UNIVERSITY OF CALIFORNIA

Los Angeles

Essays in Labor and the Economics of Education

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Economics

by

Calvin Kuo

2024

ABSTRACT OF THE DISSERTATION

Essays in Labor and the Economics of Education

by

Calvin Kuo

Doctor of Philosophy in Economics

University of California, Los Angeles, 2024

Professor Martha Jane Bailey, Chair

This dissertation consists of two chapters at the intersection of labor and the economics of education. In Chapter 1, I examine the role of management in the public sector, focusing on school principals and how they impact student outcomes. I find that a one standard-deviation increase in principal quality raises average student achievement by about 5% of a standard-deviation corresponding to a $15,000 increase in expected lifetime earnings. To identify effective principals, I construct principal value-added estimates and provide the first evidence that they are forecast unbiased. I use these individual-level estimates to examine the correlates of value-added and the mechanisms through which principal effects operate. I find that test-score boosting principals are stronger at recruiting and retaining their strongest staff members, even though they cannot offer teachers higher salaries, and are more likely to assign their best teachers to larger classrooms, increasing overall student learning. I argue that this finding reflects differences in soft skills and management practices as I document a robust positive relationship between value-added and various measures of leadership and teacher empowerment. In Chapter 2, I study how changes in job security affect worker selection and effort in the context of tenure removal for public school teachers. My analysis reveals that in the years after North Carolina eliminated tenure, newly hired teachers had smaller effects on student math and reading achievement as measured by value-added. These effects are

driven by selection into the workforce, as teacher observable characteristics remained largely unchanged. Additionally, I find that tenure receipt is not associated with changes in effort, as value-added in the years after tenure parallels the pre-tenure years.

The dissertation of Calvin Kuo is approved.

Natalie Danielle Bau

Felipe M. Gonçalves

Meredith Phillips

Yotam Shem-Tov

Martha Jane Bailey, Committee Chair

University of California, Los Angeles

2024

To my parents, for their unconditional love and support

Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

Throughout my PhD journey, I have received tremendous support from my academic advisors, peers, family, and friends. Academically, I am deeply grateful for the guidance of my committee, led by Martha Bailey and Natalie Bau. Martha's optimism was contagious—even during challenging times, her unwavering support was always there. Natalie's guidance was equally invaluable. Each meeting with her inspired me to push further, as I learned creative ways to tackle difficult problems. Her technical expertise was exceptional, but it was her encouragement and belief in me that made the greatest impact.

Rounding out the committee are Felipe Gonçalves and Yotam Shem-Tov. Both provided continuous guidance, from helping advance my projects to polishing my presentation slides. I value the bond we've built and consider them both as my big brothers. While my project never quite fit with any single advisor, they all provided insightful comments and helped me build confidence during moments of self-doubt.

Next, I have to thank all of my friends in the PhD program. A big thank you to Manu Navjeevan who helped me pass the qualifying exams in the first year and provided much needed levity during stressful times. Many thanks to my classmates in the applied microeconomics group: Sungwoo Cho, Diana Flores, Johnny Huynh, and Joaquin Serrano—I learned so much from each of you and valued your insights. I thank my peers Domenico Fabrizi, Jesper Böjeryd, Seungyub Han, Danny Ober-Reynolds, Fatih Ozturk, Mariano Palleja, and Daniel Perez for adding life and enjoyment to the program. Special thanks to the current and future UCLA alumni. Ariadna Jou, Tomas Guanzirolo[1], and Fernanda Rojas for being great mentors but even better friends. To Maria Elena Guerrero, Vicky Wang, and Huihuang Zhu, I'm grateful for your support and am excited to see your future achievements. I am incredibly proud to have been a part of this great economics community and hope we will keep in touch.

[1] Email me if I have misspelled your name

Lastly, I must thank my family and friends who have supported me throughout this journey. Amid the chaos of the PhD, peace, familiarity, and humor were invaluable. I cannot thank my parents, Yu-Chi Kuo and Hua Shih, enough for their unwavering love and support in my pursuit of becoming a researcher. A big shoutout to my friends who lifted my spirits and provided diverse perspectives: Dr. Chris Boughter, Sean Cinkilic, Ned Daniel, Sean Gallup, Michael G. Manna, Saptak Ray, and Richard Warren. It takes more than a village to "raise an economist", and I appreciate your support throughout my academic journey.

As resident of the California Center for Population Research, I am grateful for the professors and staff for providing invaluable resources. Having access to a spacious desk, printer, and other amenities cannot be understated. Finally, I acknowledge financial support from the Economics Department at UCLA.

VITA

Calvin Kuo

Education

University of California, Los Angeles (UCLA)

M.A. in Economics	2020

University of Maryland, College Park

B.A. in Economics, Departmental Honors	2015
B.S. in Mathematics	2015

Research in Progress

"Principal Quality and Student Outcomes: Evidence from North Carolina"

"The Impact of Tenure Removal on Teacher's Labor Supply Responses"

Relevant Positions

University of California, Los Angeles (UCLA)

Research Assistant for Professor Natalie Bau	2020
Research Assistant for Professor Felipe Gonçalves	2020
Teaching Assistant	2019-2024

Federal Reserve Bank of Boston

Senior Research Assistant for Bo Zhao	2015-2018

Fellowships, Honors, and Awards

Lewis L. Clark Graduate Fellowship Fund, UCLA	2021, 2023, 2024
Doctoral Travel Grant, UCLA	2023
Graduate Summer Research Mentorship, UCLA	2021

Graduate Division Fellowship, UCLA 2018-2023

Presentations

Center for Naval Analyses, Office of the Comptroller of the Currency, Society of Labor Economists Annual Conference 2024

UCLA, All-UC Demography Conference (Poster Session) 2023

University of Michigan Causal Inference in Education Research Seminar, Taiwan Economics Research Workshop, All-California Labor Economics Conference, SOCAE, APPAM

Introduction

Chapter 1 examines the role of management in the public sector, focusing on how public school principals affect student outcomes and shape the composition of teachers. Using a variance decomposition that exploits principal transitions across schools in the North Carolina public school system, I find that differences in principal quality explain approximately 5% of the variation in test scores. To identify effective principals, I construct principal value-added (PVA) estimates and provide the first evidence that they are forecast unbiased. I use these individual-level estimates to examine the correlates of PVA and the mechanisms through which principal effects operate. My results suggest that previous effective teaching strongly predicts subsequent PVA. To elucidate potential mechanisms, I employ an event-study design around principal transitions, finding that more effective principals excel at attracting better teachers and retaining their best staff. Furthermore, they are more likely to assign their schools' best teachers to larger classrooms, which increases overall student learning. School survey data allow me to unpack why effective principals attract and retain high-quality teachers. I document a robust relationship between PVA and various measures of leadership and teacher empowerment, suggesting that test score–boosting principals also possess certain characteristics that make them more appealing supervisors relative to others.

Chapter 2 examines how changes in job protections affect worker selection and effort. I answer these questions by studying a statewide reform in North Carolina that abruptly eliminated tenure paths for recently arrived and newly hired teachers. Using administrative data, I link teachers to classrooms and examine how decreased job security affects teacher impacts on student test scores. I find that productivity does not decline after receiving tenure as teacher value-added parallel the preceding years. Regarding labor supply, I find that teachers entering after the reform were less effective than their older peers, even after controlling for experience, with differences in quality widening over time. Teachers entering after the policy were 4% and 2% of a standard-deviation less effective on math and reading value-added, respectively. I argue that these declines reflect teacher selection as observable characteristics are stable across cohorts.

Chapter 1

Principal Quality and Student Outcomes: Evidence from North Carolina

1. Introduction

The United States spends approximately $870 billion dollars per year in public K–12 education, much of which principals invest in facilities, the hiring of academic staff, and the development and implementation of curricula and programs (NCES, 2020b). Principals' extensive involvement in every aspect of schools' operation——from teacher and student assignments, parent engagement, and the curation of school climate and culture—suggests that they play an important role in all aspects of educational quality. Despite this extensive engagement, principals remain relatively understudied. Quantifying principal quality requires sufficient variation in principal mobility, and uncovering potential mechanisms requires data on classroom assignments to reveal principals' impacts on teachers.

