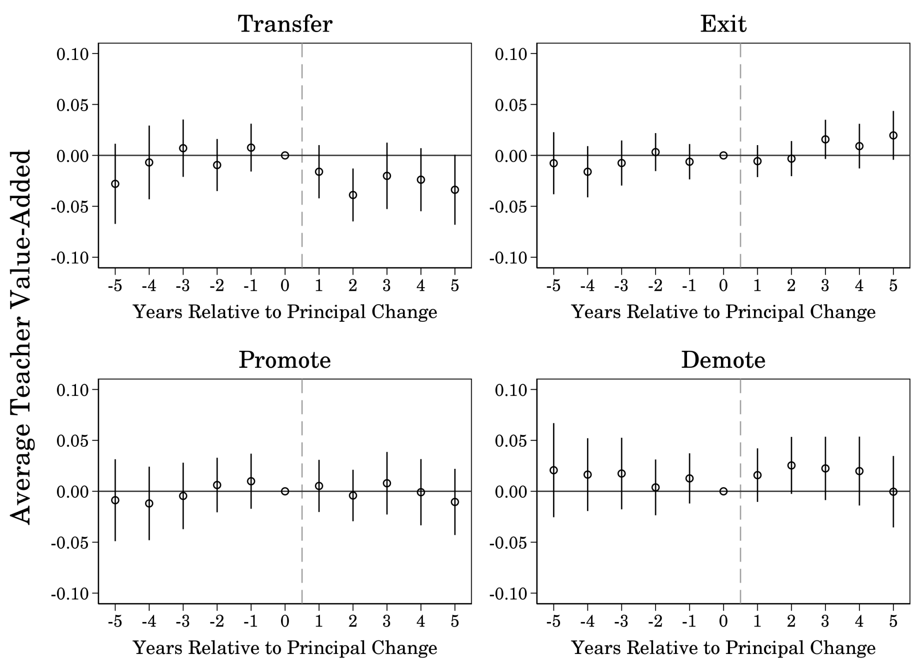
This dip in a measure of principal effectiveness is informative but something of a black box. To shed some additional light, we examine two potential avenues through which principals can impact schools: effects on teacher quality and effects on school climate. For instance, principals can increase the quality of their teaching staff by hiring more effective teachers (Loeb et al., 2012) or strategically targeting their retention efforts towards high-performing teachers (Grissom and Bartanen, 2019b). We examine changes in teacher quality (among Tennessee teachers in tested classrooms) in Figure III.6. Specifically, we estimate teacher VA using the leave-year-out, drift-adjusted approach described in Chetty et al. (2014). To produce a school-level measure, we first standardize the teacher VA estimates within subject and year and then compute the school-by-year mean. We then estimate our main difference-in-differences models with this school-by-year mean as the outcome variable.

Figure III.6 panel A shows that, on average, principal turnover has little to no effect on the average quality of the school's teaching staff. While we find a negative and statistically significant coefficient (-0.013 SD) in the second year after a principal transition, the magnitude is very small.[11] Turning to panel B, we find some evidence that transfers lead to a

[11] Benchmarked against the estimated variance of teacher effects on student achievement in Tennessee (0.14 SD in math and 0.08 SD in reading), this decrease in teacher quality would lower student scores by roughly 0.002 and 0.001 student-level standard deviations).



(a) All Turnover



(b) Turnover Types

Figure III.6: Impact of Principal Turnover on Teacher Quality in Tennessee

Notes: Each plot shows the focal coefficients (Treat x Time) from a difference-in-differences model for the listed turnover type. Average teacher value-added is expressed in teacher-level standard deviations (i.e., a value of 0.10 would indicate that the average teacher in the school is 0.10 SD above the mean in the statewide distribution of value-added in Tennessee.).

small temporary drop in average teacher quality, whereas the results for exits, promotions, and demotions show no evidence of effects.

In Appendix Figures III.A7–III.A10, we examine effects of principal turnover on quality of newly hired teachers and the retention of high-, middle- and low-VA teachers. Similar to Figure III.6, we find no changes in the quality of new hires in response to principal turnover. The exception is for demotions, where new hire VA increases by roughly 0.15 SD in the first two years after the transition. For retention rates, we find negative first-year effects for all teachers, but the subsequent effects are driven by low-VA teachers, which could in part explain why achievement rebounds relatively quickly in schools that change principals in Tennessee. For high-VA teachers, the negative effect on retention is largest when principals transfer across schools, whereas retention effects for low-VA teachers are largest for demotions, though these estimates are relatively imprecise.

Table III.4 examines changes in school climate before and after principal turnover. As part of yearly statewide survey of teachers, Tennessee includes items that assess teachers' perceptions of school climate, which we use to construct a standardized school-level measure.[12] Column 1 shows that there is a decline in school climate prior to principal turnover. For instance, in the final year of the departing principal teachers' perceptions of climate are -0.11 SD lower than during periods of stability. Under the new principal, climate remains depressed for at least one additional year. As with our previous analyses, we find that this average relationship masks heterogeneity by turnover type. In the case of transfers and promotions, we find no statistically significant change in climate before or after changing principals, though the estimates are somewhat imprecise. For exits, climate is significantly lower in the two years preceding the transfer, but significantly *higher* in the first year after the transition ($p < 0.10$). The clearest pattern is for demotions. Teacher perceptions of climate are substantially lower in the two years preceding a demotion. In the year of the demotion, for instance, climate is 0.47 SD lower. However, there is an immediate recovery

[12]Details on the construction of the climate measure are available in Appendix III.9.

to normal levels after a principal change.[13]

III.5 Discussion and Conclusions

Increasing recognition of the importance of school leaders for student outcomes and high rates of principal mobility highlight the need to understand the causes and consequences of principal turnover. Although a handful of existing studies document a negative correlation between principal transitions and school performance, this study fills an important gap by isolating the (arguably) causal effect of changing principals. Using a combination of matching and difference-in-differences, we compare schools that changed principals to schools with similar trajectories that kept their principal for at least one additional year. Additionally, we estimate the effects of different *types* of principal turnover, which prior studies have failed to consider. Finally, we probe various mechanisms that could explain the effects of principal turnover on school performance, including the composition of the teaching staff, school climate, and the effectiveness of departing and replacement principals.

Our results demonstrate that, on average, principal turnover has a negative effect on school performance. Compared to their performance had they kept the same principal, schools that change principals have lower achievement in math and reading and higher rates of teacher turnover. The magnitudes of the achievement effects are modest; the first-year effects for both math and reading are roughly 7% of the typical amount of variation in achievement scores within a typical school in Missouri and Tennessee.

Given convincing evidence that teacher turnover harms student achievement, a natural question is whether the impact of principal turnover on teacher turnover explains the negative effects on achievement. Our findings, however, suggest that the spike in teacher turnover explains only a fraction of the negative effects of principal turnover on student achievement. Further, we find that principal transitions, on average, have little impact on the quality of

[13] We also examined the possibility that principal turnover is associated with changes in survey response, which could confound the patterns in Table III.4. We do find evidence that school-level response rates drop in the final year of a departing principal, but the magnitude is small (1 percentage point). We also estimate school climate models that control for school-level response rates and find virtually identical results to those in Table III.4. See Appendix III.9 for details.

Table III.4: School Climate Before and After Principal Turnover

	(1)	(2)
All Principal Turnover		
One Year Before	-0.050	
	(0.042)	
Current Year	-0.113***	
	(0.043)	
One Year After	0.050	
	(0.040)	
Two Years After	-0.011	
	(0.037)	
Principal Transfer		
One Year Before		0.117
		(0.089)
Current Year		0.013
		(0.086)
One Year After		0.047
		(0.078)
Two Years After		-0.022
		(0.074)
Principal Exit		
One Year Before		-0.111*
		(0.057)
Current Year		-0.130**
		(0.057)
One Year After		0.099*
		(0.058)
Two Years After		0.013
		(0.051)
Principal Promotion		
One Year Before		0.035
		(0.086)
Current Year		0.102
		(0.087)
One Year After		-0.021
		(0.090)
Two Years After		-0.079
		(0.081)
Principal Demotion		
One Year Before		-0.190**
		(0.086)
Current Year		-0.473***
		(0.103)
One Year After		0.053
		(0.086)
Two Years After		0.042
		(0.083)
N	6508	6508
R^2	0.552	0.558

Notes: Standard errors clustered by school shown in parentheses. Models include school fixed effects, district-by-year fixed effects, and school demographic controls. "Current year" is the year of the principal turnover event (i.e., the principal leaves between year t and year $t+1$.) Details on the factor analysis that produces the climate score is available in Appendix Table III.9.
* p < 0.1, ** p < 0.05, *** p < 0.01.

a school's teaching staff as measured by value-added. Instead, a major driver is principal quality. Specifically, replacement principals have less experience than their predecessors, and accounting for the positive relationship between principal experience and student achievement drives our effect estimates to zero.

We also find that the average effect of principal turnover masks substantial heterogeneity by the type of principal transition. The negative effects on achievement are largest when principals transfer to a different school or are promoted to a central office position. In contrast, when principals exit from the state education system, schools experience a small short-term decline in performance but show evidence of a positive trajectory in the medium term. Further, schools where principals are demoted to a lower school-level position (e.g., assistant principal) are not negatively affected and may actually benefit from changing principals. Interpreting these differences requires caution, however, as it seems unlikely that student outcomes are due to the type of principal transition per se. Instead, transition type may proxy for unobserved conditions in the school prior to the switch or the relative quality of the departing and replacement principal.

These results have implications for both policy and research. Perhaps most importantly, while districts should seek to limit principal turnover in general, this study highlights that not all principal turnover is harmful to school performance. While policymakers rightfully worry about the disruptive effects of leadership instability, our findings imply that, in some cases, the benefits of replacing a low-performing principal outweigh these costs. To the extent that district leaders have an adequate supply of potential replacement principals, moving ineffective principals out of struggling schools could help spur improvement. Of course, there is no guarantee that replacements will be more effective, but our results for demotions (which constitute 10–15% of principal turnover) demonstrate that districts in Missouri and Tennessee have seen some success in moving (likely) struggling principals to other positions, a finding that complements recent work showing that a principal dismissal policy led to higher student achievement growth in DC Public Schools (Walsh and Dotter, 2019).

On the other hand, districts should weigh the perceived benefits of principal transfers and promotions with the costs of a school losing an effective leader. Although we cannot presume that control over principals' mobility decisions lies completely with district leaders, existing work suggests that they do play an important role (Farley-Ripple et al., 2012). In particular, the finding that principal transfers are associated with lower retention of highly effective teachers highlights that promoting the stability of (effective) school leadership in disadvantaged schools matters for maintaining an effective teaching staff. At a minimum, our results caution against policies of simply rotating successful principals among schools, as there appear to be short-term costs to changing principals. Of course, the benefits of some rotation policies (e.g., moving an experienced principal into a particularly hard-to-staff school) could potentially still outweigh these costs.

This study faces several limitations. First, as with any difference-in-differences design, our interpretation of the results as causal rests on the (untestable) assumption that the post-transition trend in our comparison group serves as a valid counterfactual for what would have happened to treatment schools had they kept their principal. While employing a matched comparison group helps to rule certain threats (e.g., regression to the mean), we cannot be certain that schools that change principals are not receiving additional "treatments" in the post-transition period. For instance, districts may seek to drive improvement in low-performing schools by strategically allocating additional resources *in addition to* replacing the principal. To the extent that comparison schools (i.e., schools that are equally low-performing but keep their principal) do not receive such supports, we could potentially understate the negative effects of principal turnover. Further, this potential source of bias may increase when examining specific types of turnover, such as demotions, where the transition is more likely to be initiated by district administrators.

Second, restricting to the area of common support in our matching strategy—necessary to prevent unwarranted extrapolation—drops from the analysis schools with the highest probabilities of principal turnover. While the number of excluded schools are small in number

(1.6% in Missouri and 0.6% in Tennessee), we acknowledge that our results should not be generalized to such schools. Finally, our analysis does not address explicitly the potential consequences of principal "churn"—multiple turnover events within a short span. Here, we have treated multiple events as a modeling challenge, but frequent principal turnover may have consequences above and beyond the individual effects of each event, a useful question for future work to investigate.

We have probed some of the potential mechanisms that drive the effects of principal turnover, but we are also limited by what we can observe in our administrative data sets. We need more information about the processes that occur inside schools during years of leadership instability. An an example, while we know a positive relationship between principal experience and school performance exists, future work should examine *why* these schools perform worse under new principals. Additionally, while we have provided descriptive evidence that teacher perceptions of climate are substantially lower in years leading up to a principal transition, future work should employ more comprehensive measures of climate and satisfaction to better shed light on how turnover affects school climate. While a handful of studies, including this one, document that student achievement is declining prior to principal turnover, there may be other factors that are contributing to the increased likelihood of leadership instability. In particular, qualitative studies could provide important insight about the causes and consequences of principal turnover that is not easily captured in quantitative analyses. Finally, future work should more explicitly examine the consequences of turnover among effective versus ineffective leaders using multiple measures of principal quality, such as rubric-based observation ratings or principal value-added.

III.6 Appendix A: Supplementary Tables and Figures

Table III.A1: The Impact of Principal Turnover

	Missouri				Tennessee			
	Math	Read	Retention	New Tch	Math	Read	Retention	New Tch
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Time = -5 x Treat	0.003	-0.004	0.000	-0.028	-0.014	0.003	-0.002	0.108
	(0.010)	(0.010)	(0.002)	(0.070)	(0.021)	(0.015)	(0.003)	(0.178)
Time = -4 x Treat	-0.007	-0.003	0.000	-0.023	0.005	0.006	-0.005	-0.032
	(0.010)	(0.010)	(0.002)	(0.063)	(0.019)	(0.013)	(0.003)	(0.137)
Time = -3 x Treat	0.003	0.002	-0.001	0.010	0.013	0.004	-0.005*	0.068
	(0.009)	(0.009)	(0.002)	(0.057)	(0.018)	(0.010)	(0.003)	(0.133)
Time = -2 x Treat	0.001	0.001	0.002	-0.011	0.001	0.007	-0.004	0.023
	(0.008)	(0.008)	(0.002)	(0.055)	(0.016)	(0.009)	(0.003)	(0.119)
Time = -1 x Treat	0.003	-0.004	-0.001	-0.010	0.005	0.008	-0.003	-0.026
	(0.007)	(0.007)	(0.002)	(0.050)	(0.014)	(0.008)	(0.003)	(0.108)
Time = 1 x Treat	-0.031***	-0.029***	-0.014***	0.345***	-0.030**	-0.019**	-0.026***	0.675***
	(0.007)	(0.006)	(0.002)	(0.053)	(0.013)	(0.008)	(0.003)	(0.125)
Time = 2 x Treat	-0.029***	-0.026***	-0.010***	0.220***	-0.004	-0.008	-0.013***	0.609***
	(0.008)	(0.008)	(0.002)	(0.054)	(0.015)	(0.009)	(0.003)	(0.121)
Time = 3 x Treat	-0.019**	-0.016**	-0.003*	0.217***	0.024	0.001	-0.012***	0.324***
	(0.009)	(0.008)	(0.002)	(0.055)	(0.017)	(0.010)	(0.003)	(0.121)
Time = 4 x Treat	-0.020**	-0.014*	-0.005**	0.202***	0.010	0.014	-0.009***	0.207
	(0.009)	(0.008)	(0.002)	(0.055)	(0.019)	(0.011)	(0.003)	(0.134)
Time = 5 x Treat	-0.001	0.001	-0.001	0.135**	0.019	0.013	-0.004	0.091
	(0.010)	(0.009)	(0.002)	(0.057)	(0.019)	(0.012)	(0.003)	(0.138)
R^2	0.820	0.842	0.625	0.673	0.814	0.929	0.531	0.699

Notes: Standard errors clustered by school shown in parentheses. Models include school characteristics, school fixed effects, and district-by-year fixed effects. Coefficients show the effect of principal turnover relative to performance in the final year of the departing principal (Time = 0).
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table III.A2: The Impact of Types of Principal Turnover on School Achievement

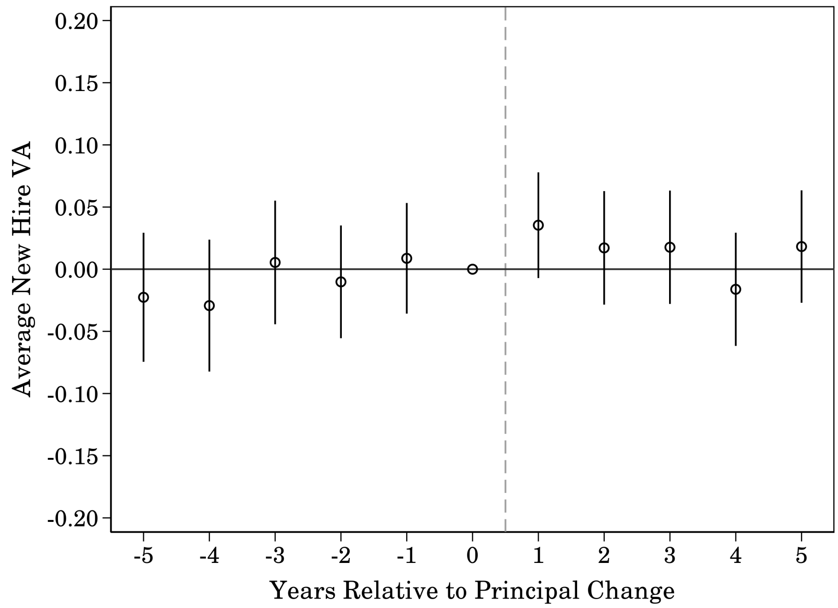
	Transfer		Exit		Promote		Demote	
	Math	Read	Math	Read	Math	Read	Math	Read
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Time = -5 x Treat	-0.005	-0.005	0.019	-0.001	-0.004	0.000	0.045*	0.007
	(0.018)	(0.016)	(0.015)	(0.014)	(0.018)	(0.016)	(0.026)	(0.022)
Time = -4 x Treat	-0.021	-0.002	0.011	0.001	-0.001	0.004	0.038	0.034*
	(0.017)	(0.016)	(0.014)	(0.013)	(0.017)	(0.015)	(0.025)	(0.019)
Time = -3 x Treat	-0.004	0.011	0.013	0.001	0.009	0.007	0.037	0.023
	(0.015)	(0.014)	(0.014)	(0.011)	(0.016)	(0.014)	(0.024)	(0.018)
Time = -2 x Treat	-0.002	-0.001	0.002	-0.004	0.009	0.007	0.027	0.015
	(0.014)	(0.012)	(0.012)	(0.010)	(0.015)	(0.012)	(0.020)	(0.018)
Time = -1 x Treat	0.008	-0.002	0.009	-0.003	0.013	0.016	0.013	0.010
	(0.012)	(0.011)	(0.010)	(0.008)	(0.012)	(0.010)	(0.018)	(0.014)
Time = 1 x Treat	-0.050***	-0.040***	-0.011	-0.024***	-0.040***	-0.020**	0.012	0.001
	(0.012)	(0.010)	(0.010)	(0.008)	(0.012)	(0.010)	(0.018)	(0.014)
Time = 2 x Treat	-0.025*	-0.031**	-0.011	-0.014	-0.032**	-0.021	0.001	0.004
	(0.014)	(0.012)	(0.012)	(0.010)	(0.015)	(0.013)	(0.022)	(0.016)
Time = 3 x Treat	-0.015	-0.030**	0.002	-0.010	-0.017	-0.007	0.022	0.032*
	(0.016)	(0.013)	(0.014)	(0.011)	(0.015)	(0.013)	(0.023)	(0.017)
Time = 4 x Treat	-0.047***	-0.025*	0.007	-0.002	-0.011	0.007	0.020	0.032*
	(0.017)	(0.014)	(0.015)	(0.011)	(0.018)	(0.014)	(0.024)	(0.019)
Time = 5 x Treat	-0.015	-0.025*	0.041***	0.024*	-0.027	0.014	0.015	0.046**
	(0.017)	(0.014)	(0.015)	(0.012)	(0.018)	(0.014)	(0.025)	(0.019)
R^2	0.826	0.864	0.831	0.878	0.823	0.866	0.830	0.883

Notes: Standard errors clustered by school shown in parentheses. Models include school characteristics, school fixed effects, and district-by-year fixed effects. Coefficients show the effect of principal turnover relative to performance in the final year of the departing principal (Time = 0).
* p < 0.1, ** p < 0.05, *** p < 0.01.

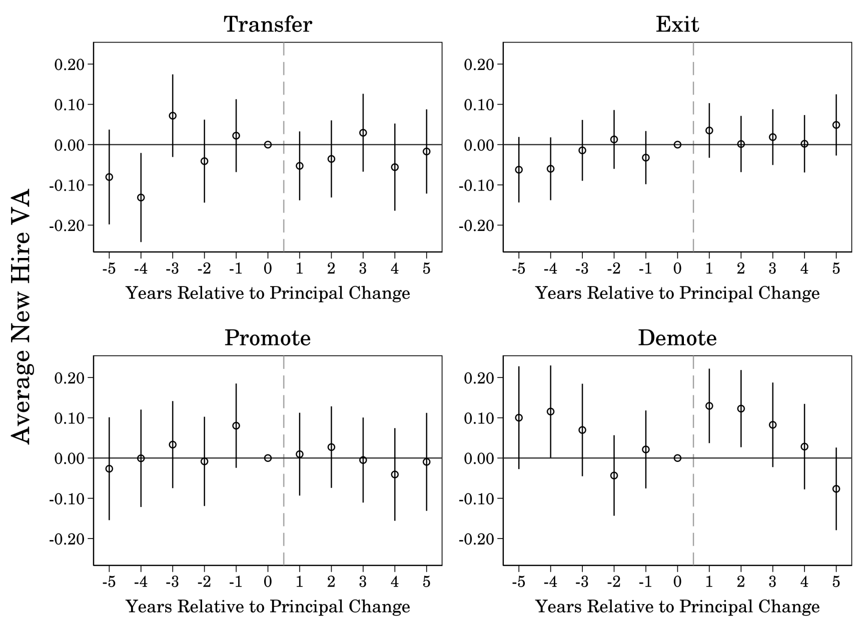
Table III.A3: The Impact of Types of Principal Turnover on Teacher Retention

	Transfer		Exit		Promote		Demote	
	Retain (1)	New Tch (2)	Retain (3)	New Tch (4)	Retain (5)	New Tch (6)	Retain (7)	New Tch (8)
Time = -5 x Treat	-0.007**	0.117	0.000	0.098	-0.009***	0.081	0.001	-0.196
	(0.003)	(0.124)	(0.003)	(0.110)	(0.003)	(0.135)	(0.005)	(0.182)
Time = -4 x Treat	-0.008**	0.094	-0.001	-0.033	-0.000	0.041	-0.010**	-0.065
	(0.003)	(0.112)	(0.003)	(0.093)	(0.003)	(0.119)	(0.005)	(0.167)
Time = -3 x Treat	-0.004	-0.010	-0.003	0.080	-0.009***	0.025	0.002	-0.049
	(0.003)	(0.099)	(0.003)	(0.093)	(0.003)	(0.123)	(0.005)	(0.143)
Time = -2 x Treat	-0.006*	0.069	0.002	0.021	-0.000	0.225*	-0.007	-0.148
	(0.003)	(0.106)	(0.002)	(0.085)	(0.003)	(0.126)	(0.005)	(0.150)
Time = -1 x Treat	-0.004	-0.042	-0.002	0.032	-0.003	0.007	-0.004	-0.246
	(0.003)	(0.091)	(0.002)	(0.081)	(0.003)	(0.117)	(0.004)	(0.163)
Time = 1 x Treat	-0.029***	0.673***	-0.015***	0.431***	-0.009***	0.376***	-0.025***	0.531***
	(0.003)	(0.112)	(0.002)	(0.082)	(0.003)	(0.111)	(0.005)	(0.155)
Time = 2 x Treat	-0.020***	0.470***	-0.009***	0.303***	-0.009***	0.349***	-0.014***	0.488***
	(0.003)	(0.106)	(0.002)	(0.084)	(0.003)	(0.113)	(0.004)	(0.169)
Time = 3 x Treat	-0.006*	0.277***	-0.007**	0.285***	-0.004	0.145	-0.021***	0.356**
	(0.003)	(0.104)	(0.003)	(0.089)	(0.003)	(0.110)	(0.005)	(0.160)
Time = 4 x Treat	-0.010***	0.288***	-0.008***	0.234**	-0.008**	0.099	-0.007	0.155
	(0.003)	(0.106)	(0.003)	(0.098)	(0.003)	(0.116)	(0.005)	(0.172)
Time = 5 x Treat	-0.008**	0.168	-0.003	0.151	-0.000	0.090	-0.003	0.051
	(0.003)	(0.111)	(0.003)	(0.096)	(0.003)	(0.128)	(0.005)	(0.183)
R^2	0.650	0.711	0.622	0.702	0.658	0.723	0.672	0.729

Notes: Standard errors clustered by school shown in parentheses. Models include school characteristics, school fixed effects, and district-by-year fixed effects. Coefficients show the effect of principal turnover relative to performance in the final year of the departing principal (Time = 0).
* p < 0.1, ** p < 0.05, *** p < 0.01.



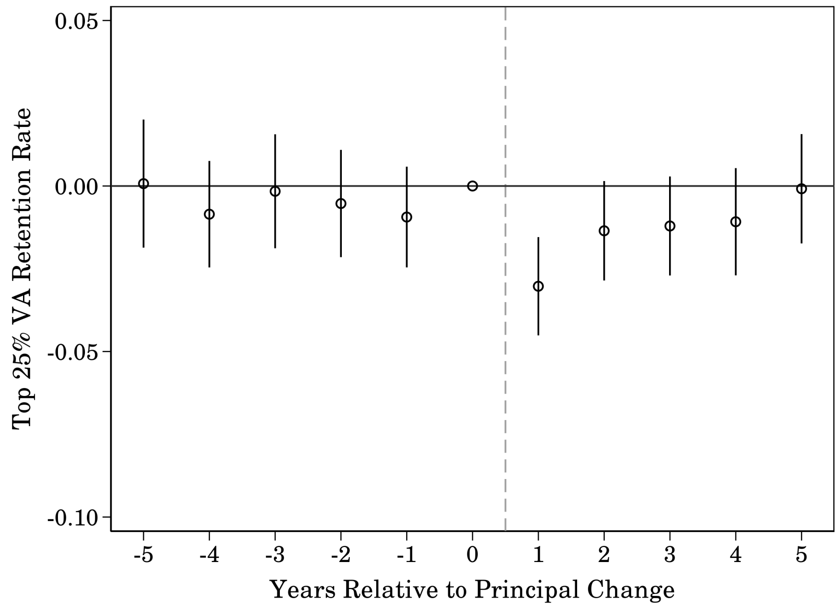
(a) All Turnover



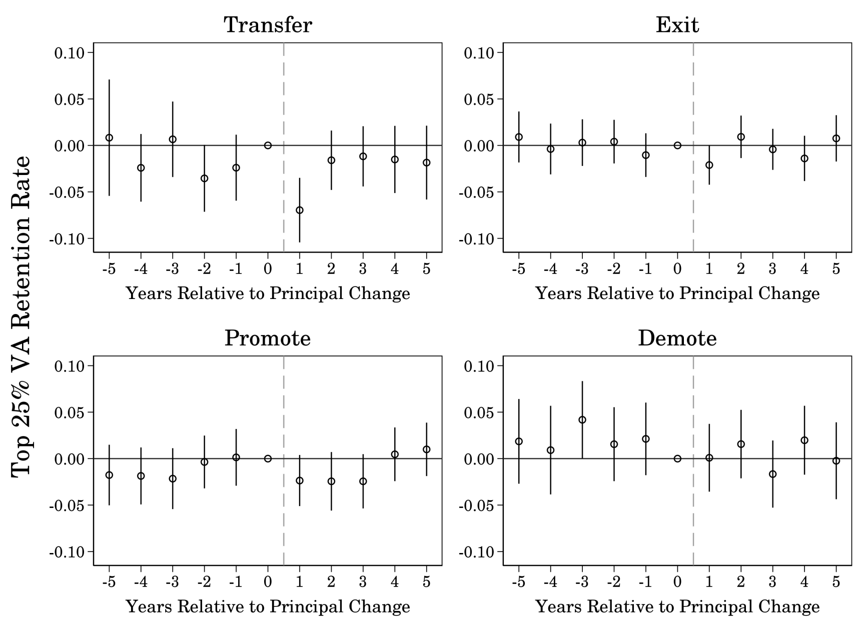
(b) Turnover Types

Figure III.A7: Impact of Principal Turnover on New Hire Quality

Notes: Each plot shows the focal coefficients (Treat x Time) from a difference-in-differences model for the listed turnover type. Average new hire value-added is expressed in teacher-level standard deviations (i.e., a value of 0.10 would indicate that the average newly hired teacher in the school is 0.10 SD above the mean in the statewide distribution of value-added in Tennessee.).



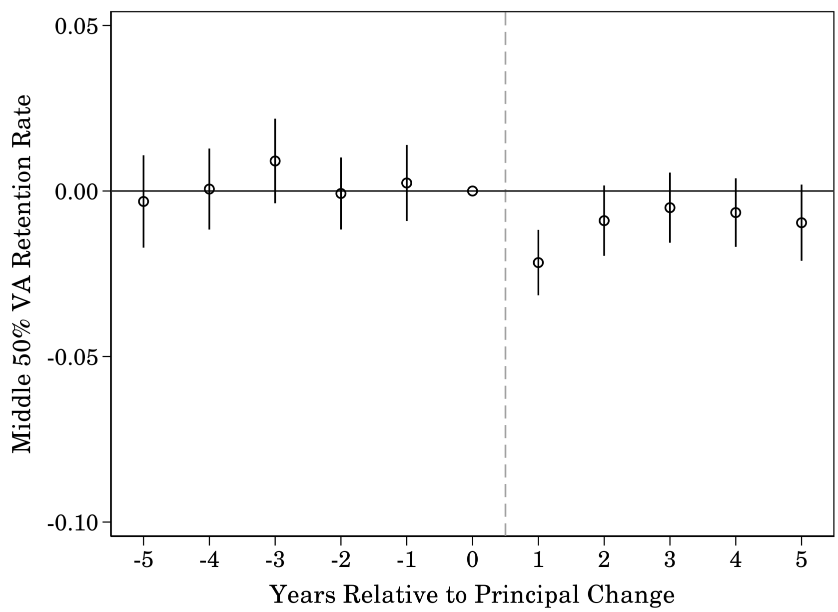
(a) All Turnover



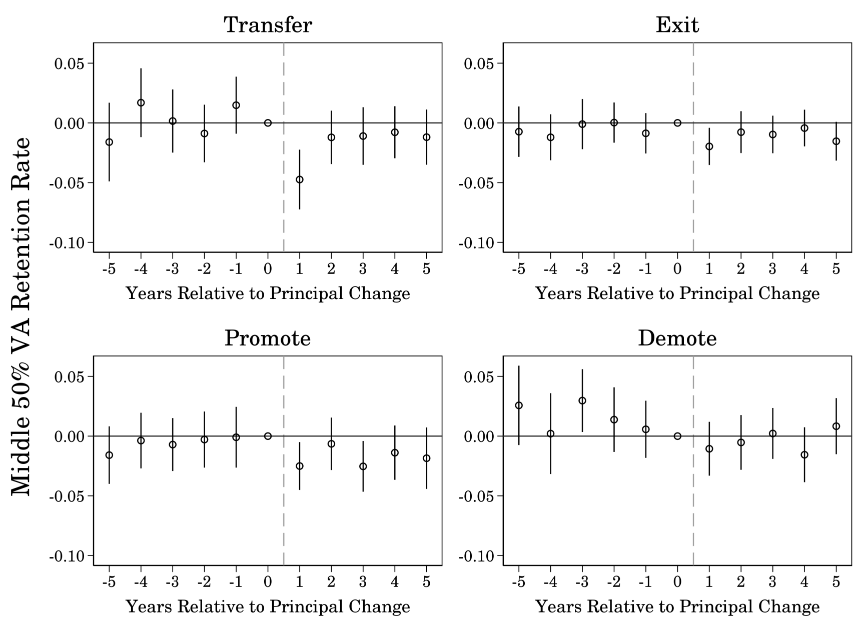
(b) Turnover Types

Figure III.A8: Impact of Principal Turnover on Retention of High-VA Teachers

Notes: Each plot shows the focal coefficients (Treat x Time) from a difference-in-differences model for the listed turnover type. High-VA is defined as top 25% in the statewide distribution of value-added in Tennessee.