This paper quantifies the share of variation in student test scores due to principals, estimates principal value-added (PVA), provides the first evidence showing the estimates' validity, and correlates PVA with novel data on personnel practices to examine how principals affect students and teachers. I find that a one standard-deviation (SD) increase in principal quality raises average student test scores by 0.047 SD. Using estimates from Chetty et al. (2014b), my results suggest that a one-SD increase in principal quality during middle school raises expected lifetime earnings by $45,825 for one student. Considering that middle school typically lasts three years and serves 575 students (NCES, 2010), the implied returns to principal quality exceed $26 million per *school*—a magnitude far greater than that of the returns to teacher quality since principals affect the entire school and not just individual classrooms.

I then compute and validate individual principal value-added (PVA) to provide three insights as to who are effective principals and how effective principals shape student outcomes. First, a history of prior effective teaching is strongly predictive of future principal success. Second, transitioning to a more effective principal is associated with improvements in average, school-level, teacher quality. Higher value-added principals are better at attracting effective teachers and more likely to keep their best staff. Third, higher PVA is associated with stronger personnel and leadership skills. Above median-PVA principals are more likely to assign their best teachers to larger classrooms which increases overall student learning. Moreover, higher PVA is associated with stronger ratings on overall school climate, leadership, and teacher empowerment based on working conditions surveys.

For these exercises, I utilize rich administrative data from the North Carolina Education Research Data Center (NCERDC) which contain records on 2.5 million students and 6,000 principals from 1996 to 2019. With the long panel, I observe over 2,000 principals moving across schools allowing me to disentangle principal effects from other confounders. Unique to my setting, the data contain survey records asking faculty various questions related to principal leadership, teacher empowerment, and overall school climate. These responses can then be linked to principals, allowing me to examine the role of soft skills and PVA.

To quantify principal effectiveness, I extend Chetty et al. (2014a), Araujo et al. (2016), and Bau and Das (2020) and allow student achievement to be a function of teacher, school, principal, and classroom effects where identification of principal effects utilizes the movement of principals across schools. I then estimate PVA and provide two tests showing that value-added indeed captures variation in principal quality and does not reflect selective sorting by students. First, I follow Bau and Das (2020) and Andrabi et al. (2022) and exploit students' switching of schools to examine whether a student's new principal fully predicts changes in her test scores. I show that future principal quality is not predictive of student achievement prior to the principal's arrival, suggesting that students are not sorting on unobserved characteristics. I then confirm that PVA fully predicts student test score gains in the year of the switch, indicating that value-added is forecast unbiased. Second, I use principal entry and exit to examine whether changes in principal quality predict changes in mean school test scores. I find, consistent with the previous test, that PVA is forecast unbiased, as the coefficient on test score changes is not statistically different from one in the year of the switch and that future principal quality does not predict lag score changes.

Having shown the validity of PVA methodologies, I use my estimates to document the correlates of value-added and whether greater effectiveness translates to higher wages for principals. I find that observable characteristics such as average experience, education quality, and highest degree obtained, explain less than 4% of the within-district variation in principal effectiveness. However, I find that principals with a history of effective teaching are more likely to be stronger principals. This result mitigates concerns regarding the efficiency of promoting the best teachers (e.g., the "Peter Principle"; see Benson et al., 2019) and has important screening implications, as over 95% of principals in the United States have prior teaching experience (National Teacher and Principal Survey 2021). Regarding compensation, I find that value-added is weakly correlated with salary, an unsurprising result as public officials' salaries typically follow an experienced-based pay scale. However, as more affluent school districts offer larger bonuses, my results suggest that more-effective principals sort into less disadvantaged districts, potentially exacerbating inequities in access to quality schooling.

To examine the mechanisms through which principal effects operate, I use an event-study design where a principal transition across schools is the "event". I show that a 1-SD increase in PVA implies that the average student has a 0.17-SD higher value-added teacher, with most of these gains coming from teacher recruitment. Beyond influencing the quality of new teachers, effective principals also reduce overall turnover and, more importantly, increase the retention of their best teachers. A 1-SD increase in principal effectiveness reduces the job separation rate of teachers with above-median value-added by 6.4% relative to that of teachers below the median. My results contribute to a growing discussion on whether principals can identify effective teachers and I show that differences in principal quality may help reconcile the mixed existing evidence (Jacob and Lefgren, 2005; Hinrichs, 2021; Bates et al., 2022).

A final set of analyses investigate how principal quality relates to personnel management and leadership. Like most public sector officials, principals cannot easily adjust salaries or alter contracts. This limitation suggests that stronger personnel management could be key to maximizing student outcomes and that leadership and empathy are important traits of principals. I show that, relative to those with below-median value-added, more effective principals assign their best teachers to larger classrooms, thereby increasing overall student learning. These effects persist even when I account for school–year-level shocks, suggesting differences in personnel management skills across principals. I then utilize the NCERDC's unique survey data to understand whether test score–boosting principals also possess stronger soft skills. I show that, across a variety of survey questions, transitioning to a higher-quality principal is associated with more teachers agreeing that school leadership (1) is effective and supports teachers, (2) empowers teachers and values their opinions, and (3) creates an engaging school culture.

This paper contributes to the literature in three ways. One puzzle in the literature on principals is the difference in magnitude between principal and school effects since the existing estimates of principal effects (Branch et al., 2012; Chiang et al., 2016; Dhuey and Smith, 2018; Bartanen, 2020), are consistently much larger than those of school effects which include principal impacts (Angrist et al., 2017, Jackson et al., 2020, Angrist et al., 2021). My variance estimates are one-third the size of most prior estimates of principal quality and are smaller than school effects, as I properly account for sampling error. This discrepancy between my estimates and previous studies arises since the common Krueger and Summers (1988) estimator to correct for sampling error underperforms in high-dimensional settings[1].

Second, this study is one of the first to show that PVA estimates are forecast unbiased, extending the applicability of value-added methodologies (see Chetty et al., 2014a, Jackson, 2018, and Bau and Das, 2020 for teachers; Angrist et al., 2017 and Angrist et al., 2021 for schools; and Mulhern, 2020 for school counselors) and allowing policymakers to evaluate principal impacts on test scores without conflating them with the school or teacher effects.

[1]See Kline et al. (2020) for a formal discussion and details on the jack-knife estimator.

Third, this study unpacks the black box of practices of effective public-sector officials. Recent work by Fenizia (2022) suggests that public-sector managers influence output by inducing older workers to exit, but it is unclear whether these workers are indeed less productive than others or are misallocated to tasks. I extend this analysis by using a clear metric of worker productivity, allowing me to examine task allocation and changes in worker composition. In addition, my analysis complements work linking school management and student academic outcomes by showing that school management is driven primarily by differences in principal quality (Bloom et al., 2015; Lemos et al., 2021). As workers value leadership and a supportive work environment (Bates et al., 2022; Maestas et al., 2023), my results suggest that differences in these skills may explain why effective principals attract and retain the best teachers.

The remainder of the paper is organized as follows. Section 2. describes the NCERDC data. Section 3. discusses the framework for quantifying the distribution of principal effects. Section 4. estimates and validates PVA. Section 5. examines the relationship between value-added and observable principal characteristics, wages, and where effective principals tend to work. Section 6. examines the mechanisms through which principal effects operate, including their relationship to leadership skills. Section 7. concludes.

2. Data and Sample Description

2.1 Data Description

I use administrative microdata from the North Carolina Education Research Data Center (NCERDC) to examine the impacts of principal quality on student outcomes. The data contain detailed information on the near universe of students enrolled in North Carolina public schools from 1995 to 2019, including background information on principals and teachers, the schools where the principals and teachers worked, assignments to classrooms, and school climate surveys of teachers and staff members.