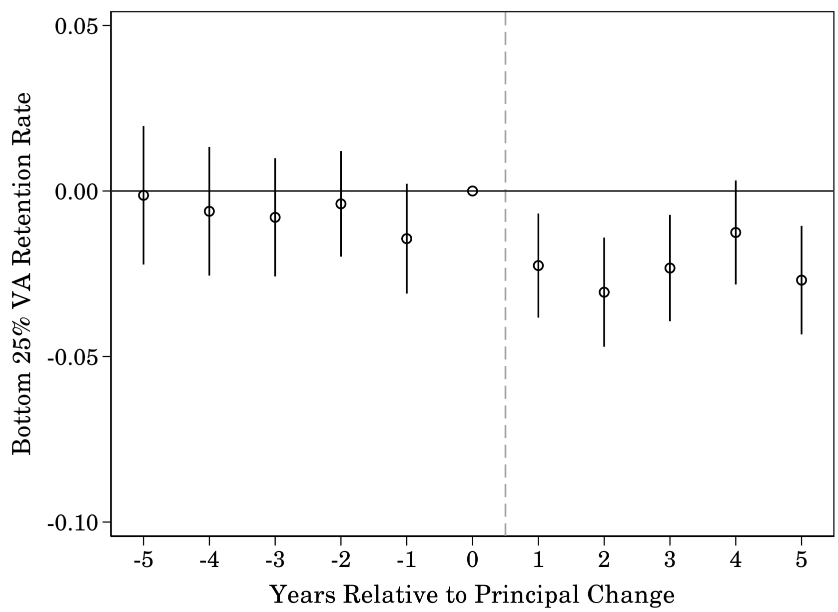
(a) All Turnover



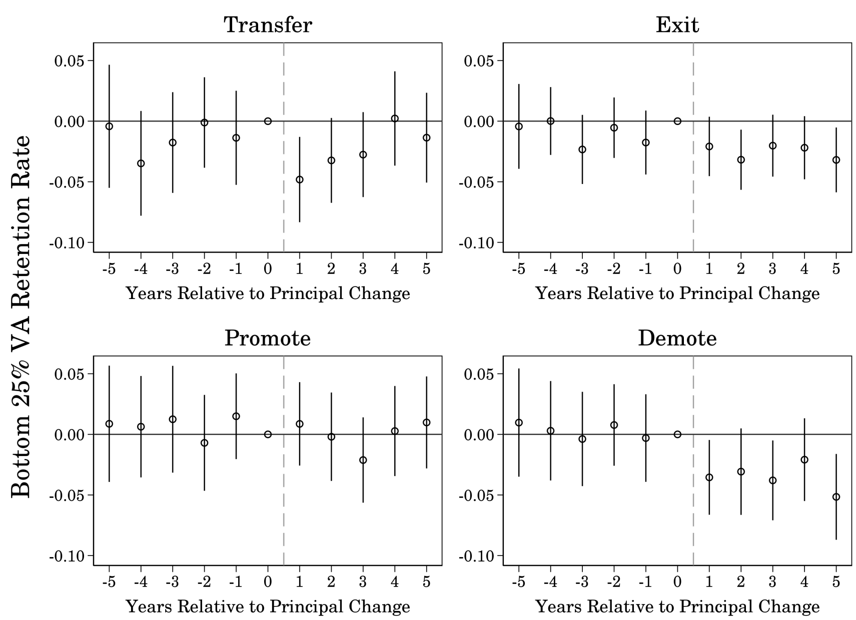
(b) Turnover Types

Figure III.A9: Impact of Principal Turnover on Retention of Middle-VA Teachers

Notes: Each plot shows the focal coefficients (Treat x Time) from a difference-in-differences model for the listed turnover type. High-VA is defined as mid 25% in the statewide distribution of value-added in Tennessee.



(a) All Turnover



(b) Turnover Types

Figure III.A10: Impact of Principal Turnover on Retention of Low-VA Teachers

Notes: Each plot shows the focal coefficients (Treat x Time) from a difference-in-differences model for the listed turnover type. High-VA is defined as low 25% in the statewide distribution of value-added in Tennessee.

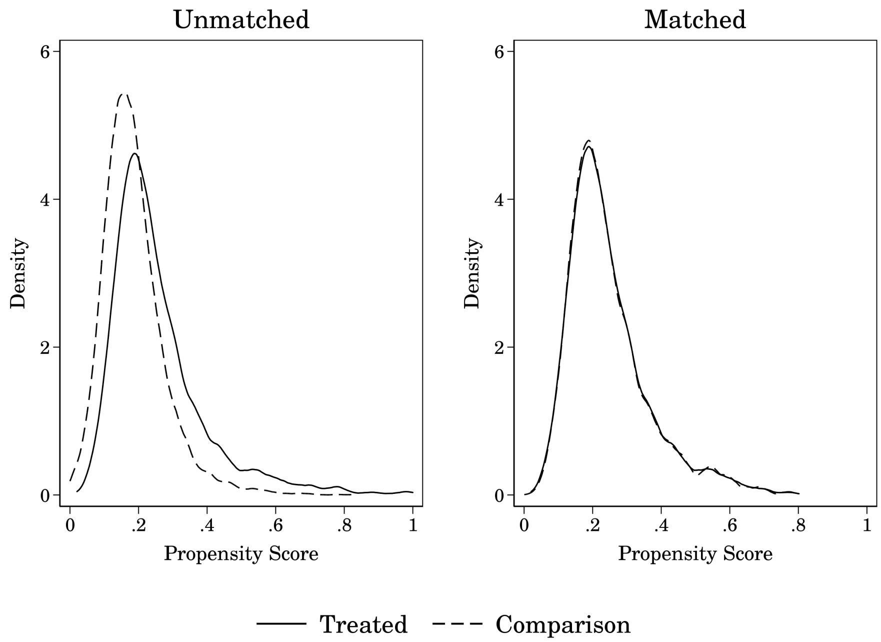
III.7 Appendix B: Matching Details

This appendix section shows the results of the procedure used to construct a matched comparison group for each type of principal turnover. The results are broken down separately by state and year.

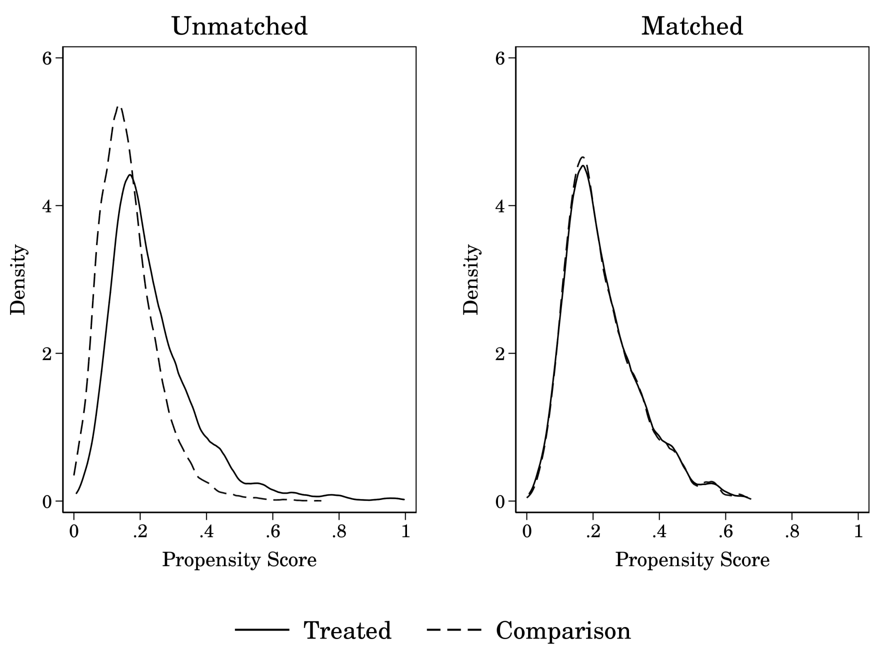
Table III.A4: Matching by Year (All Turnover)

	Treated		Comparison		
Year	Matched	Unmatched	Used	Unused	Bandwidth
Missouri					
2001	331	1	1290	44	0.006
2002	322	6	1301	37	0.011
2003	331	4	1293	15	0.011
2004	314	4	1291	15	0.018
2005	321	1	1319	22	0.030
2006	343	5	1329	5	0.030
2007	327	6	1341	7	0.025
2008	372	4	1302	4	0.018
2009	277	2	1215	1	0.055
2010	308	14	1236	18	0.011
2011	285	7	1340	8	0.014
2012	345	4	1278	6	0.025
2013	354	7	1259	12	0.025
2014	262	8	1368	4	0.030
2015	343	9	1289	29	0.025
All	4835	82	19451	227	.
Tennessee					
2007	208	5	1050	12	0.030
2008	235	3	1041	3	0.025
2009	211	9	1088	1	0.025
2010	197	5	1079	20	0.013
2011	235	6	1057	0	0.016
2012	252	1	1018	13	0.009
2013	221	1	1035	12	0.030
2014	236	5	1032	3	0.020
2015	217	7	1049	0	0.035
All	2012	42	9449	64	.

Notes: Kernel matching (epanechnikov) performed separately by year. Bandwidth chosen via cross-validation using the "kmatch" package in Stata.



(a) Missouri



(b) Tennessee

Figure III.A11: Propensity Score Distribution (All Turnover)

Table III.A5: Covariate Balance (All Turnover)

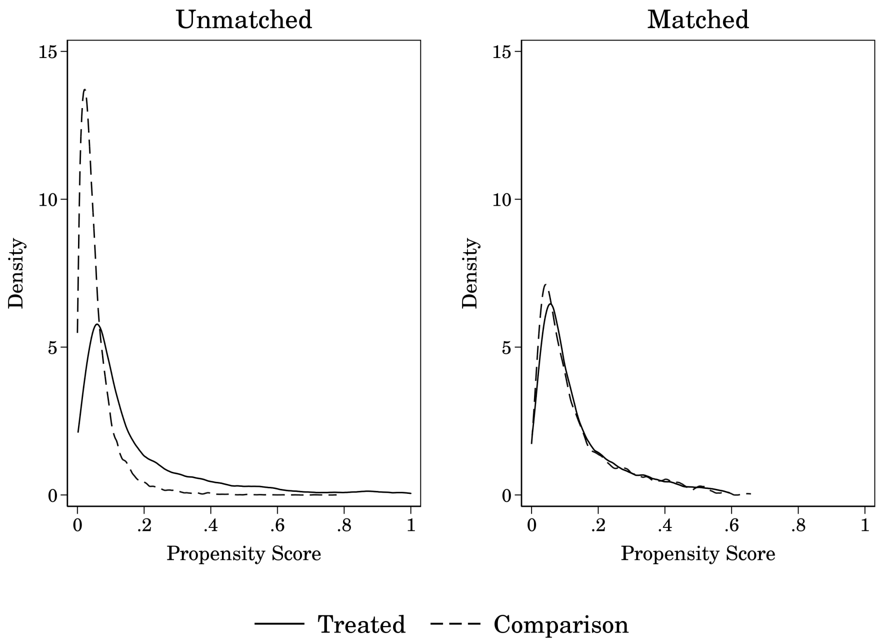
	Unmatched			Matched		
	Treat	Comp	Std Diff	Treat	Comp	Std Diff
Missouri						
Math (year t)	-0.077	0.111	-0.20	-0.050	-0.044	-0.01
Math (year $t-1$)	-0.041	0.117	-0.18	-0.018	-0.008	-0.01
Math (year $t-2$)	-0.020	0.124	-0.16	0.001	0.008	-0.01
Math (year $t-3$)	-0.010	0.125	-0.16	0.009	0.016	-0.01
Math (year $t-4$)	-0.007	0.126	-0.17	0.014	0.026	-0.02
Math (year $t-5$)	0.015	0.119	-0.14	0.033	0.040	-0.01
ELA (year t)	-0.077	0.118	-0.22	-0.046	-0.041	-0.01
ELA (year $t-1$)	-0.041	0.120	-0.18	-0.015	-0.007	-0.01
ELA (year $t-2$)	-0.014	0.121	-0.16	0.009	0.018	-0.01
ELA (year $t-3$)	-0.014	0.123	-0.17	0.004	0.018	-0.02
ELA (year $t-4$)	-0.012	0.125	-0.18	0.009	0.018	-0.01
ELA (year $t-5$)	0.004	0.119	-0.16	0.021	0.027	-0.01
Prin Transition (year $t-1$)	0.131	0.179	-0.13	0.131	0.130	0.00
Prin Transition (year $t-2$)	0.174	0.181	-0.02	0.174	0.174	0.00
Prin Transition (year $t-3$)	0.192	0.168	0.06	0.190	0.190	0.00
Prin Transition (year $t-4$)	0.170	0.167	0.01	0.170	0.166	0.01
Prin Transition (year $t-5$)	0.180	0.162	0.05	0.180	0.176	0.01
Prop New Tch (year t)	0.167	0.152	0.13	0.166	0.164	0.01
Prop New Tch (year $t-1$)	0.168	0.153	0.13	0.167	0.166	0.01
Prop New Tch (year $t-2$)	0.168	0.155	0.12	0.167	0.166	0.01
Prop New Tch (year $t-3$)	0.171	0.157	0.12	0.170	0.170	0.00
Prop New Tch (year $t-4$)	0.175	0.160	0.12	0.174	0.173	0.01
Prop New Tch (year $t-5$)	0.181	0.167	0.09	0.180	0.180	0.00
Tennessee						
Math (year t)	-0.073	0.082	-0.16	-0.052	-0.061	0.01
Math (year $t-1$)	-0.048	0.085	-0.14	-0.030	-0.043	0.01
Math (year $t-2$)	-0.029	0.087	-0.12	-0.015	-0.028	0.01
Math (year $t-3$)	-0.035	0.094	-0.14	-0.016	-0.038	0.02
Math (year $t-4$)	0.009	0.090	-0.09	0.031	0.016	0.02
Math (year $t-5$)	0.020	0.095	-0.08	0.036	0.023	0.01
ELA (year t)	-0.077	0.082	-0.17	-0.052	-0.061	0.01
ELA (year $t-1$)	-0.043	0.086	-0.14	-0.025	-0.034	0.01
ELA (year $t-2$)	-0.025	0.091	-0.12	-0.006	-0.018	0.01
ELA (year $t-3$)	-0.023	0.102	-0.13	-0.002	-0.013	0.01
ELA (year $t-4$)	0.011	0.108	-0.10	0.035	0.025	0.01
ELA (year $t-5$)	0.021	0.119	-0.11	0.039	0.023	0.02
Prin Transition (year $t-1$)	0.117	0.177	-0.17	0.117	0.121	-0.01
Prin Transition (year $t-2$)	0.155	0.174	-0.05	0.156	0.163	-0.02
Prin Transition (year $t-3$)	0.173	0.167	0.02	0.171	0.170	0.00
Prin Transition (year $t-4$)	0.160	0.161	-0.00	0.161	0.160	0.00
Prin Transition (year $t-5$)	0.168	0.153	0.04	0.168	0.169	-0.00
Prop New Tch (year t)	0.158	0.149	0.10	0.157	0.158	-0.01
Prop New Tch (year $t-1$)	0.152	0.148	0.05	0.151	0.152	-0.01
Prop New Tch (year $t-2$)	0.158	0.147	0.12	0.156	0.155	0.00
Prop New Tch (year $t-3$)	0.159	0.152	0.07	0.158	0.158	0.01
Prop New Tch (year $t-4$)	0.165	0.157	0.07	0.164	0.164	-0.00
Prop New Tch (year $t-5$)	0.259	0.260	-0.01	0.256	0.258	-0.00

Notes: "Treat" are schools who did not change principals in year t and "Comp" are comparison schools who did not change principals in year t. "Std Diff" refers to the standardized difference between treatment and comparison schools, which compares the difference in means in units of the pooled standard deviation.