Like other administrative datasets, the NCERDC dataset contains student-level variables, including end-of-grade test scores in math and reading and demographic information such as race and ethnicity, gender and indicators for whether a student is eligible for free or reduced-price lunch, academically gifted status, and whether a student repeated a grade or course. The data also allow me to link teachers to students to form classrooms, construct classroom-level controls, and estimate teacher value-added. Relevant staff-level data include demographics, experience, highest degree obtained and degree-granting institution, and total compensation. The end-of-year climate surveys ask staff members a variety of questions relating to school leadership, teacher empowerment, and overall school climate. These survey results can be linked to principals allowing me to examine whether test score–boosting principals also possess stronger leadership traits and how these skills might affect the teacher composition[2].

2.2 Sample

My analysis focuses on principals and students in public elementary and middle schools (grades 4–8), as principals usually have more influence on teachers and students in these settings. Elementary and middle school principals are also less likely than high school principals to rely on assistant principals or supporting staff, whose influence could mute or amplify principals' true effect.

Certain features of the data lead me to impose several additional sample restrictions. First, I drop cases where a principal oversees multiple schools in a given academic year

[2] Section 6.3 describes the survey data in more detail.

(3.4% of all school observations). Second, in circumstances where a school's principal data are missing, I use the previous year's primary principal (if one exists) to proxy for the principal in the current year. With this method and restriction, I am able to identify nearly 6,600 principals covering 2,100 schools, where less than 2% of the principals are identified using the proxy method.

To identify classroom assignments, I follow Jackson (2018) and Rose et al. (2022) and use the NCERDC provided "Course Membership" files, which directly link teachers to students for academic years 2007–2019. For the remaining years from 1995 to 2006, I follow Rothstein (2017) and use the provided "End-of-Grade" files, which directly link students to end-of-year testing proctors as proxies for the student's true teacher.

After matching teachers to students, I impose additional sample restrictions. I limit my sample to students with valid test scores in the prior year (prior-year test scores are key controls in my construction of value-added). To mitigate potential mismatches of teachers to students, I limit the sample to classrooms with 15–100 students (Rose et al., 2022). Finally, I drop observations in which teachers teach at multiple schools or in multiple grades in a year as their effects since their students are only partially exposed to their effects or may capture the impacts of substitute teachers. After I impose these restrictions, my estimation sample consists of 11,624,281 student–subject–year observations.

2.3 Sample Description

The first two columns of Table 1.1 provide summary statistics for my analytic sample. The sample covers approximately 2.6 million students, 6,600 principals, and 2,100 schools. The principals in my sample typically serve for four years and are more likely to be white than nonwhite. These descriptive characteristics are consistent with those reported in other studies examining principal impacts on student achievement (Branch et al., 2012, Bartanen et al., 2022). The principals in the sample are also slightly more likely to be women. Women's representation in education leadership is slightly larger than that in other public-sector occupations and vastly greater than that in the private sector (Beaman et al., 2009), but it remains lower than one might anticipate given that elementary and middle school teachers—the pool of potential principals—are disproportionately female. Having an advanced degree is one of the pathways to becoming a principal, and nearly all principals in the sample have at least a master's.

Turning to school-level characteristics, the typical school undergoes nearly five principal transitions, consistent with the data period spanning 25 years and the typical principal serving approximately four years. Approximately 38% of the student population is Black (26%) or Hispanic or Latinx (12%), 52% are eligible for free or reduced-price lunch, and approximately 5% are classified as having limited English proficiency.

Since disentangling principal from school effects requires the same principal to be observed at multiple schools, columns (3) and (4) report summary statistics for principals with employment histories at multiple schools. Principals who have worked at multiple schools are comparable to the average principal in terms of demographics and school quality. Movers do have more experience and are slightly more likely to hold a doctorate degree but have lower annual compensation than the average principal. While the descriptives are similar in magnitude, it is important to note that, for most observable characteristics, the difference between movers and the average principal is statistically different. Finally, the school characteristics for principal movers largely parallel those of the overall sample.

3. Variance Decomposition

I divide the task of quantifying principal effects into three parts. Section 3.1 provides the theoretical framework on the underlying process of student achievement and the interpretation of principal effects. It then describes how to estimate the main parameters from observational data. Section 3.2 reports the results and discusses their magnitude and implications for student outcomes. Finally, Section 3.3 benchmarks my estimates against results in the existing literature and discusses why they might differ.

3.1 Theoretical Framework and Model Setup

To quantify the importance of principal quality, I extend Araujo et al. (2016) and Bau and Das (2020) and allow student achievement to be a function of teacher, school, principal, and classroom effects and observable and unobservable student heterogeneity:

$$Y_{i,s,j,p,t} = \theta^j + \theta^s + \theta^p + \theta^{j,s,p,t} + X'_{i,t}\tau + \epsilon_{i,s,j,p,t} \tag{1.1}$$

where i denotes a student, s denotes a school, j denotes a teacher, p denotes a principal, and t denotes a year. $Y_{i,s,j,p,t}$ are end-of-year test scores for either math or reading and are standardized at the grade–year level. θ^j is a time-invariant teacher-specific effect, θ^s is a time-invariant school effect, θ^p is a time-invariant principal effect, and $\theta^{j,s,p,t}$ is the classroom-specific effect for teacher j in school s in year t. $X'_{i,t}$ are observable student characteristics, while $\epsilon_{i,s,j,p,t}$ reflects unobserved student–year-level shocks[3].

The parameters of interest are σ^2_j, σ^2_s, σ^2_p, and $\sigma^2_{j,s,p,t}$, corresponding to the variances of teacher, school, principal, and classroom effects. By construction, equation 1.1 does not

[3]Superscripts on the θs are used to emphasize their place as structural parameters.

account for complementarities between each of the causal θ parameters, ruling out potential match effects between teachers and principals. Equation 1.1 provides the framework for understanding how a variety of classroom and school forces affect student achievement. To map this to observable data, I define “observational" effects as the population projection of equation 1.1:

$$Y_{i,s,j,p,t} = \sum D_{i,t}\delta_{j,s,p,t} + X'_{i,t}\tau + \mu_{i,s,j,p,t} \tag{1.2}$$

where $D_{i,t}$ is an indicator for whether student i was assigned to a specific classroom in year t. $\delta_{j,s,p,t}$ are classroom fixed effects that subsume teacher-, school-, principal-, and classroom-level shocks for a given year[4]. The estimated classroom effects can then be used to quantify the variances of interest under the following assumptions (McCaffrey et al., 2009). First, if $E[D_{i,t}\epsilon_{i,s,j,g,p,t}] = 0$ holds, implying that classroom assignments are uncorrelated with unobserved student–year-level shocks, then the observational framework maps directly to equation 1.1. Second, if $cov(\theta_m, \theta_{-m} = 0, \forall m)$, then the covariance of classroom effects across different years and agent allows me to identify σ_j^2, σ_s^2, σ_p^2, and $\sigma_{j,s,p,t}^2$ without estimating the individual θ parameters. I discuss each assumption in greater detail below.

Assumption 1. *Conditional Independence:* $E[D_{i,t}\epsilon_{i,s,j,g,p,t}] = 0$

Conditional independence requires that classroom assignments be uncorrelated with unobserved determinants of student achievement. Violations of Assumption 1 would occur if students who learn faster are systematically assigned to particular teachers.