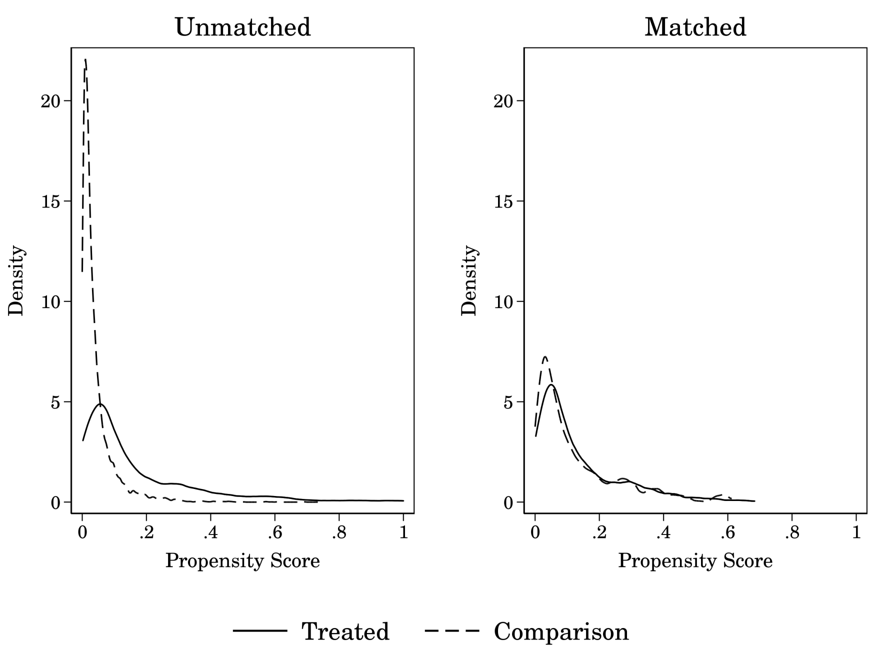
Table III.A6: Matching by Year (Transfers)

	Treated		Comparison		
Year	Matched	Unmatched	Used	Unused	Bandwidth
Missouri					
2001	102	1	1332	2	0.020
2002	80	4	1333	5	0.025
2003	92	2	1302	6	0.025
2004	68	6	1282	24	0.014
2005	88	3	1340	1	0.050
2006	94	4	1325	9	0.009
2007	91	3	1345	3	0.035
2008	99	3	1305	1	0.070
2009	65	8	1215	1	0.040
2010	82	5	1252	2	0.055
2011	78	3	1347	1	0.030
2012	91	11	1283	1	0.030
2013	89	8	1263	8	0.017
2014	71	4	1372	0	0.070
2015	86	14	1314	4	0.011
All	1276	79	19610	68	.
Tennessee					
2007	43	4	1032	30	0.009
2008	50	4	1017	27	0.007
2009	46	3	1069	2	0.030
2010	44	3	1013	1	0.019
2011	52	3	1043	0	0.055
2012	63	3	1030	1	0.045
2013	54	3	1031	0	0.050
2014	39	4	998	1	0.055
2015	52	1	1027	1	0.080
All	443	28	9260	63	.

Notes: Kernel matching (epanechnikov) performed separately by year. Bandwidth chosen via cross-validation using the "kmatch" package in Stata.



(a) Missouri



(b) Tennessee

Figure III.A12: Propensity Score Distribution (Transfers)

Table III.A7: Covariate Balance (Transfers)

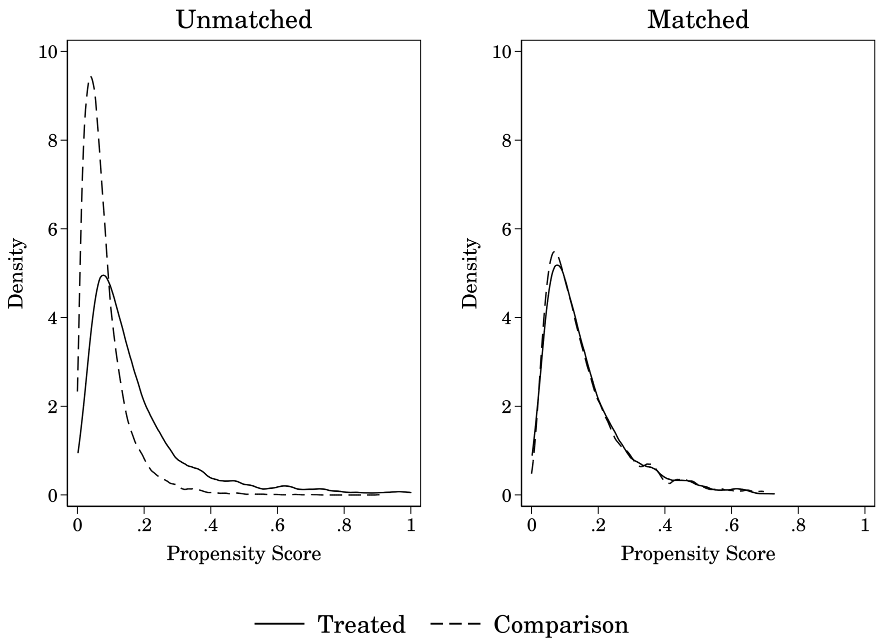
	Unmatched			Matched		
	Treat	Comp	Std Diff	Treat	Comp	Std Diff
Missouri						
Math (year t)	-0.022	0.111	-0.14	0.024	0.038	-0.01
Math (year $t-1$)	0.018	0.117	-0.11	0.057	0.069	-0.01
Math (year $t-2$)	0.022	0.124	-0.12	0.067	0.079	-0.01
Math (year $t-3$)	0.033	0.125	-0.11	0.069	0.080	-0.01
Math (year $t-4$)	0.035	0.126	-0.12	0.055	0.068	-0.02
Math (year $t-5$)	0.047	0.119	-0.10	0.068	0.077	-0.01
ELA (year t)	-0.048	0.118	-0.18	0.004	0.006	-0.00
ELA (year $t-1$)	-0.017	0.120	-0.16	0.031	0.032	-0.00
ELA (year $t-2$)	-0.016	0.121	-0.16	0.032	0.049	-0.02
ELA (year $t-3$)	0.014	0.123	-0.13	0.058	0.065	-0.01
ELA (year $t-4$)	-0.008	0.125	-0.17	0.020	0.028	-0.01
ELA (year $t-5$)	0.002	0.119	-0.16	0.028	0.036	-0.01
Prin Transition (year $t-1$)	0.160	0.179	-0.05	0.160	0.161	-0.00
Prin Transition (year $t-2$)	0.216	0.181	0.09	0.215	0.209	0.01
Prin Transition (year $t-3$)	0.228	0.168	0.15	0.226	0.220	0.01
Prin Transition (year $t-4$)	0.190	0.167	0.06	0.193	0.185	0.02
Prin Transition (year $t-5$)	0.201	0.162	0.10	0.206	0.198	0.02
Prop New Tch (year t)	0.173	0.152	0.18	0.168	0.170	-0.01
Prop New Tch (year $t-1$)	0.175	0.153	0.18	0.170	0.171	-0.01
Prop New Tch (year $t-2$)	0.177	0.155	0.19	0.172	0.174	-0.02
Prop New Tch (year $t-3$)	0.178	0.157	0.17	0.176	0.172	0.02
Prop New Tch (year $t-4$)	0.184	0.160	0.18	0.179	0.177	0.01
Prop New Tch (year $t-5$)	0.191	0.167	0.16	0.188	0.186	0.01
Tennessee						
Math (year t)	-0.090	0.085	-0.18	-0.065	-0.058	-0.01
Math (year $t-1$)	-0.049	0.087	-0.14	-0.028	-0.027	-0.00
Math (year $t-2$)	-0.041	0.090	-0.14	-0.041	-0.015	-0.03
Math (year $t-3$)	-0.085	0.099	-0.20	-0.080	-0.053	-0.03
Math (year $t-4$)	-0.080	0.094	-0.19	-0.053	-0.015	-0.04
Math (year $t-5$)	-0.044	0.099	-0.16	-0.042	-0.018	-0.03
ELA (year t)	-0.154	0.085	-0.25	-0.112	-0.099	-0.01
ELA (year $t-1$)	-0.108	0.087	-0.20	-0.089	-0.072	-0.02
ELA (year $t-2$)	-0.079	0.094	-0.18	-0.074	-0.039	-0.04
ELA (year $t-3$)	-0.058	0.107	-0.17	-0.052	-0.004	-0.05
ELA (year $t-4$)	-0.065	0.111	-0.19	-0.041	0.001	-0.04
ELA (year $t-5$)	-0.071	0.123	-0.21	-0.060	-0.015	-0.05
Prin Transition (year $t-1$)	0.144	0.177	-0.09	0.147	0.159	-0.04
Prin Transition (year $t-2$)	0.191	0.171	0.05	0.196	0.192	0.01
Prin Transition (year $t-3$)	0.221	0.166	0.14	0.217	0.201	0.04
Prin Transition (year $t-4$)	0.210	0.159	0.13	0.199	0.202	-0.01
Prin Transition (year $t-5$)	0.185	0.152	0.09	0.181	0.188	-0.02
Prop New Tch (year t)	0.174	0.149	0.25	0.170	0.173	-0.03
Prop New Tch (year $t-1$)	0.158	0.147	0.11	0.158	0.158	0.01
Prop New Tch (year $t-2$)	0.166	0.147	0.19	0.164	0.161	0.03
Prop New Tch (year $t-3$)	0.172	0.152	0.19	0.167	0.168	-0.00
Prop New Tch (year $t-4$)	0.174	0.156	0.17	0.171	0.172	-0.01
Prop New Tch (year $t-5$)	0.265	0.261	0.01	0.259	0.265	-0.02

Notes: "Treat" are schools who did not change principals in year t and "Comp" are comparison schools who did not change principals in year t. "Std Diff" refers to the standardized difference between treatment and comparison schools, which compares the difference in means in units of the pooled standard deviation.

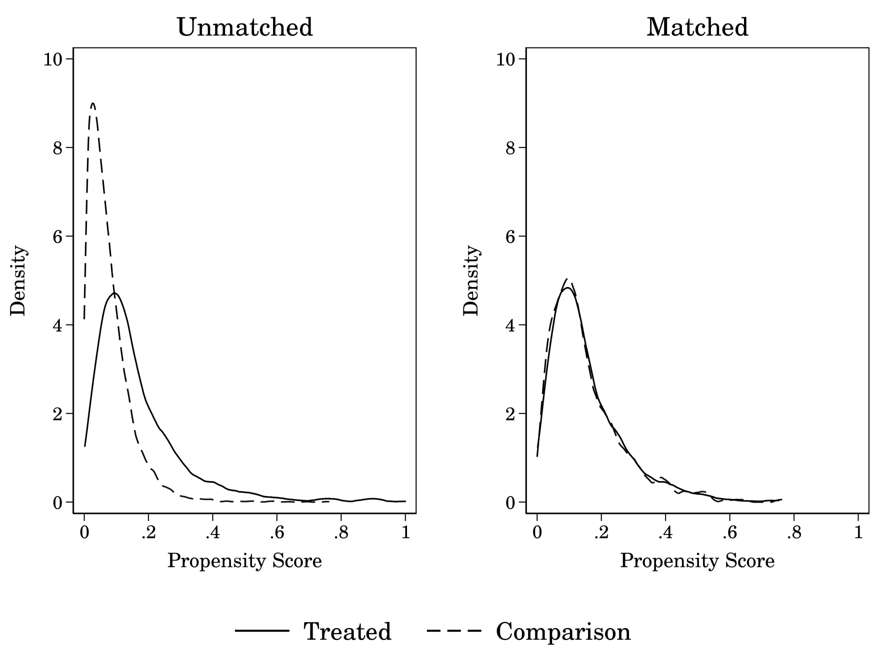
Table III.A8: Matching by Year (Exits)

Year	Treated		Comparison		Bandwidth
	Matched	Unmatched	Used	Unused	
Missouri					
2001	122	0	1332	2	0.035
2002	129	6	1180	158	0.005
2003	123	2	1307	1	0.030
2004	138	1	1306	0	0.055
2005	113	5	1276	65	0.006
2006	113	9	1292	42	0.013
2007	109	6	1344	4	0.016
2008	169	1	1305	1	0.035
2009	124	4	1208	8	0.015
2010	119	14	1251	3	0.025
2011	104	5	1317	31	0.013
2012	146	1	1283	1	0.035
2013	141	5	1270	1	0.035
2014	93	3	1371	1	0.017
2015	117	5	1314	4	0.030
All	1860	67	19356	322	.
Tennessee					
2007	90	1	1013	49	0.005
2008	91	4	1042	2	0.030
2009	91	2	1087	2	0.045
2010	87	2	1066	9	0.009
2011	101	1	1047	2	0.055
2012	108	0	1030	1	0.035
2013	82	3	1028	2	0.040
2014	106	1	1023	2	0.017
2015	90	4	1034	1	0.050
All	846	18	9370	70	.

Notes: Kernel matching (epanechnikov) performed separately by year. Bandwidth chosen via cross-validation using the "kmatch" package in Stata.



(a) Missouri



(b) Tennessee

Figure III.A13: Propensity Score Distribution (Exits)

Table III.A9: Covariate Balance (Exits)

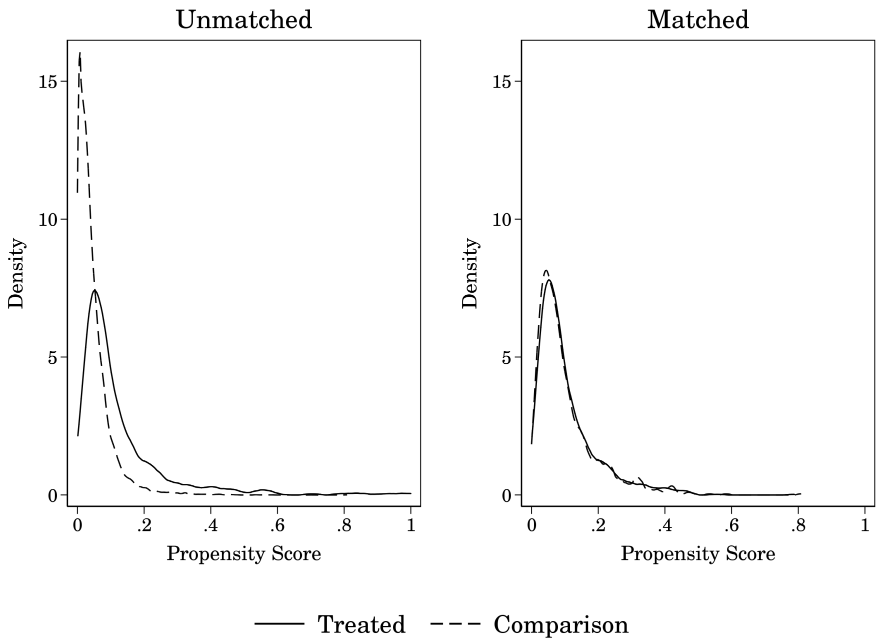
	Unmatched			Matched		
	Treat	Comp	Std Diff	Treat	Comp	Std Diff
Missouri						
Math (year t)	-0.145	0.111	-0.27	-0.087	-0.094	0.01
Math (year $t-1$)	-0.106	0.117	-0.24	-0.054	-0.064	0.01
Math (year $t-2$)	-0.082	0.124	-0.22	-0.031	-0.041	0.01
Math (year $t-3$)	-0.066	0.125	-0.22	-0.016	-0.023	0.01
Math (year $t-4$)	-0.051	0.126	-0.21	-0.002	-0.005	0.00
Math (year $t-5$)	-0.028	0.119	-0.19	0.016	0.003	0.02
ELA (year t)	-0.144	0.118	-0.28	-0.076	-0.091	0.02
ELA (year $t-1$)	-0.100	0.120	-0.24	-0.047	-0.060	0.01
ELA (year $t-2$)	-0.059	0.121	-0.20	-0.012	-0.016	0.01
ELA (year $t-3$)	-0.058	0.123	-0.21	-0.015	-0.029	0.02
ELA (year $t-4$)	-0.053	0.125	-0.22	-0.000	-0.015	0.02
ELA (year $t-5$)	-0.026	0.119	-0.19	0.015	0.009	0.01
Prin Transition (year $t-1$)	0.103	0.179	-0.22	0.099	0.106	-0.02
Prin Transition (year $t-2$)	0.135	0.181	-0.13	0.135	0.139	-0.01
Prin Transition (year $t-3$)	0.155	0.168	-0.04	0.151	0.149	0.01
Prin Transition (year $t-4$)	0.147	0.167	-0.06	0.146	0.144	0.01
Prin Transition (year $t-5$)	0.154	0.162	-0.02	0.154	0.155	-0.00
Prop New Tch (year t)	0.163	0.152	0.10	0.161	0.161	-0.00
Prop New Tch (year $t-1$)	0.164	0.153	0.09	0.162	0.162	0.00
Prop New Tch (year $t-2$)	0.162	0.155	0.06	0.160	0.161	-0.00
Prop New Tch (year $t-3$)	0.166	0.157	0.07	0.164	0.164	0.00
Prop New Tch (year $t-4$)	0.170	0.160	0.08	0.168	0.168	-0.00
Prop New Tch (year $t-5$)	0.175	0.167	0.05	0.173	0.173	0.01
Tennessee						
Math (year t)	-0.048	0.084	-0.14	-0.041	-0.046	0.01
Math (year $t-1$)	-0.031	0.087	-0.12	-0.022	-0.048	0.03
Math (year $t-2$)	-0.018	0.090	-0.12	-0.011	-0.020	0.01
Math (year $t-3$)	-0.019	0.098	-0.12	-0.011	-0.045	0.03
Math (year $t-4$)	0.012	0.093	-0.09	0.026	-0.017	0.05
Math (year $t-5$)	0.015	0.099	-0.09	0.034	-0.024	0.06
ELA (year t)	-0.027	0.084	-0.11	-0.017	-0.032	0.02
ELA (year $t-1$)	-0.005	0.088	-0.09	0.003	-0.012	0.02
ELA (year $t-2$)	0.002	0.094	-0.10	0.013	0.000	0.01
ELA (year $t-3$)	0.000	0.106	-0.11	0.012	-0.001	0.01
ELA (year $t-4$)	0.042	0.111	-0.07	0.064	0.023	0.04
ELA (year $t-5$)	0.043	0.122	-0.08	0.065	-0.001	0.07
Prin Transition (year $t-1$)	0.072	0.176	-0.32	0.073	0.081	-0.03
Prin Transition (year $t-2$)	0.110	0.173	-0.18	0.111	0.118	-0.02
Prin Transition (year $t-3$)	0.130	0.167	-0.10	0.126	0.128	-0.01
Prin Transition (year $t-4$)	0.127	0.161	-0.10	0.125	0.140	-0.04
Prin Transition (year $t-5$)	0.139	0.153	-0.04	0.138	0.139	-0.00
Prop New Tch (year t)	0.154	0.149	0.05	0.153	0.156	-0.02
Prop New Tch (year $t-1$)	0.151	0.148	0.03	0.149	0.152	-0.03
Prop New Tch (year $t-2$)	0.154	0.147	0.07	0.150	0.151	-0.02
Prop New Tch (year $t-3$)	0.156	0.152	0.04	0.154	0.157	-0.03
Prop New Tch (year $t-4$)	0.159	0.157	0.02	0.157	0.161	-0.03
Prop New Tch (year $t-5$)	0.262	0.261	0.00	0.260	0.264	-0.01

Notes: "Treat" are schools who did not change principals in year t and "Comp" are comparison schools who did not change principals in year t. "Std Diff" refers to the standardized difference between treatment and comparison schools, which compares the difference in means in units of the pooled standard deviation.

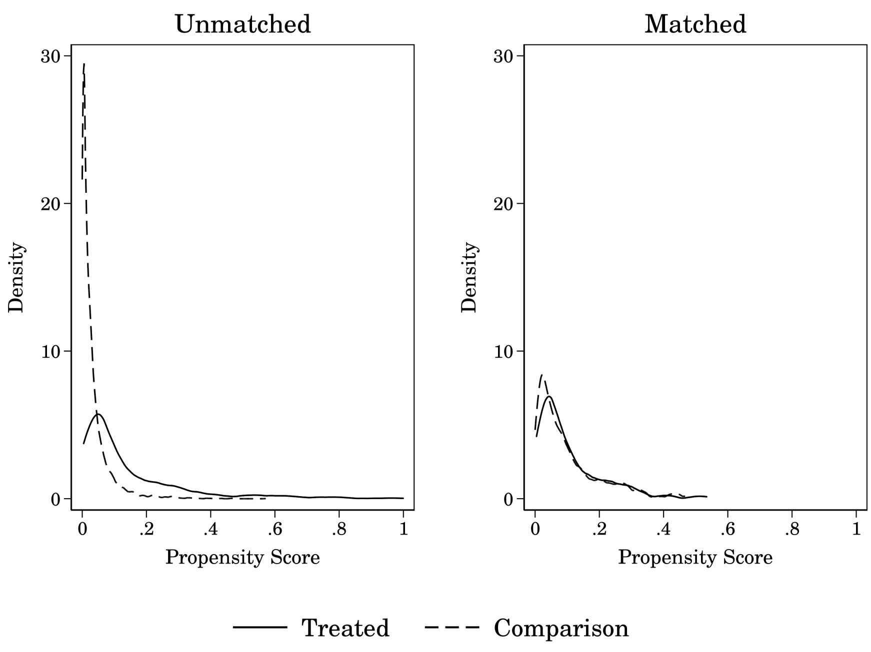
Table III.A10: Matching by Year (Promotions)

	Treated		Comparison		
Year	Matched	Unmatched	Used	Unused	Bandwidth
Missouri					
2001	66	0	1155	130	0.005
2002	63	2	1337	1	0.035
2003	62	5	1303	5	0.014
2004	71	0	1299	7	0.015
2005	61	5	1336	5	0.019
2006	86	0	1334	0	0.030
2007	69	1	1347	1	0.030
2008	71	1	1305	1	0.080
2009	46	4	1185	31	0.008
2010	60	3	1194	60	0.005
2011	69	1	1344	4	0.045
2012	57	2	946	338	0.007
2013	76	3	1270	1	0.030
2014	57	3	1366	6	0.014
2015	77	6	1302	16	0.014
All	991	36	19023	606	.
Tennessee					
2007	40	4	1036	26	0.011
2008	37	2	1021	23	0.011
2009	34	4	1070	19	0.011
2010	18	4	1037	24	0.013
2011	47	0	1034	0	0.070
2012	40	1	964	5	0.025
2013	37	3	988	9	0.011
2014	41	3	996	1	0.080
2015	40	6	1002	28	0.009
All	334	27	9148	135	.

Notes: Kernel matching (epanechnikov) performed separately by year. Bandwidth chosen via cross-validation using the "kmatch" package in Stata.



(a) Missouri



(b) Tennessee

Figure III.A14: Propensity Score Distribution (Promotions)

Table III.A11: Covariate Balance (Promotions)

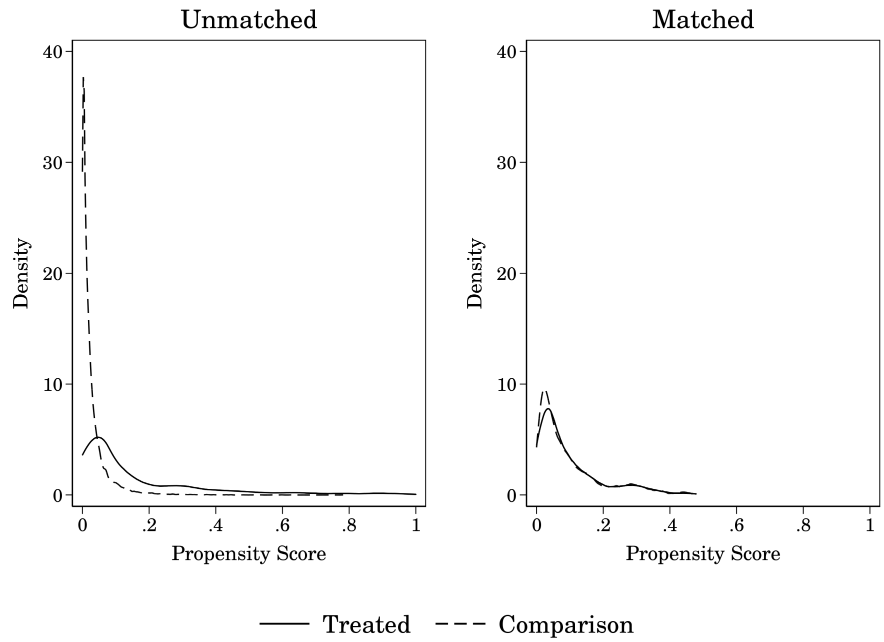
	Unmatched			Matched		
	Treat	Comp	Std Diff	Treat	Comp	Std Diff
Missouri						
Math (year t)	0.048	0.112	-0.08	0.073	0.050	0.03
Math (year $t-1$)	0.062	0.118	-0.07	0.093	0.071	0.03
Math (year $t-2$)	0.088	0.125	-0.05	0.106	0.090	0.02
Math (year $t-3$)	0.083	0.125	-0.05	0.094	0.074	0.03
Math (year $t-4$)	0.053	0.126	-0.10	0.076	0.059	0.02
Math (year $t-5$)	0.066	0.119	-0.07	0.086	0.075	0.02
ELA (year t)	0.067	0.118	-0.06	0.094	0.075	0.03
ELA (year $t-1$)	0.106	0.120	-0.02	0.132	0.104	0.04
ELA (year $t-2$)	0.132	0.122	0.01	0.161	0.137	0.03
ELA (year $t-3$)	0.076	0.123	-0.06	0.096	0.078	0.03
ELA (year $t-4$)	0.088	0.125	-0.05	0.110	0.096	0.02
ELA (year $t-5$)	0.087	0.119	-0.05	0.104	0.093	0.02
Prin Transition (year $t-1$)	0.105	0.179	-0.21	0.103	0.107	-0.01
Prin Transition (year $t-2$)	0.142	0.180	-0.10	0.141	0.144	-0.01
Prin Transition (year $t-3$)	0.173	0.168	0.01	0.170	0.166	0.01
Prin Transition (year $t-4$)	0.176	0.167	0.02	0.174	0.178	-0.01
Prin Transition (year $t-5$)	0.195	0.162	0.08	0.193	0.186	0.02
Prop New Tch (year t)	0.164	0.152	0.11	0.162	0.161	0.00
Prop New Tch (year $t-1$)	0.163	0.153	0.09	0.161	0.162	-0.01
Prop New Tch (year $t-2$)	0.163	0.155	0.07	0.161	0.160	0.01
Prop New Tch (year $t-3$)	0.167	0.157	0.09	0.165	0.164	0.01
Prop New Tch (year $t-4$)	0.166	0.160	0.05	0.163	0.165	-0.01
Prop New Tch (year $t-5$)	0.176	0.167	0.06	0.173	0.177	-0.03
Tennessee						
Math (year t)	0.052	0.088	-0.04	0.103	0.089	0.02
Math (year $t-1$)	0.095	0.091	0.00	0.123	0.106	0.02
Math (year $t-2$)	0.129	0.095	0.04	0.162	0.129	0.04
Math (year $t-3$)	0.112	0.102	0.01	0.151	0.135	0.02
Math (year $t-4$)	0.162	0.098	0.07	0.192	0.191	0.00
Math (year $t-5$)	0.225	0.103	0.13	0.259	0.240	0.02
ELA (year t)	0.028	0.088	-0.06	0.082	0.065	0.02
ELA (year $t-1$)	0.097	0.092	0.01	0.125	0.107	0.02
ELA (year $t-2$)	0.116	0.098	0.02	0.149	0.130	0.02
ELA (year $t-3$)	0.125	0.110	0.02	0.154	0.139	0.02
ELA (year $t-4$)	0.167	0.116	0.06	0.189	0.177	0.02
ELA (year $t-5$)	0.214	0.127	0.10	0.227	0.211	0.02
Prin Transition (year $t-1$)	0.100	0.176	-0.22	0.093	0.100	-0.03
Prin Transition (year $t-2$)	0.150	0.171	-0.06	0.150	0.157	-0.02
Prin Transition (year $t-3$)	0.152	0.165	-0.04	0.153	0.156	-0.01
Prin Transition (year $t-4$)	0.161	0.160	0.00	0.156	0.161	-0.01
Prin Transition (year $t-5$)	0.163	0.151	0.03	0.159	0.162	-0.01
Prop New Tch (year t)	0.141	0.148	-0.08	0.143	0.146	-0.03
Prop New Tch (year $t-1$)	0.143	0.147	-0.06	0.141	0.146	-0.06
Prop New Tch (year $t-2$)	0.149	0.147	0.03	0.147	0.147	0.00
Prop New Tch (year $t-3$)	0.150	0.152	-0.02	0.152	0.150	0.02
Prop New Tch (year $t-4$)	0.162	0.156	0.05	0.158	0.157	0.01
Prop New Tch (year $t-5$)	0.264	0.262	0.01	0.259	0.267	-0.03

Notes: "Treat" are schools who did not change principals in year t and "Comp" are comparison schools who did not change principals in year t. "Std Diff" refers to the standardized difference between treatment and comparison schools, which compares the difference in means in units of the pooled standard deviation.