To ensure that Assumption 1 is satisfied, I use vector $X'_{i,t}$, which contains a rich set of student covariates commonly used in estimating teacher value-added (Rockoff, 2004; Kane

[4]One may interpret $\delta_{j,s,p,t}$ as the average effect of the school, teacher, principal, and classroom composition or classroom-specific resources.

et al., 2008; Rothstein, 2010, Chetty et al., 2014a): limited English proficiency status, free or reduced-price lunch status, indicators for whether a student is academically gifted or has repeated a grade or subject, race, gender and, most importantly, lagged student test scores interacted by grade. I also include school-level means of each of these characteristics. Prior-year test scores are particularly important, as they have been shown to address teacher sorting (Chetty et al., 2014a) and within-school sorting of students to classrooms (Jackson, 2018)[5], strengthening the likelihood that the conditional independence assumption is satisfied. If values are missing for a particular student, then the variable is equal to zero, and an indicator denoting missing is included. Additionally, grade and year fixed effects are included to control for grade-specific shocks (e.g., changes in testing for a particular grade) and temporal shocks.

Assumption 2. *Uncorrelated Effects: $cov(\theta_m, \theta_{-m} = 0, \forall m)$*

Assumption 2 implies that, conditional on the rich set of controls, teachers and principals do not differentially sort into particular schools. This means, for example, that highly effective principals are not more likely to sort into particular types of schools, conditional on our accounting for covariates. This assumption is supported across a variety of settings, including North Carolina, as prior work notes that teachers have strong preferences for working in schools with fewer disadvantaged students and that in general, more affluent districts can attract stronger academic staff, which I control for (Branch et al., 2012; Biasi et al., 2021; Bates et al., 2022) .

To estimate the variance of principal effects, I consider the covariance of classroom effects for principal p who works at school s in year t with teacher j with classroom effects for the

[5]Chetty et al., 2014a find that 85% of the variation in teacher quality is within rather than between schools.

same principal who moves in year t' to a different school, s' and has a different teacher j' : $cov(\delta_{j,s,p,t}, \delta_{j',s',p,t'})$. Under Assumptions 1 and 2, the covariance can be rewritten as

$$cov(\delta_{j,s,p,t}, \delta_{j',s',p,t'}) = cov(\theta_j + \theta_s + \theta_p + \theta_{j,s,p,t}, \theta_j' + \theta_s' + \theta_p + \theta_{j',s',p,t'}) \tag{1.3}$$

$$= \sigma^2_{\theta_p} + cov(\theta_p, \theta_{j'}) + cov(\theta_p, \theta_{s'}) + \dots \tag{1.4}$$

$$= \sigma^2_{\theta_p} \tag{1.5}$$

Using this result, we can apply similar logic to obtain the variance of school effects. Comparing the variance of classroom effects for the same principal p at school s but with different teachers yields $cov(\theta_j + \theta_s + \theta_p + \theta_{j,s,p,t}, \theta_{j'} + \theta_s + \theta_p + \theta_{j',s',p,t'}) = \sigma^2_{\theta_s} + \sigma^2_{\theta_p}$. Solving for $\sigma^2_{\theta_s}$ implies that $\sigma^2_{\theta_s} = cov(\delta_{j,s,p,t}, \delta_{j',s,p,t'}) - cov(\delta_{j,s,p,t}, \delta_{j',s',p,t'})$. Applying a similar logic implies that the variance of teacher effects can be obtained by calculating $\sigma^2_j = cov(\hat{\delta}_{j,s,p,t}, \hat{\delta}_{j,s,p,t'}) - cov(\hat{\delta}_{j,s,p,t}, \hat{\delta}_{j',s,p,t})$. Finally, to identify the variance of classroom effects, note that $var(\hat{\delta}_{j,s,p,t}) = \sigma^2_{j,s,p,t} + \sigma^2_s + \sigma^2_p + \sigma^2_j + \phi$, where ϕ is the sampling error that occurs since we observed estimated $\hat{\delta}_{j,s,p,t}$. To correct for sampling bias, I follow Bau and Das (2020) and assume that the student residuals are homoskedastic to create a closed-form solution for ϕ[6].

The Conditional Independence and Uncorrelated Effects assumptions provide identification of the variances of interest. To show that my estimates are robust, Appendix 1.2 provides an alternative decomposition that identifies the variance of principal effects under weaker assumptions. Following Kane et al. (2008), Chetty et al. (2014a), Kline et al. (2020), and Rose et al. (2022), I assume that cohort or school shocks are uncorrelated over time, and construct mean principal-year-level residuals to quantify the variance of principal effects. This approach yields nearly identical estimates. Additionally, since identification of $\sigma^2_{\theta_p}$ rests on principals who work across multiple schools, a natural question is what deter-

[6]See Appendix 1.1 for full details.

mines principal mobility. Appendix Section 1.5, explores this in greater detail, though note that, even if trends in academic achievement predict principal turnover, this would not bias the estimated variances, which capture average changes in classroom achievement associated with principal switchers.

3.2 Results: Magnitudes and Implications

Table 1.2 reports the estimated effects of a one-SD improvement in classroom, school, principal, and teacher effects on student test scores in math, reading, and the average of these two subjects. I find that a one-SD increase in principal quality raises student math and reading achievement by 0.05 and 0.043 SD, respectively; these estimates are approximately half the size of the Black–white testing gap after kindergarten (Fryer Jr and Levitt, 2004). Under standard normality assumptions, these estimates imply that a rise in principal quality from the fifth percentile to the ninety-fifth percentile would raise average student achievement by approximately 0.15 SD.

Figure 1.1 provides a visual comparison of how my estimates compare to some from the existing literature examining principal, school, and teacher effects. In general, my estimated principal effects are markedly smaller than existing estimates of principal quality, as small as 87% of those in Dhuey and Smith (2014). They are also smaller than those in prior work examining school effects (0.06 SD, Angrist et al., 2017; Jackson et al., 2022) which is consistent since these estimates do not disentangle the impact of schools from principals. If anything, my estimates of the variance of principal effects suggest that prior estimates of school quality largely reflect underlying differences in principals. To demonstrate this, I omit θ_p from equation 1.1 and then estimate its population analogue. I find that a one-SD increase in school quality raises average student achievement by 0.058 SD, up by 0.011 from the estimate in Table 1.2, which suggests that over one-third of the observed variation in school effects, in North Carolina, is driven by differences in principal quality with potentially larger differences in other settings[7]. This finding has important implications since household sorting patterns are partially influenced by school quality (Bayer et al., 2007; Agostinelli et al., 2021) and suggests that the potential benefits from a student's attending a certain school may be temporary if the principal exits the school.

While principals' effects on individual students are modest, principal quality remains economically important since they oversee entire schools[8]. When I use estimates from Chetty et al. (2014b), the financial value of having a principal with one-SD higher value-added is approximately \$2,740 per year, equivalent to a lifetime earnings gain of \$15,275[9]. Given that the average principal tenure is approximately four years and that students are exposed to the same principal for multiple years, having a high-value-added principal has major consequences for long-run student outcomes. For instance, a principal with one-SD higher value-added would raise lifetime earnings by \$45,825 (\$15,275 $\times$ 3) since middle school education typically lasts three years in this sample[10]. When I extend this calculation to the entire school, the implied returns to principal quality exceed \$26 million, a magnitude far greater than the estimated returns to teacher quality.

3.3 Comparison to Existing Literature

Existing work on principal quality tends to find that a one-SD increase in principal quality raises student test scores by at least 0.13 SD (Dhuey and Smith, 2014; Grissom et al., 2015; Dhuey and Smith, 2018; Bartanen, 2020). These estimates are considerably larger than my

[7]To arrive at this number, I calculate $(1 - \frac{0.047^2}{0.058^2})$, where the numerator is the explained *variance* of test scores from the original estimates.

[8]Grissom et al. (2021) note that the average elementary school has 483 students.

[9]To arrive at this number, I note that, in Chetty et al. (2014a), a one-SD increase in teacher quality raises average student achievement by 0.12 SD. I then scale the estimates in Chetty et al. (2014b) by 0.047/0.12 to gauge how principals affect student lifetime earnings. I calculate lifetime earnings assuming a 2% growth rate and a 0% discount rate as in Chetty et al. (2014b)

[10]In the United States, middle schools typically have three grades and serve students aged 11 through 14.

findings as well as results from prior work on school effects. To quantify principal effects, these papers estimate a regression similar to equation 1.6 but also include school fixed effects, and correct for sampling error using the estimator proposed by Krueger and Summers (1988)[11]. Recent work by Kline et al. (2020) discusses the difficulty of using this estimator in two-way fixed-effect (TWFE) models and in particular, how in high-dimensional settings (e.g., when the number of principals and schools grows with the sample size), the estimator may underperform.

To illustrate this methodological limitation, I also report estimates using the estimation approaches of Dhuey and Smith (2018) and Bartanen (2020), who use the Krueger and Summers (1988) sampling error estimator[12]. With this estimator, a one-SD increase in principal quality raises student test scores by nearly 0.222 and 0.132 SD for math and reading; these are closer to the original estimates (0.172 and 0.117 SD for math and reading) found by Dhuey and Smith (2018), who use a subset of the same NCERDC dataset[13]. The large difference in magnitude highlights the challenge of using the Krueger and Summers (1988) estimator to adjust for sampling error when simultaneously estimating both the principal and school fixed-effect. I obtain smaller estimates on the variance of principal effects since the sampling error is at the classroom-level and not at the principal-level.

[11] Similar estimators have been used to estimate the variance of teacher effects, as in Jacob and Lefgren (2005) and Aaronson et al. (2007).

[12] Specifically, I first estimate $y_{i,t} = \sum_a \beta_a y_{i,t-1}\mathbf{I}\{\text{g} = a\} + X'_{i,t}\tau + X'_{s,t}\tau + \theta_p + \theta_s + \epsilon_{i,t}$. I then obtain the standard deviation of the "true" principal effects by calculating $\sqrt{var(\hat{\theta}_p) - \frac{1}{N_p}\sum_p \hat{\sigma}_p^2}$, where $\hat{\sigma}_p$ is the standard error of the *individual* principal fixed-effect estimates and N_p is the total number of principals. The second term under the square root is similar to the sampling correction term proposed by Krueger and Summers (1988).

[13] Their analysis consists of test score records spanning 1995–2011

4. Principal Value-Added

Having shown that principal quality can significantly shape the outcomes of an entire school, a natural question is *who* effective principals are and *how* they influence student outcomes. Answering these questions requires constructing individual-level estimates of principal quality. To this end, I estimates PVA to quantify each principal's impact on test scores and provide a battery of validation checks that indicate that these estimates are forecast unbiased. After establishing the validity of my PVA estimates, I document the correlates of principal effectiveness by examining what observable characteristics predict principal quality and assess whether higher value-added is correlated with higher wages. Section 4.1 describes the empirical model used to estimate individual-level principal effectiveness. Section 4.2 examines whether my estimates of value-added are valid and unbiased. Section 5.1 documents which observable characteristics are predictive of principal effectiveness. Finally, Section 5.2 details whether principal quality is correlated with wages.

4.1 Estimating Principal Value-Added

To estimate PVA, I follow the teacher value-added literature and estimate value-added as a time-invariant fixed effect (Harris and Sass (2006) Kane et al. (2008) Chetty et al. (2014a)). I estimate the equation below for each subject separately:

$$y_{i,g,p,t} = \sum_a \beta_a y_{i,t-1} \mathbf{I}\{\text{grade} = a\} + X'_{i,t}\delta + X'_{j,s,t}\tau + \alpha_t + \alpha_g + \theta_p + \epsilon_{i,g,p,t} \tag{1.6}$$

where $y_{i,g,p,t}$ are contemporaneous test scores and are standardized at the grade–year level. Following the teacher literature, the vectors $X'_{i,t}$ and $X'_{j,s,t}$ are the set of student characteristics, as defined in equation 1.2, and their corresponding school- and classroom-level means[14].

[14] See Chetty et al. (2014a) and Rose et al. (2022), who use school-level means to account for students' sorting across schools. The argument follows that in Altonji and Mansfield (2018).

These controls, along with lagged student test scores $y_{i,t-1}$, help address students' sorting into specific schools or classrooms. α_g a grade fixed effect and α_t a year fixed effect. θ_p is the principal fixed effect.

$\hat{\theta}_p$ is the estimate of a principal's value-added and reflects the average test score gains of a student assigned to principal p conditional on observable characteristics. This estimate is unbiased if students do not sort into assignment to specific principals based on unobservable characteristics. It is important to note that $\hat{\theta}_p$ reflects both the average effect of principals on student achievement and the independent teacher effect. The decision to omit teacher fixed effects is motivated by the following. First, estimating principal effects is difficult in and of itself since there are relatively fewer principals than teachers. Second, adding teacher fixed effects may significantly bias estimates of PVA unless principal mobility is uncorrelated with time-varying residual components of test scores[15].

Even if the estimates of principal value-added are unbiased, $\hat{\theta}_p$ still contains sampling error. This implies that when using PVA is an explanatory variable, its coefficient will be attenuated. To address this, I construct empirical Bayes estimates of value-added using the variances from Table 1.2, with full details described in Appendix 1.3. When value-added is the outcome variable, I use the principal fixed effects from estimating equation 1.6.

Throughout this analysis, I focus on the mean PVA across math and reading. An exception is when the outcome variable relates to teacher quality, where I use the PVA corresponding to math scores. For instance, when examining principal effects on teacher composition as well as differences in attrition rates by teacher quality, I use value-added on math scores. This occurs since, in middle school, students typically have different teachers for different

[15] See Card et al. (2018) for a detailed discussion.

subjects, making it difficult to use the mean teacher value-added. To ease interpretation, I use math value-added for both teachers and principals[16].

4.2 Validating PVA

One can use estimates of principal value-added to identify effect principals and learn how they shape student outcomes. However, estimates of PVA are informative so long as they are indeed predictive of test score gains and are unbiased. I first show that bias from students' sorting on twice-lagged test scores is minimal, an important concern as Rothstein (2010) finds significant scope for bias using a subset of the same North Carolina data. I then implement two out-of-sample prediction tests in the spirit of Bau and Das (2020) and Chetty et al. (2014a) to show that PVA predicts test scores when students move to a new school and when a principal moves to a different school. These exercises test for forecast unbiasedness and allow me to check whether students are sorting on unobservable characteristics.

Value-Added Prediction and Omitted Observables

I begin with an out-of-sample prediction of PVA and student test scores. Let $\hat{\theta}_p^{-t}$ denote the estimate of θ_p that is estimated excluding data from year t. If value-added is an unbiased predictor of student test scores, then there should be a one-to-one relationship between $\hat{\theta}_p^{-t}$ and test scores in year t ($y_{i,t}$). I test this relationship by regressing $y_{i,t}$ against $\hat{\theta}_p^{-t}$ and including subject (math versus reading) by school type (elementary versus middle) fixed effects.

Panel A of Figure 1.2 plots the conditional expectation between $y_{i,t}$ and $\hat{\theta}_p^{-t}$ after residualizing both variables on the subject–school type fixed effects. The binned scatter plot is divided into 40 equally sized bins of $\hat{\theta}_p^{-t}$ with the mean value of $y_{i,t}$ plotted in each bin. The coefficient estimate corresponds to the linear regression utilizing the student-level data, and

[16] The results are robust to my using teacher and PVA for reading.

standard errors are clustered at the school level. The conditional expectation reveals a linear relationship between $y_{i,t}$ and $\hat{\theta}_p^{-t}$, with a slope estimate of 0.918, which suggests that PVA is indeed highly predictive of student test scores.

Since Rothstein (2010) finds that students sort on twice-lagged test scores, I examine whether the observed relationship between $y_{i,t}$ and $\hat{\theta}_p^{-t}$ is driven by such sorting patterns. This exercise focuses on the subsample of students with valid twice-lagged scores, which primarily eliminates students in fourth grade from the analysis. To assess the degree of bias, I predict test scores in year t based on performance in year $t-2$ and examine whether these predicted values are correlated with $\hat{\theta}_p^{-t}$. The process begins by first residualizing $y_{i,t}$ and $y_{i,t-2}$ on the same set of controls variables described in equation 1.6. After calculating these residuals, I obtain fitted values from a regression of the residualized $y_{i,t}$ on the residualized $y_{i,t-2}$. Finally, I regress the predicted values on $\hat{\theta}_p^{-t}$.