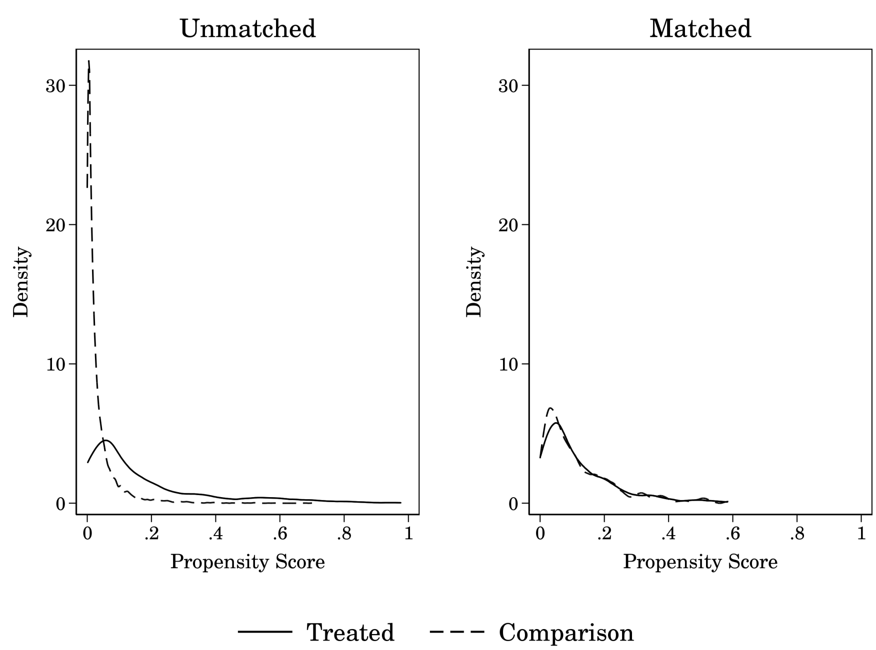
Table III.A12: Matching by Year (Demotions)

	Treated		Comparison		
Year	Matched	Unmatched	Used	Unused	Bandwidth
Missouri					
2001	39	2	1226	13	0.013
2002	37	8	1239	12	0.011
2003	50	7	1301	7	0.014
2004	29	8	1293	13	0.010
2005	45	2	1319	22	0.007
2006	43	4	1292	42	0.008
2007	49	6	1336	12	0.008
2008	33	5	1280	26	0.012
2009	28	6	1199	17	0.012
2010	49	6	1232	22	0.009
2011	26	10	1274	74	0.006
2012	34	7	1147	75	0.008
2013	36	6	1256	15	0.008
2014	35	6	1230	28	0.009
2015	43	7	1267	51	0.009
All	576	90	18891	429	.
Tennessee					
2007	34	2	1020	5	0.014
2008	51	2	931	100	0.008
2009	33	8	1051	38	0.008
2010	48	2	1036	9	0.015
2011	34	6	956	64	0.007
2012	34	5	986	11	0.013
2013	41	3	1027	12	0.014
2014	43	9	992	20	0.014
2015	29	7	1015	17	0.014
All	347	44	9014	276	.

Notes: Kernel matching (epanechnikov) performed separately by year. Bandwidth chosen via cross-validation using the "kmatch" package in Stata.



(a) Missouri



(b) Tennessee

Figure III.A15: Propensity Score Distribution (Demotions)

Table III.A13: Covariate Balance (Demotions)

	Unmatched			Matched		
	Treat	Comp	Std Diff	Treat	Comp	Std Diff
Missouri						
Math (year t)	-0.303	0.113	-0.44	-0.123	-0.123	-0.00
Math (year $t-1$)	-0.240	0.119	-0.39	-0.085	-0.095	0.01
Math (year $t-2$)	-0.209	0.127	-0.37	-0.071	-0.092	0.03
Math (year $t-3$)	-0.181	0.128	-0.36	-0.048	-0.053	0.01
Math (year $t-4$)	-0.154	0.127	-0.34	-0.002	-0.024	0.03
Math (year $t-5$)	-0.073	0.120	-0.25	0.049	0.024	0.04
ELA (year t)	-0.296	0.120	-0.45	-0.105	-0.114	0.01
ELA (year $t-1$)	-0.259	0.122	-0.41	-0.075	-0.097	0.03
ELA (year $t-2$)	-0.224	0.123	-0.39	-0.070	-0.096	0.03
ELA (year $t-3$)	-0.193	0.126	-0.38	-0.061	-0.074	0.02
ELA (year $t-4$)	-0.146	0.127	-0.34	-0.019	-0.033	0.02
ELA (year $t-5$)	-0.111	0.121	-0.30	0.012	0.009	0.00
Prin Transition (year $t-1$)	0.201	0.180	0.05	0.188	0.199	-0.03
Prin Transition (year $t-2$)	0.252	0.181	0.17	0.247	0.234	0.03
Prin Transition (year $t-3$)	0.260	0.169	0.22	0.245	0.251	-0.02
Prin Transition (year $t-4$)	0.191	0.168	0.06	0.186	0.189	-0.01
Prin Transition (year $t-5$)	0.188	0.164	0.06	0.189	0.186	0.01
Prop New Tch (year t)	0.177	0.151	0.22	0.169	0.172	-0.02
Prop New Tch (year $t-1$)	0.181	0.153	0.25	0.174	0.176	-0.02
Prop New Tch (year $t-2$)	0.182	0.155	0.23	0.171	0.175	-0.03
Prop New Tch (year $t-3$)	0.177	0.157	0.17	0.171	0.174	-0.03
Prop New Tch (year $t-4$)	0.186	0.160	0.21	0.176	0.177	-0.00
Prop New Tch (year $t-5$)	0.185	0.167	0.12	0.180	0.182	-0.01
Tennessee						
Math (year t)	-0.316	0.086	-0.42	-0.258	-0.220	-0.04
Math (year $t-1$)	-0.278	0.088	-0.38	-0.200	-0.180	-0.02
Math (year $t-2$)	-0.269	0.090	-0.38	-0.199	-0.185	-0.02
Math (year $t-3$)	-0.229	0.098	-0.35	-0.164	-0.167	0.00
Math (year $t-4$)	-0.151	0.096	-0.27	-0.068	-0.052	-0.02
Math (year $t-5$)	-0.202	0.098	-0.33	-0.121	-0.090	-0.04
ELA (year t)	-0.301	0.086	-0.40	-0.245	-0.208	-0.04
ELA (year $t-1$)	-0.271	0.089	-0.37	-0.207	-0.192	-0.02
ELA (year $t-2$)	-0.243	0.093	-0.35	-0.171	-0.143	-0.03
ELA (year $t-3$)	-0.268	0.105	-0.39	-0.189	-0.158	-0.03
ELA (year $t-4$)	-0.245	0.112	-0.37	-0.135	-0.115	-0.02
ELA (year $t-5$)	-0.235	0.122	-0.38	-0.138	-0.113	-0.03
Prin Transition (year $t-1$)	0.217	0.177	0.10	0.222	0.210	0.03
Prin Transition (year $t-2$)	0.217	0.172	0.12	0.216	0.199	0.04
Prin Transition (year $t-3$)	0.228	0.164	0.16	0.219	0.239	-0.05
Prin Transition (year $t-4$)	0.179	0.159	0.05	0.173	0.164	0.02
Prin Transition (year $t-5$)	0.225	0.153	0.18	0.222	0.224	-0.00
Prop New Tch (year t)	0.169	0.148	0.20	0.163	0.163	0.01
Prop New Tch (year $t-1$)	0.162	0.147	0.16	0.158	0.156	0.02
Prop New Tch (year $t-2$)	0.171	0.147	0.25	0.166	0.167	-0.01
Prop New Tch (year $t-3$)	0.162	0.152	0.11	0.160	0.160	-0.01
Prop New Tch (year $t-4$)	0.169	0.156	0.11	0.165	0.169	-0.04
Prop New Tch (year $t-5$)	0.246	0.258	-0.04	0.246	0.253	-0.02

Notes: "Treat" are schools who did not change principals in year t and "Comp" are comparison schools who did not change principals in year t. "Std Diff" refers to the standardized difference between treatment and comparison schools, which compares the difference in means in units of the pooled standard deviation.

III.8 Appendix C: Modeling Multiple Principal Transitions

A challenge to estimating the causal effect of principal turnover is that most schools experience principal transitions fairly frequently. In a difference-in-differences framework, this means that in any given year, a school can be in both the pre-treatment and post-treatment period. In prior work (e.g., Miller, 2013), this problem is addressed by modeling each principal transition as a separate event. Under this approach, school-by-year observations are duplicated by the number of principal transitions. "Stacking" the data allows for the inclusion of all principal transitions in the model.

To test the stacking approach, we construct a simulation that mirrors the observed patterns of principal turnover in Missouri and Tennessee. To be specific, we test two types of scenarios. In the first, principal turnover is uncorrelated with year-to-year fluctuations in achievement. This provides a basic test of whether stacking is unbiased in simplified circumstances. The second scenario is likely more realistic; principal turnover is correlated changes in school performance (see Figure III.A16 for examples of both scenarios). In each case, roughly 20% of schools experience principal turnover each year. The simulation has 1800 schools observed across 15 years, with 100 repetitions per scenario. Regardless of the scenario, we employ the same difference-in-differences model (combined with the matching strategy) used in our main analysis.

The data generating mechanism is as follows:

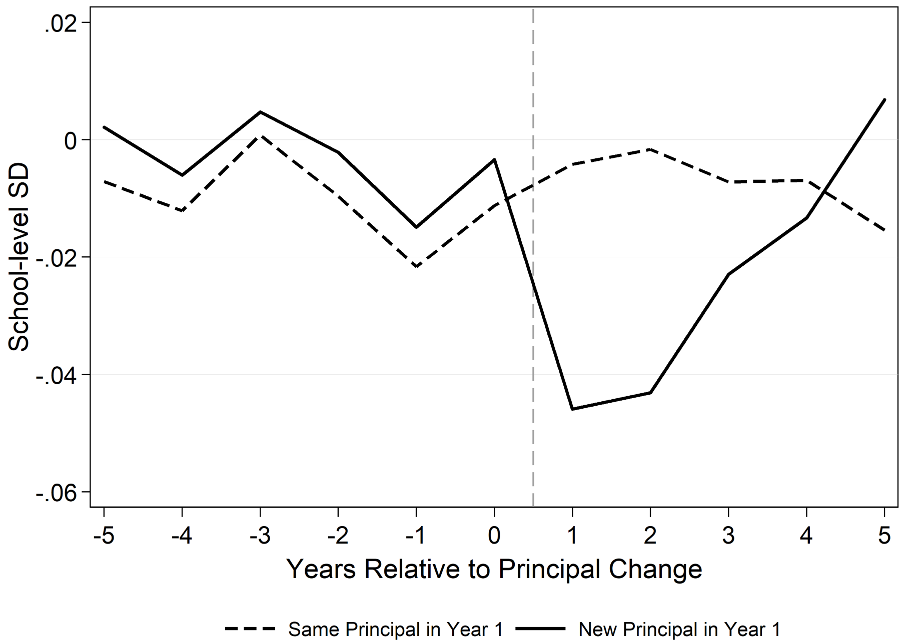
$$Y_{it} = \sum_{k=1}^{5} \delta_k + \alpha_i + \gamma_{it} + \epsilon_{it} \tag{III.2}$$

δ_k are the effects of a principal transition for up to five years. We test two scenarios of effects. In the first, δ_1 starts at -0.05 and decays linearly each year to -0.01 in the fifth year after the transition. In the second scenario, the negative effects fade out more quickly, such that δ_k takes values of -0.05, -0.03, and -0.01 in the first three years after a transition. α_i is fixed school heterogeneity distributed $\sim N(0, 0.9)$. γ_{it} is an AR1 term ($\varphi = 0.8$) distributed $\sim N(0, 0.3)$ that allows school performance to drift over time. ϵ_{it} is a random error distributed $\sim N(0, 0.3)$. An important assumption of this simulation is that the effects of turnover events are additively separable—there is no cumulative effect beyond the individual effects of each turnover event. Additionally, one instance of principal turnover does not increase the future likelihood of principal turnover. We do plan to extend our simulation to these more complicated scenarios in the future.

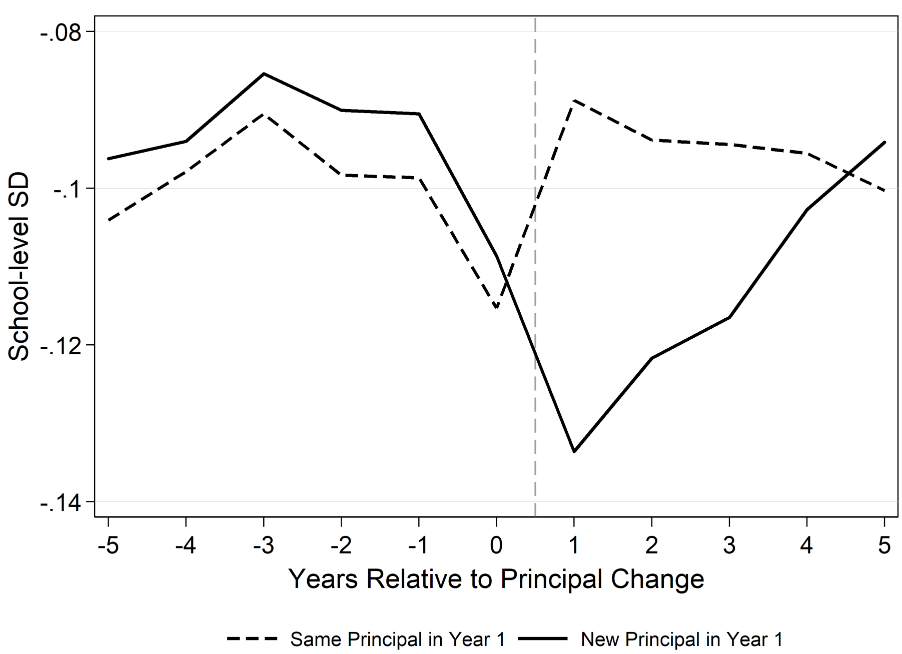
Tables III.A14 and III.A15 demonstrate that stacking turnover events does not produce biased results when turnover events are random. The ratio of the estimates to the true effects are all very close to one. However, the standard errors (which are clustered by school) are biased upwards by an average of 15 percent.

Tables III.A16 and III.A17 portray a more realistic scenario, perhaps, where turnover events are more common in schools that have downturns in performance. Here, we find that our analytic approach produces estimates that are larger in magnitude than the true effects. This bias grows larger (in proportional terms) for each additional lag. However, the overall magnitude of the bias is fairly small. For instance, the year 1 and 2 effects are biased by 4%

and 7% of the magntitude of the true effect.



(a) Example of Estimation with Random Turnover



(b) Example of Estimation with Correlated Turnover

Figure III.A16: Simulation Examples

Table III.A14: Simulation Results (Random Turnover)

	True	Estimate	Ratio
Coefficients			
Time = 1 x Treat	-0.050	-0.051	1.010
Time = 2 x Treat	-0.040	-0.041	1.021
Time = 3 x Treat	-0.030	-0.030	1.013
Time = 4 x Treat	-0.020	-0.019	0.954
Time = 5 x Treat	-0.010	-0.010	0.981
Standard Errors			
Time = 1 x Treat	0.006	0.008	1.182
Time = 2 x Treat	0.007	0.008	1.146
Time = 3 x Treat	0.007	0.008	1.208
Time = 4 x Treat	0.008	0.008	1.122
Time = 5 x Treat	0.008	0.009	1.118

Notes: "True" standard errors are the standard deviation of the coefficient estimates across the 100 simulations.

Table III.A15: Simulation Results (Random Turnover)

	True	Estimate	Ratio
Coefficients			
Time = 1 x Treat	-0.050	-0.051	1.019
Time = 2 x Treat	-0.030	-0.031	1.017
Time = 3 x Treat	-0.010	-0.010	1.023
Time = 4 x Treat	0.000	0.002	.
Time = 5 x Treat	0.000	0.001	.
Standard Errors			
Time = 1 x Treat	0.006	0.008	1.241
Time = 2 x Treat	0.007	0.008	1.194
Time = 3 x Treat	0.006	0.008	1.286
Time = 4 x Treat	0.007	0.008	1.222
Time = 5 x Treat	0.007	0.009	1.217

Notes: "True" standard errors are the standard deviation of the coefficient estimates across the 100 simulations.

Table III.A16: Simulation Results (Turnover Correlated with Performance Fluctuations)

	True	Estimate	Ratio
Coefficients			
Time = 1 x Treat	-0.050	-0.052	1.040
Time = 2 x Treat	-0.040	-0.043	1.070
Time = 3 x Treat	-0.030	-0.032	1.064
Time = 4 x Treat	-0.020	-0.022	1.112
Time = 5 x Treat	-0.010	-0.012	1.183
Standard Errors			
Time = 1 x Treat	0.007	0.008	1.119
Time = 2 x Treat	0.006	0.008	1.285
Time = 3 x Treat	0.007	0.008	1.109
Time = 4 x Treat	0.007	0.008	1.134
Time = 5 x Treat	0.008	0.009	1.055

Notes: "True" standard errors are the standard deviation of the coefficient estimates across the 100 simulations.