Panel B examines the relationship between predicted test scores using $y_{i,t-2}$ against $\hat{\theta}_p^{-t}$, where each bin plots the mean value of the predicted test scores. The regression coefficient is -0.030 with a standard error of 0.002, indicating that, in magnitude, the upper bound of the 95% confidence interval attributable to the bias from twice-lagged test scores is 0.046. Since the estimates of PVA may also capture school effects, panel C plots the same relationship but for the subset of principals who are observed at multiple schools to partial out these influences. The magnitude of this estimate is even smaller than that of the estimate in panel B, at -0.022, with an upper bound on the 95% confidence interval of -0.042. In line with Chetty et al. (2014a), the bias arising from students' sorting on twice-lagged scores is small, as the baseline controls in equation 1.6 capture much of the variation from $y_{i,t-2}$. Broadly, the magnitude of bias in my setting is larger than, although comparable to, that of the result of Chetty et al. (2014a), who find upper-bound effects of 0.026 when examining teacher value-added.

Bias from Student Sorting on Unobservables

Having shown that PVA is highly predictive of student test scores and is not driven by students' sorting on twice-lagged test scores, I assess whether PVA reflects bias from unobservable student characteristics.

For the first exercise, I focus on students who switch schools and assess whether their test scores can be fully explained by their new principals, conditional on lagged test scores, observable student characteristics, and other school characteristics. I estimate an event-study regression at the student–year level where, at time $t = 0$, a student switches to a new school with a different principal:

$$y_{i,s,t} = \alpha_0 + \sum_{\tau \neq -1} \beta_\tau \hat{\theta}_p^{EB} + \mathbf{X}_{i,s,t} + \alpha_g + \alpha_t + \alpha_s + \epsilon_{i,s,t} \tag{1.7}$$

where $\hat{\theta}_p^{EB}$ is the mean value-added across math and reading of the principal at time $t = 0$ (the future principal), scaled by the shrinkage factor discussed in the previous section. To break the mechanical correlation between $\hat{\theta}_p^{EB}$ and student test scores at $t = 0$, I follow Chetty et al. (2014a) and construct leave-out estimates of PVA omitting data corresponding to the time of the switch. $\mathbf{X}_{i,s,t}$ is the vector of student and school controls as specified in equation 1.6, α_g is a grade fixed effect to account for potential differences in testing across grades, α_t is a year fixed effect to account for common shocks across time, and α_s is a school fixed effect for the school that the student attends at $t = 0$.

Since the transition from elementary to middle school represents the vast majority of student school switches, including multiple periods prior to the school switch is infeasible. This is because students transition to middle school after completing fifth grade and lagged test scores are available only from fourth grade and after. To address this limitation, equation 1.7 includes two preperiod indicators for different periods prior to the student switch. The first is an indicator for the year immediately prior to the switch, $\tau = -1$, which typically corresponds to the year when students are in fifth grade. The second is an indicator for *at least* two years prior to the switch, $\tau = -2$, which typically corresponds to when students are in fourth grade.

The parameters of interest are β_τ, which capture the effect of a one-SD increase in PVA on mean student test scores. The estimates are normalized relative to $\tau = -1$, the year prior to the move. If the PVA estimates are unbiased, then $\hat{\beta}_0$ should equal 1 in expectation. Furthermore, if $\mathbf{X}_{i,s,t}$ sufficiently controls for student sorting, then future principal quality should not predict prior test scores. In other words, β_τ should be indistinguishable from zero for $\tau = -2$[17].

Figure 1.3 provides a visual representation of the estimates from equation 1.7. Consistent with forecast unbiasedness, PVA fully predicts student student test score gains in the year of the student's move. The null hypothesis testing β_0 equals unity cannot be rejected at the α of 5% level. Moreover, future PVA has no impact on prior student achievement. The point estimate for the indicator of two or more years prior to the switch is indistinguishable from zero which strengthens the evidence for the underlying assumption that students are not sorting based on unobservable characteristics.

Table 1.3 reports the regression estimates and examines their sensitivity to the omission of certain controls. Column 1 is the baseline specification that uses the full set of student

[17]This test is similar to the logic behind the parallel trends assumption in the difference-in-differences literature.

and school controls and corresponds to the analysis in Figure 1.3. Column 2 omits student and school controls, but still includes lagged-test scores. Even without these controls, future principal quality fully predicts student-test scores at the time of the switch and still cannot explain prior achievement. Column 3 estimates the most parsimonious version of equation 1.7 by further omitting school fixed-effects. This omission is crucial. Failing to account for school-effects overstates the impact of future principal quality on student test scores as estimated principal effects absorb the influence of schools. The estimate for principal impacts at the time of the switch is statistically different (larger) than one suggesting that $\hat{\beta}_0$ captures both principal and school forces. Combined Table 1.3 provides evidence that PVA is forecast unbiased and assuages concerns that observed principal effects are driven by differences in unobserved student sorting patterns.

As an additional test, I follow Chetty et al. (2014a) and examine whether changes in principal quality predict changes in average student test scores. Since $\hat{\theta}_p$ may be influenced by time-invariant school-level shocks, I focus on the subsample of principals who move to disentangle potential school effects from PVA. Furthermore, since movers have more experience, on average, than nonmovers, focusing on this subsample allows me to obtain more precise estimates of $\hat{\theta}_p$. I estimate the following regression at the school–year-level focusing on year t in which a principal switch occurs:

$$\Delta Y_{s,t} = \beta_0 + \beta_1 \Delta Q_{s,t} + \alpha_s + \alpha_t + \epsilon_{s,t} \tag{1.8}$$

where $\Delta Y_{s,t} = Y_{s,t} - Y_{s,t-1}$, capturing changes in average student test scores at school s in a given year t. $\Delta Q_{s,t} = \hat{\theta}^{EB}_{p,in} - \hat{\theta}^{EB}_{p,out}$ captures the difference in shrunken PVA between the incoming and outgoing principals and each of the individual estimates. Similarly those in to the previous exercise, the constructed PVA estimates are mechanically correlated with the outcome variable since value-added is constructed with the same set of test scores. To break this correlation, $\hat{\theta}^{EB}_{p,in}$ and $\hat{\theta}^{EB}_{p,out}$ are estimated without data for years t and $t-1$, respectively.

The first three columns of Table 1.4 report the regression results from equation 1.8. While limiting the analysis to principal switchers reduces the scope for school effects, the restriction leads to a smaller sample size and hence more imprecise estimates. Nonetheless, across both subjects and their pooled means, I find evidence that changes in principal quality are highly predictive of test score gains. The null hypothesis that $\beta_1 = 1$ cannot be rejected at most conventional levels, affirming the event-study estimates indicating that PVA is indeed forecast unbiased despite the greater imprecision of the estimates relative to that of the estimates in the previous validity check. Columns 4 through 6 run a placebo test and examine whether changes in principal quality predict changes in average test scores *prior* to the principal move. Reassuringly, the estimates of β_1 are not distinguishable from zero, mitigating concerns that the changes in principal quality are picking up spurious changes in student performance or that students are sorting based on academic gains.

Finally, Appendix Figure A.4 presents event-study estimates using the analogue of equation 1.7 to directly test student sorting on observables. Consistent with the previous analysis, changes in principal quality are not associated with changes in school-level means of student characteristics. This result is perhaps unsurprising given the difficulties involved in moving schools or school districts, and that schools, within the same district, are similar in terms of student demographics.

5. Who are the Effective Principals

This section uses the PVA estimates to understand what effective principals look like, whether more effective principals are rewarded with higher salaries, and where effective principals tend to work.

5.1 Correlates of Principal Value-Added

I now examine whether observable characteristics can predict principal quality and quantify the amount of variation in PVA explained by these observables. Documenting a meaningful association between observed characteristics and principal quality could allow district superintendents to better identify effective candidates and to potentially create incentives to obtain certain credentials.

Table 1.5 describes the raw data for principals with above- and below- median value-added. These two groups differ in most observable characteristics: most noticeably, above-median-PVA principals have higher annual salaries, more experience, are more likely to be female, white, graduate from a school ranked in the *US News & World Rerport's* university ranking, and were more likely to have been effective teachers if they had previously taught in the North Carolina public school system. To formally test the extent to which whether observable characteristics explain principal effectiveness, I estimate the following equation, where the dependent variable is the raw value-added estimates obtained in equation 1.6:

$$\hat{\theta}_p = \beta_0 + \tau X_p + \hat{\theta}^{eb}_{p(j)} + \epsilon_p \tag{1.9}$$

where X_p is a vector of time-invariant principal characteristics such as gender, race, highest degree obtained, and indicators for whether the principal graduated from a school ranked among *US News & World Report*'s top universities and whether a principal previously held a teaching position in North Carolina. For principals previously observed teaching, I use empirical Bayes shrunken estimates of their teacher value-added (TVA) to directly account for teacher quality[18].

[18] TVA is calculated similarly to PVA but replaces the principal fixed effects with teacher fixed effects.

To account for time-varying characteristics, I use the mean or most common value for that principal. For example, I control for age by using the mean age observed and include an indicator for whether a principal was observed for at least four years of experience. In some specifications, district fixed effects are included to account for potential endogenous sorting or assignment of principals across districts.

Table 1.6 presents the results from this specification. Column 1 reports the regression results across all principals, and Column 2 reports the estimates for principals who previously taught in North Carolina. Columns 3 and 4 include district fixed effects, while Columns 5 and 6 include school fixed effects to account for potential sorting across schools. Across the estimation samples, certain patterns emerge. First principals who were effective at teaching are more likely to become stronger principals in the future. This result is in contrast to the idea of the "Peter Principle" (Peter et al., 1969), that suggests promotions based on current performance are misaligned, and mitigates concerns that promoting the strongest teachers dampens student learning since the pool of teachers becomes weaker[19]. Additionally, this result stands in contrast with existing literature assessing teacher quality. A well-documented phenomenon in the education literature is the difficulty of using observable characteristics to predict teacher quality (Rockoff, 2004; Rockoff et al., 2011; Rothstein, 2015). However, these estimates suggest that prior teaching quality could be an informative signal in the hiring and future success of principals.

Second, principals with at least four years of experience are more effective by 1.7% of a standard deviation than their less experienced counterparts. This role of experience is

The shrinkage parameter for the teacher fixed effects follows from Kane and Staiger (2008), Jackson (2018), and Bau and Das (2020).

[19]The "Peter principle" contends that workers are promoted up to the point where they are no longer competent suggesting that using current performance is a poor predictor of success at the later job.

consistent with findings from studies assessing the correlates of teacher and managerial effectiveness (Rockoff, 2004, Bau and Das, 2020, Fenizia, 2022). Additionally, graduating from an in-state institution is associated with higher future principal quality, which, similarly to the estimates on prior teaching, suggests that institutional features could play a role in principal effectiveness, as over 80% of North Carolina graduates stay in North Carolina[20].

Third, observable characteristics explain very little of the variation in principal effects. The within R^2 in specifications excluding district fixed effects never exceeds 7%, and a formal F-test cannot reject the null hypothesis that the covariates are jointly statistically different from zero[21], suggesting that observable characteristics play a small part in determining principal effectiveness. Nonetheless, the estimates from Table 1.6 shed light on what attributes school district officials could value, as prior teaching experience is strongly correlated with future principal quality.

5.2 Are Wages Commensurate with PVA

Estimates of PVA allow me to obtain an ordinal ranking of principals and to examine the relationship between value-added and principal compensation as well as where effective principals tend to work. While standard labor economics models indicate that wages are correlated with productivity, public sector salaries are experience based potentially limiting the correlation between wages and PVA.

.

Table 1.5 suggests that the labor market rewards more effective principals; however, differences in principal compensation might reflect differences in bonus salary compensation

[20]See data from https://tower.nc.gov/.

[21]A separate explanatory calculation examines the difference in adjusted R^2 of a regression of PVA on school fixed effects and a regression of PVA on covariates and the fixed effects (i.e., Column 1), this approach yields a quantitatively similar result.

across school districts, with more affluent districts typically offering larger bonuses to offset higher living costs[22]. To test whether more effective principals are rewarded with higher salaries, I regress log salaries on principal characteristics:

$$log(\overline{salary})_p = \beta_0 + \hat{\theta}_p^{eb} + \tau \mathbf{X}_p + \alpha_d + \epsilon_p \tag{1.10}$$

where $log(\overline{salary})_p$ is the log of the average salary of principal p and $\mathbf{X}_p$ is the set of observable characteristics in equation 1.9. α_d is a district fixed effect and, in some specifications, a school fixed effect. Table 1.7 reports the estimation results.

Several patterns emerge for the relationship between principal salary and observable characteristics. Across all specifications, there is an approximately 6% increase in salary for principals with at least four years of experience. Second, conditional on either district or school fixed effects, mean PVA explains very little in observed principal compensation as the point estimates are not distinguishable from zero and the difference in adjusted R^2 does not change with the inclusion or exclusion of PVA (Columns 2 and 3). Finally, in isolation, Column 1 suggests a public sector premium for higher PVA. However, this effect shrinks in size and is no longer statistically significant when including either district or school fixed-effects (Columns 3 and 4) suggesting that cross-district variation in salary bonuses partially explain the observed relationship between principal compensation and PVA.

Appendix Figure A.1 and Appendix Table A.4 explore the spatial allocation of principals in greater detail. Figure A.1 plots the distribution of PVA, separately, for schools in various quartiles based on the share of students eligible for FRPL, while Table A.4 provides principal- and school-level summary statistics based on the same quartiles. Several consistent

[22]See the North Carolina Department of Public Instruction for specific information on the local salary supplement provided by each school district

patterns emerge. First, schools with less disadvantaged students have stronger principals on average. Schools in the first quartile (least disadvantage) have an average PVA of 0.023 standard-deviations above the mean whereas those in the most disadvantage (fourth quartile) have principals 0.008 standard-deviations below the mean. Second, principals working in less disadvantaged schools have larger salaries which reflects differences in local salary supplements and further supports the finding that stronger principals tend to sort into more affluent schools. These descriptive statistics are consistent with Branch et al. (2012) who find that the strongest principals reside in the least disadvantaged schools in Texas.

This result has important equity implications regarding the principal labor market. It suggests that more effective principals sort into, or are potentially allocated to, less distressed areas since more affluent districts offer larger salary bonuses. My results are consistent with Ba et al. (2021), who find that more effective police officers tend to work in low-crime areas, and suggest potential equity gains from incentivizing stronger principals' working in more disadvantaged areas. In contrast to Das et al. (2016) who examine the public sector wage premium for doctors in India, this exercise on PVA and compensation does not necessarily suggest an absence of a PVA premium (across districts), instead it reveals effective principals sort into less disadvantaged areas.

6. What Makes for an Effective Principal

In this section, I explore what practices are espoused by effective principals and how they shape student outcomes. I examine how within-school changes in principal quality affect the teacher composition and whether effective principals possess certain skills that differentiate them from their peers.

Panels B and C of Table 1.5 contrast the school and teacher characteristics of principals with above- and below-median value-added. While teacher experience is comparable, there are large differences in overall teacher quality and in the quality of arriving and exiting teachers. While the scope for principals to directly hire teachers varies across states and school districts, principals in North Carolina typically have the authority to signal which teachers they are interested in. Using teacher application data from one large school district in North Carolina, Bates et al. (2022) note that principals directly assess teacher candidates, have records on which candidates they interview, and also have records on which candidates are hired, suggesting that principals are directly involved in the teacher labor market and are aware of the job status of prospective candidates.