Table III.A17: Simulation Results (Turnover Correlated with Performance Fluctuations)

	True	Estimate	Ratio
Coefficients			
Time = 1 x Treat	-0.050	-0.050	1.005
Time = 2 x Treat	-0.030	-0.032	1.073
Time = 3 x Treat	-0.010	-0.012	1.157
Time = 4 x Treat	0.000	-0.004	.
Time = 5 x Treat	0.000	-0.001	.
Standard Errors			
Time = 1 x Treat	0.005	0.008	1.631
Time = 2 x Treat	0.006	0.008	1.358
Time = 3 x Treat	0.007	0.008	1.139
Time = 4 x Treat	0.008	0.009	1.050
Time = 5 x Treat	0.008	0.009	1.095

Notes: "True" standard errors are the standard deviation of the coefficient estimates across the 100 simulations.

III.9 Measuring School Climate

To create a measure of teacher perception of school climate, we draw on teacher responses from a low-stakes survey administered to Tennessee teachers beginning in 2011–12. For the first three years, the survey was part of the "First to the Top" (FTTT) initiative conducted by the Tennessee Consortium on Research, Evaluation, and Development (TNCRED) at Vanderbilt University, which was the precursor organization to the Tennessee Education Research Alliance (TERA). Beginning in 2014–15, the survey became a joint partnership between the Tennessee Department of Education and TERA, and is now called the "Tennessee Educator Survey" (TES). While the FTTT and TES surveys both contain items that ask teachers about their perceptions of school climate, the specific items change across years. The overall response rates for the surveys are (ascending by year) 24.8%, 38.7%, 41.9%, 55.3%, 48.1%, and 56.2%.

Our goal in constructing the climate measure was to ensure consistency across years while maintaining a high level of reliability within each year. To this end, we created two measures: a "conservative" measure that only included items that were identical or very similar across surveys and our preferred measure that supplemented the common items with survey- or year-specific items that also assessed school climate. The list of items is shown in Table III.A18, with the common items in panel A and the supplementary items in panel B. For each item, teacher responses are on a four-point Likert scale (strongly disagree, disagree, agree, strongly agree).

We show the factor analysis results (principal-factor method) for our preferred measure (i.e., using all of the items shown in Table III.A18) for each year in Tables III.A19–III.A24. The reliability (Cronbach's alpha) is above 0.90 in each year. We use the weights from the factor analysis to produce a teacher-level factor score in each year. To construct the school-by-year level score, we average teacher scores within each school-by-year cell and standardize these school-by-year averages within year, which becomes the dependent variable used in Table III.4. Figure III.A17 compares the factor scores from our preferred and conservative measures. In each year, they are highly correlated both at the teacher- and school-average level. The major difference between the preferred and conservative measure is that the reliability is lower for the conservative measure (Cronbach's alpha is 0.82, 0.84, 0.82, 0.73, 0.75, 0.79, ascending by year). For robustness, we show the school climate results using the conservative measure in Table III.A25. As would be implied by the high correlations between the measures, the results are very similar to Table III.4, which gives us confidence that differences in the items across years do not confound our main results.

We also examined survey non-response, which could potentially bias our estimates if teachers are systematically more or less likely to respond to the survey as a function of principal turnover. Table III.A26 shows the school-level response rates (i.e., the proportion of teachers in the school that responded to the survey each year) before and after principal transitions. There is a very small drop (1 percentage point) in the response rate in the final year of a departing principal, which appears to be driven by exits and promotions. The small magnitude of these results suggests that survey non-response is not a major issue for our school climate analysis. As an additional piece of evidence, Table III.A27 shows that including response rate as an additional predictor produces nearly identical results to those in Table III.4.

Table III.A18: List of Survey Items Used in Factor Analysis by Year

Survey Item (FTTT)	2012	2013	2014	Questionnaire Item (TES)	2015	2016	2017
(A) Very Similar or Identical Items							
The staff at this school like being here; I would describe us as a satisfied group.	✓	✓	✓	The staff at this school like being here; I would describe us as a satisfied group.	✓	✓	✓
I like the way things are run at this school.	✓	✓	✓	I like the way things are run at this school.			✓
Teachers in this school have high expectations for our students' achievement.	✓	✓	✓	Teachers hold all students to high academic standards.	✓	✓	✓
The principal at my school sets high standards for teaching.	✓	✓	✓	Administrators hold teachers to high professional standards for delivering instruction.	✓	✓	
(B) Supplementary Items							
Teachers' involvement in policy or decision-making is taken seriously.	✓	✓	✓	The staff feels comfortable raising issues and concerns that are important to them with school leaders.	✓	✓	✓
Leaders in this school trust the professional judgment of teachers.	✓	✓	✓	There is an atmosphere of trust and mutual respect within this school.	✓	✓	✓
Teachers in this school encourage students to keep trying even when the work is challenging.	✓	✓		Most of my colleagues share my beliefs and values about what the central mission of the school should be.	✓	✓	
Teachers in this school let students know that making mistakes is OK as long as they are learning and improving.	✓	✓	✓	School leadership makes a sustained effort to address staff concerns.	✓	✓	✓
The principal at my school sets high standards for student learning.	✓	✓	✓	School leadership provides useful feedback about my instructional practices.	✓	✓	✓
Leaders value teachers' ideas.	✓	✓	✓	School leadership proactively seeks to understand the needs of teachers and staff.		✓	✓
Leaders take time to praise teachers that perform well.	✓	✓	✓				
Teachers and leaders regularly engage in conversations about improving instruction.	✓	✓	✓				

Checkmarks indicate that the given item was included on the survey for the given year. FTTT = First to the Top Survey, TES = Tennessee Educator Survey

Table III.A19: Results of Factor Analysis for Teacher Perception of School Climate (2012)

Survey Item	Factor Loading
The staff at this school like being here; I would describe us as a satisfied group.	0.66
I like the way things are run at this school.	0.79
Teachers in this school have high expectations for our students' achievement.	0.48
The principal at my school sets high standards for teaching.	0.74
Teachers' involvement in policy or decision-making is taken seriously.	0.80
Leaders in this school trust the professional judgment of teachers.	0.83
Teachers in this school encourage students to keep trying even when the work is challenging.	0.46
Teachers in this school let students know that making mistakes is OK as long as they are learning and improving.	0.46
The principal at my school sets high standards for student learning.	0.76
Leaders value teachers' ideas.	0.83
Leaders take time to praise teachers that perform well.	0.77
Teachers and leaders regularly engage in conversations about improving instruction.	0.77

N = 2,272 teacher responses. Eigenvalue for factor = 6.0. Reliability of scale = 0.91.

Table III.A20: Results of Factor Analysis for Teacher Perception of School Climate (2013)

Survey Item	Factor Loading
The staff at this school like being here; I would describe us as a satisfied group.	0.70
I like the way things are run at this school.	0.80
Teachers in this school have high expectations for our students' achievement.	0.48
The principal at my school sets high standards for teaching.	0.76
Teachers' involvement in policy or decision-making is taken seriously.	0.82
Leaders in this school trust the professional judgment of teachers.	0.85
Teachers in this school encourage students to keep trying even when the work is challenging.	0.49
Teachers in this school let students know that making mistakes is OK as long as they are learning and improving.	0.45
The principal at my school sets high standards for student learning.	0.79
Leaders value teachers' ideas.	0.84
Leaders take time to praise teachers that perform well.	0.79
Teachers and leaders regularly engage in conversations about improving instruction.	0.77

N = 2,750 teacher responses. Eigenvalue for factor = 6.3. Reliability of scale = 0.92.

Table III.A21: Results of Factor Analysis for Teacher Perception of School Climate (2014)

Survey Item	Factor Loading
The staff at this school like being here; I would describe us as a satisfied group.	0.65
I like the way things are run at this school.	0.79
Teachers in this school have high expectations for our students' achievement.	0.47
The principal at my school sets high standards for teaching.	0.77
Teachers' involvement in policy or decision-making is taken seriously.	0.81
Leaders in this school trust the professional judgment of teachers.	0.82
Teachers in this school let students know that making mistakes is OK as long as they are learning and improving.	0.44
The principal at my school sets high standards for student learning.	0.78
Leaders value teachers' ideas.	0.82
Leaders take time to praise teachers that perform well.	0.78
Teachers and leaders regularly engage in conversations about improving instruction.	0.77

N = 3,543 teacher responses. Eigenvalue for factor = 5.9. Reliability of scale = 0.92.

Table III.A22: Results of Factor Analysis for Teacher Perception of School Climate (2015)

Survey Item	Factor Loading
The staff at this school like being here; I would describe us as a satisfied group.	0.76
Teachers hold all students to high academic standards.	0.59
Administrators hold teachers to high professional standards for delivering instruction.	0.68
The staff feels comfortable raising issues and concerns that are important to them with school leaders.	0.82
There is an atmosphere of trust and mutual respect within this school.	0.85
Most of my colleagues share my beliefs and values about what the central mission of the school should be.	0.66
School leadership makes a sustained effort to address staff concerns.	0.83
School leadership provides useful feedback about my instructional practices.	0.76

N = 31,897 teacher responses. Eigenvalue for factor = 4.5. Reliability of scale = 0.91.

Table III.A23: Results of Factor Analysis for Teacher Perception of School Climate (2016)

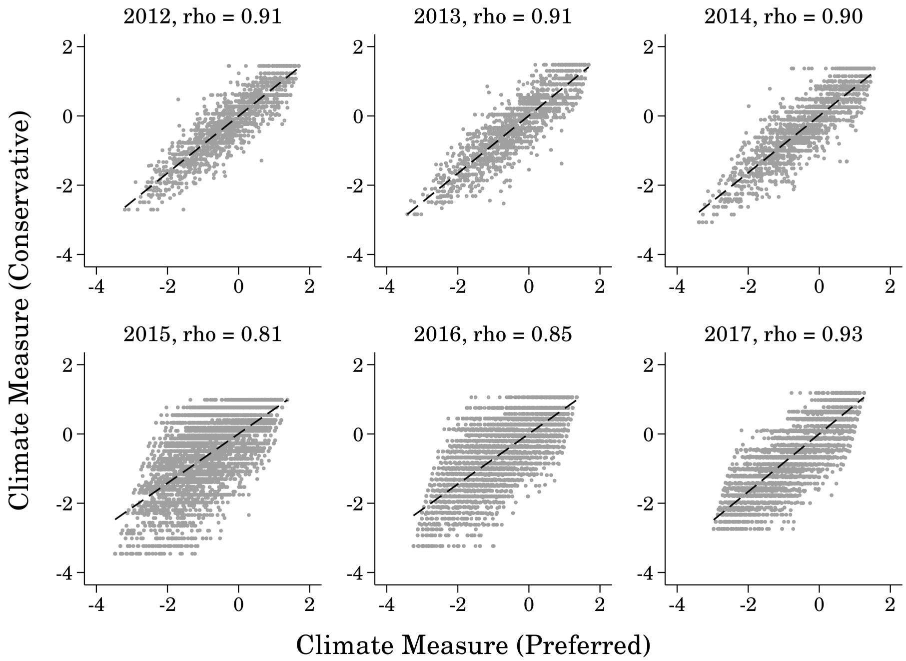
Survey Item	Factor Loading
The staff at this school like being here; I would describe us as a satisfied group.	0.78
Teachers hold all students to high academic standards.	0.58
Administrators hold teachers to high professional standards for delivering instruction.	0.72
The staff feels comfortable raising issues and concerns that are important to them with school leaders.	0.83
There is an atmosphere of trust and mutual respect within this school.	0.86
Most of my colleagues share my beliefs and values about what the central mission of the school should be.	0.65
School leadership makes a sustained effort to address staff concerns.	0.89
School leadership provides useful feedback about my instructional practices.	0.79
School leadership proactively seeks to understand the needs of teachers and staff.	0.89

N = 27,422 teacher responses. Eigenvalue for factor = 5.5. Reliability of scale = 0.93.

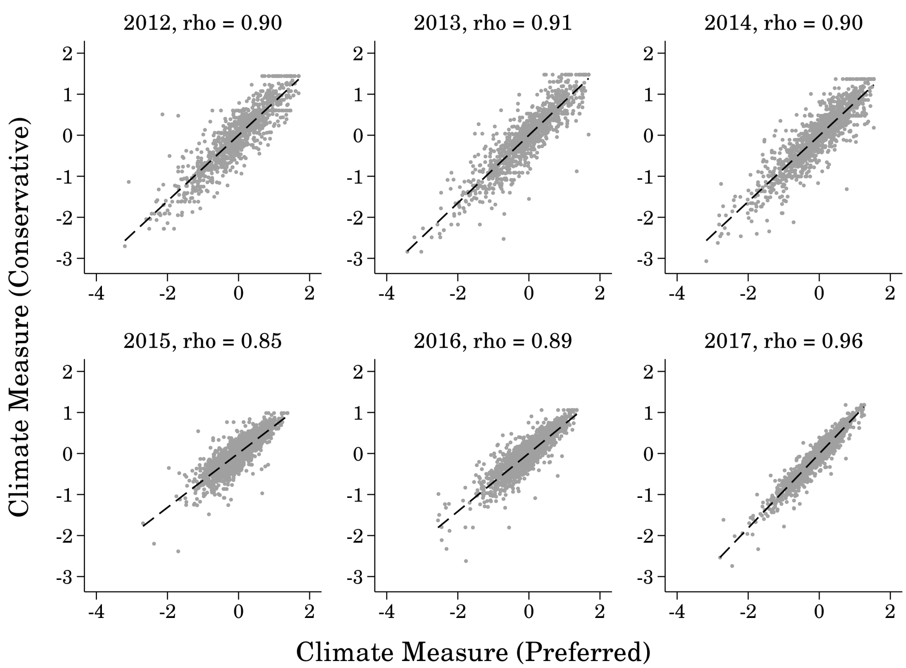
Table III.A24: Results of Factor Analysis for Teacher Perception of School Climate (2017)

Survey Item	Factor Loading
The staff at this school like being here; I would describe us as a satisfied group.	0.80
I like the way things are run at this school.	0.88
Teachers hold all students to high academic standards.	0.51
The staff feels comfortable raising issues and concerns that are important to them with school leaders.	0.84
There is an atmosphere of trust and mutual respect within this school.	0.87
School leadership makes a sustained effort to address staff concerns.	0.91
School leadership provides useful feedback about my instructional practices.	0.79
School leadership proactively seeks to understand the needs of teachers and staff.	0.90

N = 33,398 teacher responses. Eigenvalue for factor = 5.4. Reliability of scale = 0.94.



(a) Teacher-level Scores



(b) School-level Scores

Figure III.A17: Comparison of Preferred and Conservative Measures of School Climate

Table III.A25: Robustness Check: School Climate Results Using "Conservative" Measure

	(1)	(2)
All Principal Turnover		
One Year Before	-0.079**	
	(0.039)	
Current Year	-0.102**	
	(0.041)	
One Year After	0.027	
	(0.039)	
Two Years After	-0.046	
	(0.035)	
Principal Transfer		
One Year Before		0.072
		(0.085)
Current Year		0.059
		(0.088)
One Year After		0.035
		(0.078)
Two Years After		-0.087
		(0.071)
Principal Exit		
One Year Before		-0.145***
		(0.055)
Current Year		-0.135**
		(0.054)
One Year After		0.034
		(0.058)
Two Years After		-0.012
		(0.050)
Principal Promotion		
One Year Before		0.022
		(0.082)
Current Year		0.077
		(0.087)
One Year After		0.016
		(0.080)
Two Years After		-0.138*
		(0.077)
Principal Demotion		
One Year Before		-0.200**
		(0.085)
Current Year		-0.419***
		(0.107)
One Year After		0.067
		(0.086)
Two Years After		0.039
		(0.082)
N	6508	6508
R^2	0.567	0.572

Notes: Standard errors clustered by school shown in parentheses. Models include school fixed effects, district-by-year fixed effects, and school demographic controls. "Current year" is the year of the principal turnover event (i.e., the principal leaves between year t and year $t+1$.) Details on the factor analysis that produces the climate score is available in Appendix Table III.9.
* p < 0.1, ** p < 0.05, *** p < 0.01.

Table III.A26: Robustness Check: School-level Teacher Response Rates on Survey Before and After Principal Turnover

	(1)	(2)
All Principal Turnover		
One Year Before	-0.005	
	(0.004)	
Current Year	-0.010**	
	(0.005)	
One Year After	-0.006	
	(0.005)	
Two Years After	-0.002	
	(0.004)	
Principal Transfer		
One Year Before		-0.001
		(0.009)
Current Year		0.001
		(0.009)
One Year After		-0.010
		(0.009)
Two Years After		-0.003
		(0.008)
Principal Exit		
One Year Before		-0.007
		(0.007)
Current Year		-0.012*
		(0.007)
One Year After		0.001
		(0.007)
Two Years After		-0.002
		(0.006)
Principal Promotion		
One Year Before		-0.004
		(0.009)
Current Year		-0.017*
		(0.010)
One Year After		-0.013
		(0.009)
Two Years After		0.004
		(0.010)
Principal Demotion		
One Year Before		-0.008
		(0.009)
Current Year		-0.007
		(0.009)
One Year After		-0.012
		(0.010)
Two Years After		-0.007
		(0.009)
N	7884	7884
R^2	0.809	0.809

Notes: Standard errors clustered by school shown in parentheses. Models include school fixed effects, district-by-year fixed effects, and school demographic controls. "Current year" is the year of the principal turnover event (i.e., the principal leaves between year t and year $t+1$.)
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table III.A27: Robustness Check: School Climate Results Controlling for Response Rate

	(1)	(2)
Response Rate	0.281**	0.278**
	(0.119)	(0.118)
All Principal Turnover		
One Year Before	-0.048	
	(0.041)	
Current Year	-0.111***	
	(0.043)	
One Year After	0.056	
	(0.040)	
Two Years After	-0.010	
	(0.036)	
Principal Transfer		
One Year Before		0.116
		(0.088)
Current Year		0.010
		(0.086)
One Year After		0.058
		(0.078)
Two Years After		-0.023
		(0.074)
Principal Exit		
One Year Before		-0.110*
		(0.057)
Current Year		-0.123**
		(0.058)
One Year After		0.105*
		(0.058)
Two Years After		0.012
		(0.052)
Principal Promotion		
One Year Before		0.043
		(0.086)
Current Year		0.102
		(0.087)
One Year After		-0.019
		(0.091)
Two Years After		-0.079
		(0.081)
Principal Demotion		
One Year Before		-0.185**
		(0.086)
Current Year		-0.472***
		(0.104)
One Year After		0.056
		(0.086)
Two Years After		0.046
		(0.082)
N	6508	6508
R^2	0.555	0.561

Notes: Standard errors clustered by school shown in parentheses. Models include school fixed effects, district-by-year fixed effects, and school demographic controls. "Current year" is the year of the principal turnover event (i.e., the principal leaves between year t and year $t+1$.). Response rate is the proportion of teachers in the school-by-year cell that responded to the survey.
* p < 0.1, ** p < 0.05, *** p < 0.01.

BIBLIOGRAPHY

Adnot, M., Dee, T., Katz, V., and Wyckoff, J. (2017). Teacher Turnover, Teacher Quality, and Student Achievement in DCPS. *Educational Evaluation and Policy Analysis*, 39(1):1–23.

Allensworth, E. M. and Easton, J. Q. (2007). What Matters for Staying On-Track and Graduating in Chicago Public High Schools A Close Look at Course Grades, Failures, and Attendance in the Freshman Year What Matters for Staying On-Track and Graduating in Chicago Public High Schools. Technical report, Consortium on Chicago School Research, Chicago, IL.

Aucejo, E. M. and Romano, T. F. (2016). Assessing the effect of school days and absences on test score performance. *Economics of Education Review*, 55:70–87.

Backes, B. and Hansen, M. (2018). The Impact of Teach For America on Non-Test Academic Outcomes. *Education Finance and Policy*, 13(2):168–193.

Barnes, G., Crowe, E., and Schaefer, B. (2007). The cost of teacher turnover in five states: A pilot study. *National Commision on Teaching and America's Future*, pages Washington, D.C.

Bartanen, B., Grissom, J. A., and Rogers, L. K. (2019). The Impacts of Principal Turnover. *Educational Evaluation and Policy Analysis*.

Bastian, K. C. and Henry, G. T. (2015). The Apprentice: Pathways to the Principalship and Student Achievement. *Educational Administration Quarterly*, 51(4):600–639.

Becker, G. S. (1962). Investment in Human Capital : A Theoretical Analysis. *Journal of Political Economy*, 70(5):9–49.

Bergman, P. (2015). Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment.

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

Black, A. T., Seder, R. C., and Kekahio, W. (2014). Review of Research on Student Nonenrollment and Chronic Absenteeism: A Report for the Pacific Region. Technical report, U.S. Department of Education, Institute of Education Sciences, Naitonal Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Pacific, Washington, D.C.

Blazar, D. and Kraft, M. A. (2017). Teacher and Teaching Effects on Students Attitudes and Behaviors. *Educational Evaluation and Policy Analysis*, 39(1):146–170.

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

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

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

Brown, K. M. and Wynn, S. R. (2009). Finding, Supporting, and Keeping: The Role of the Principal in Teacher Retention Issues. *Leadership and Policy in Schools*, 8(1):37–63.

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.

Carroll, T. G. (2007). Policy brief: The high cost of teacher turnover. Technical report, National Commission on Teaching and America's Future.

Cattan, S., Kamhöfer, D. A., Karlsson, M., and Nilsson, T. (2017). The Short- and Long-Term Effects of Student Absence: Evidence from Sweden.

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

Chiang, H., Lipscomb, S., and Gill, B. (2016). Is School Value Added Indicative of Principal Quality? *Education Finance and Policy*, 11(3):283–309.

Clark, D., Martorell, P., and Rockoff, J. (2009). School Principals and School Performance.

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

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.

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

Dhuey, E. and Smith, J. (2018). How school principals influence student learning. *Empirical Economics*, 54:851–882.

Eberts, R. W. and Stone, J. A. (1988). Student achievement in public schools: Do principals make a difference? *Economics of Education Review*, 7(3):291–299.

Epstein, J. L. and Sheldon, S. B. (2002). Present and accounted for: Improving student attendance through family and community involvement. *Journal of Educational Research*, 95(5):308–318.

Farley-Ripple, E. N., Raffel, J. A., and Welch, J. C. (2012). Administrator career paths and decision processes. *Journal of Educational Administration*, 50(6):788–816.

Fuller, E. and Young, M. D. (2009). Tenure and Retention of Newly Hired Principals in Texas. Technical report, University Council for Educational Administration.

Gates, S. M., Ringel, J. S., Santibañez, L., Chung, C. H., and Ross, K. E. (2003). Who is Leading our Schools? An Overview of School Administrators and Their Careers. Technical report, RAND Corporartion, Santa Monica.

Gershenson, S. (2016). Linking Teacher Quality, Student Attendance, and Student Achievement. *Education Finance and Policy*, 11(2):125–149.

Gershenson, S., Jacknowitz, A., and Brannegan, A. (2017). Are Student Absences Worth the Worry in U.S. Primary Schools? *Education Finance and Policy*, 12(2):137–165.

Goldring, R. and Taie, S. (2018). Principal Attrition and Mobility: Results from the 2016-17 Principal Follow-up Survey (NCES 2018-066). Technical report, National Center for Education Statistics, Washington, DC.

Goodman, J. (2014). Flaking Out: Student Absences and Snow Days as Disruptions of Instructional Time.

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

Grissom, J. A. and Bartanen, B. (2019a). Principal Effectiveness and Principal Turnover. *Education Finance and Policy*, 14(3):1–28.

Grissom, J. A. and Bartanen, B. (2019b). Strategic Retention: Principal Effectiveness and Teacher Turnover in Multiple-Measure Teacher Evaluation Systems. *American Educational Research Journal*, 56(2):514–555.

Grissom, J. A., Bartanen, B., and Mitani, H. (2019). Principal Sorting and the Distribution of Principal Quality. *AERA Open*, 5(2):1–21.

Grissom, J. A., Blissett, R. S. L., and 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., and Loeb, S. (2015a). Using Student Test Scores to Measure Principal Performance. *Educational Evaluation and Policy Analysis*, 37(1):3–28.

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

Grissom, J. A., Loeb, S., and Master, B. (2013). Effective Instructional Time Use for School Leaders: Longitudinal Evidence From Observations of Principals. *Educational Researcher*, 42(8):433–444.

Grissom, J. A., Loeb, S., and 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., Viano, S. L., and Selin, J. L. (2015b). Understanding Employee Turnover in the Public Sector: Insights from Research on Teacher Mobility. *Public Administration Review*, 76(2):241–251.

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

Hallfors, D., Vevea, J. L., Iritani, B., Cho, H., Khatapoush, S., and Saxe, L. (2002). Truancy, Grade Point Average, and Sexual Activity: A Meta-Analysis of Risk Indicators for Youth Substance Use. *Journal of School Health*, 72(5):205–211.

Hallinger, P. and Heck, R. H. (1998). Exploring the Principal's Contribution to School Effectiveness: 1980-1995. *School Effectiveness and School Improvement*, 9(2):157–191.

Harris, D. N. and Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95(7-8):798–812.

Heckman, J. and Kautz, T. (2013). Fostering and Measuring Skills: Interventions That Improve Character and Cognition.

Henry, K. L. and Thornberry, T. P. (2010). Truancy and Escalation of Substance Use During Adolescence. *Journal of Studies on Alcohol and Drugs*, 71(1):115–124.

Horng, E., Klasik, D., and Loeb, S. (2010). Principal's Time Use and School Effectiveness. *American Journal of Education*, 116.

Jackson, C. K. (2012). Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina.

Jackson, C. K. (2016). What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes.

Jacob, B. (2011). Do Principals Fire the Worst Teachers? *Educational Evaluation and Policy Analysis*, 33(January):403–434.

Johnson, S. M. and Birkeland, S. E. (2003). Pursuing a "Sense of Success": New Teachers Explain Their Career Decisions. *American Educational Research Journal*, 40(3):581–617.

Jordan, P. W., Fothergill, S., and Rosende, M. (2018). Writing the Rules: Ensuring Chronic Absenteeism Data Works for Schools and Students. Technical report, FutureEd, Washington, D.C.

Jordan, P. W. and Miller, R. (2017). Who's In: Chronic Absenteeism Under the Every Student Succeeds Act. Technical report, FutureEd, Washington, D.C.

Kane, T. J., Rockoff, J. E., and Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, 27(6):615–631.

Kraft, M. A., Marinell, W. H., and Yee, D. (2016). School Organizational Contexts, Teacher Turnover, and Student Achievement: Evidence from Panel Data. *American Educational Research Journal*, 53(5):1411–1449.

Kraft, M. A. and Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, 30(4):476–500.

Kraft, M. A. and Rogers, T. (2015). The underutilized potential of teacher-to-parent communication: Evidence from a field experiment. *Economics of Education Review*, 47:49–63.

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.

Ladd, H. F. and Sorensen, L. C. (2017). Returns to Teacher Experience: Student Achievement and Motivation in Middle School. *Education Finance and Policy*, 12(2).

Leithwood, K. and Jantzi, D. (2000). The effects of transformational leadership on organizational conditions and student engagement with school. *Journal of Educational Administration*, 38(2):112–129.

Liu, J. and Loeb, S. (2017). Engaging Teachers: Measuring the Impact of Teachers on Student Attendance in Secondary School.

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

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

Mansfield, R. K. (2015). Teacher Quality and Student Inequality. *Journal of Labor Economics*, 33(3):751–788.

Marks, H. M. and Printy, S. M. (2003). Principal Leadership and School Performance: An Integration of Transformational and Instructional Leadership. *Educational Administration Quarterly*, 39(3):370–397.

Master, B. (2014). Staffing for Success: Linking Teacher Evaluation and School Personnel Management in Practice. *Educational Evaluation and Policy Analysis*, 36(2):207–227.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., and Hamilton, L. S. (2003). Evaluating Models for Teacher Accountability. Technical report, RAND Corporation, Santa Monica, CA.

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

Monk, D. H. and Ibrahim, M. A. (1984). Patterns of Absence and Pupil Achievement. *American Educational Research Journal*, 21(2):295–310.

Mueser, P. R., Troske, K. R., and Gorislavsky, A. (2007). Using State Administrative Data To Measure Program Performance. *The Review of Economics and Statistics*, 89(4):761–783.

Papay, J. P. and Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, 130:105–119.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2):417–458.

Robinson, V. M. J., Lloyd, C. A., and Rowe, K. J. (2008). The Impact of Leadership on Student Outcomes: An Analysis of the Differential Effects of Leadership Types. *Educational Administration Quarterly*, 44(5):635–674.

Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *Papers and Proceedings of the One Hundred Sixteenth Annual Meetings of the American Economic Assocation San Diego, CA*, 94(2):247–252.

Rockoff, J. E., Staiger, D. O., Kane, T. J., and Taylor, E. S. (2010). Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools. *NBER Working Paper*.

Roderick, M., Arney, M., Axelman, M., Dacosta, K., Steiger, C., Stone, S., Villarreal-sosa, L., and Waxman, E. (1997). Habits hard to break: A new look at truancy in Chicago's public high schools. Technical report, School of Social Service Administration, University of Chicago, Chicago, IL.

Rogers, T., Duncan, T., Wolford, T., Ternovski, J., Subramanyam, S., and Reitano, A. (2017). A Randomized Experiment Using Absenteeism Information to 'Nudge' Attendance. Technical report, U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Education Laboratory Mid-Atlantic., Washington, D.C.

Rogers, T. and Feller, A. (2018). Reducing student absences at scale by targeting parents' misbeliefs. *Nature Human Behaviour*, 2(5):335–342.

Ronfeldt, M., Loeb, S., and Wyckoff, J. (2013). How Teacher Turnover Harms Student Achievement. *American Educational Research Journal*, 50(1):4–36.

Rothstein, J. (2010). Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement. *Quarterly Journal of Economics*, 125(1):175–214.

Sandler, D. H. and Sandler, R. (2014). Multiple event studies in public finance and labor economics: A simulation study with applications. *Journal of Economic and Social Measurement*, 39(1-2):31–57.

Schoeneberger, J. A. (2012). Longitudinal Attendance Patterns: Developing High School Dropouts. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 85(1):7–14.

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

Sheldon, S. B. and Epstein, J. L. (2004). Getting Students to School: Using Family and Community Involvement to Reduce Chronic Absenteeism. *School Community Journal*, 14:39–56.

Simon, N. S. and Johnson, S. M. (2015). Teacher Turnover in High-Poverty Schools: What We Know and Can Do. *Teachers College Record*, 117(3):1–36.

Urick, A. and Bowers, A. J. (2014). What Are the Different Types of Principals Across the United States? A Latent Class Analysis of Principal Perception of Leadership. *Educational Administration Quarterly*, 50(1):96–134.

Walsh, E. and Dotter, D. (2019). The Impact of Replacing Principals on Student Achievement in DC Public Schools. *Education Finance and Policy*, pages 1–53.

Witziers, B., Bosker, R. J., and Krüger, M. L. (2003). Educational Leadership and Student Achievement: The Elusive Search for an Association. *Educational Administration Quarterly*, 39(3):398–425.