Given principals' scope to influence teacher hiring decisions, the next subsections examines how principals affect teacher composition and whether principals who are effective at boosting test scores also have stronger leadership and soft skills.

6.1 Impacts on Teacher Composition

I begin by asking whether effective principals affect student test scores by shaping teacher composition. As a first pass, I examine how changes in principal quality affect the average teacher value-added for a given school. I estimate the following regression at the school–year level:

$$\overline{TVA}_{s,t} = \beta_0 + \sum_{\tau \neq -1} \beta_\tau \Delta P + X'_{s,t}\lambda + \alpha_s + \alpha_t + \epsilon_{s,t} \tag{1.11}$$

where a school experiences a principal transition at time $\tau = 0$. The main parameters of interest are β_τ, which capture the effect of changes in principal quality (ΔP) on outcomes.

While the previous exercise estimates individual principal effectiveness and allows us to compute changes in principal quality, teacher and principal quality are mechanically correlated since they are estimated from the same underlying data. In other words, an effective principal at school s will seemingly have a positive β_τ estimate for her teachers since teacher and principal effects are estimated from same set of strong test score gains.

We address this issue by estimating principal quality and teacher quality "out-of-sample". For a given principal p, I estimate her value-added using all test score data prior to her arriving at school s at time $\tau = 0$. This restriction means that all of the variation in her value-added comes from the test score data from some school $s' \neq s$. I impose a similar restriction for teachers. I estimate a given teacher's value-added using the test score data from prior to principal p's arrival, modifying equation 1.6 by replacing the principal fixed effects with a teacher fixed effects[23].

These restrictions, while conservative, address the following issues. First, deriving estimates from data on principals at a different set of schools from teachers ensures that principal and teacher value-added are not mechanically correlated. Second, it addresses the potential for changes in test scores to be driven primarily by teacher as opposed to principal-teacher match quality. Together, these restrictions imply that PVA captures whether a principal who is effective at school s' is as effective at school s. Teachers' value-added can be interpreted as teacher effectiveness without the incorporation of match effects or changes in their development from working with a given principal.

In addition to the estimation restrictions, I require exiting and arriving principals to be observed for at least four consecutive years. This restriction ensures that principal quality

[23]Less than 4% of all teachers work for the same principal multiple times, limiting the concern that teacher value-added is estimated from the same data as principal effectiveness.

is estimated with sufficient data and thus is more precisely estimated and that principals are not simply temporary hires, and it allows sufficient time for principals to exert their influence. However, this imposition limits the sample to 498 principals. While these restrictions might be too demanding, Appendix Figure A.5 provides the estimation results for the sample without this restriction. The estimates, while smaller, remain qualitatively similar.

Given these restrictions, I estimate a modified version of 1.11 focusing on how changes in principal quality affect the average TVA in mathematics[24]:

$$\begin{gathered}\overline{TVA}_{s,t} = \beta_0 + \sum_{\tau \neq -1} \beta_\tau \widehat{\Delta P} + X'_{s,t}\lambda + \alpha_s + \alpha_t + \epsilon_{s,t} \\ \text{where } \widehat{\Delta P} = \hat{\theta}^{eb,-s}_{p,\text{incoming}} - \hat{\theta}^{eb}_{p,\text{outgoing}}\end{gathered} \tag{1.12}$$

where $\hat{\theta}^{eb,-s}_{p,\text{incoming}}$ is the value-added of the incoming principal at time $\tau = 0$ estimated in data from schools other than s. $\hat{\theta}^{eb}_{p,\text{outgoing}}$ is the value-added of the outgoing principal, and both parameters are shrunken by means of the previously described procedure. $\overline{TVA}_{s,t}$ is the average TVA of school s at time t and is estimated from the sample subject to the aforementioned restrictions. β_τ captures changes in average teacher quality attributable to entry and exit of existing teachers but not to differences in teacher development. $X'_{s,t}$ is a vector of time-varying school means of student demographics. α_s is a school fixed effect that accounts for time-invariant features of the school that may influence school outcomes, and α_t is a year fixed effect that addresses annual-level shocks.

Figure 1.4a plots the event-study coefficients. Schools gaining more effective principals see improvements in average teacher quality. Differences-in-differences estimates of 1.12 indicate that a one-SD increase in principal quality raises average teacher quality by 0.17 SD.

[24] As discussed in Section 4.1, I use TVA in math since, in middle school, students have multiple teachers, which makes it difficult to construct the mean value-added for a given teacher.

The effects on teacher composition take time to materialize, reflecting the general difficulty of dismissing public officials and the fact that principals may require time to exert their influence on staff.

I next examine whether these compositional effects are driven by differences in hiring or dismissals. I re-estimate equation 1.12 at the teacher–year level with ex ante value-added as the outcome variable. Panels 1.4b and 1.4c examine these changes. Changes in teacher effectiveness primarily come from teacher recruitment, as a one-SD increase in principal effectiveness increases the ex ante value-added of entering teachers by 0.357 SD. On the other hand, composition effects attributable to changes in dismissal appear less likely, as exiting teacher quality largely mirrors dismissal quality prior to the arrival of an effective principal.

To address whether changes in teacher quality reflect a change in the volume of entries and exits, Appendix Figure A.3 plots the corresponding event-study estimates under the same sample restrictions but with the dependent variable replaced in 1.12 with the number of entering and exiting teachers. Figures A.2a and A.2b provide evidence that entering and exiting teacher counts remain constant following principal transitions. Figure A.3 provides some evidence that effective principals reduce the number of exiting teachers, as the differences-in-differences estimate, though noisy, shows a modest effect on the number of exiting teachers, and I provide additional evidence that effective principals reduce turnover in Section 6.2. Finally, Figure A.3b plots binned scatter plots of the relation between the number of teacher exits and principal quality across the full sample of principals. Consistent with those from the restricted sample, these results indicate that, if anything, transitioning to a more effective principal is associated with a decrease in the number of teachers leaving, which suggests that the positive estimates in Figure 1.4 reflect the recruitment of stronger teachers—and not mere dismissal of more teachers by principals—boosting average teacher quality.

The characteristics of entering teachers differ across principal quality, as well. Figure 1.5a presents suggestive evidence that more effective principals can more easily fill a job vacancy with experienced teachers. Among existing teachers who transition to a new school, a one-SD increase in principal quality leads to a nearly 3-year increase in experience (though not statistically significant)[25], but in general teachers who move tend to be relatively experienced (with 10.6 years of experience, on average). Prior work has documented that newly arrived teachers or those with just one year of experience tend to be the least effective. Figure 1.5b indicates that test score–boosting principal may increase average teacher quality by relying on fewer inexperienced teachers to fill a vacancy, as a one-SD increase in value-added decreases the share of new teachers and those with one year of experience by nearly 20%. To show robustness to alternative definitions of inexperienced teachers, Appendix Figure A.6b shows that this relationship still holds when extending the analysis to include teachers with less than three years of experience or teachers who did not receive tenure at their prior school[26].

This subsection provides evidence that higher-PVA principals raise aggregate teacher quality. While compositional changes largely reflect stronger teacher recruitment, as shown in Figure 1.4c, the analysis does not address whether higher-PVA principals can identify stronger teachers or can better retain the best teachers or how they might maximize the output of teaching staff.

[25]To be consistent with the specification in equation 1.12, Appendix Figure A.6a re-examines this relationship using PVA in math as the independent variable. Results are similar in magnitude but statistically significant at the $\alpha = 10\%$ level.

[26]Prior to 2013, teachers in North Carolina were eligible for career status or "tenure" after four years of service.

6.2 *Impacts on Existing Teachers*

An important question is whether effective principals reduce turnover and, in particular, whether they are more likely to keep the best teachers. Reducing general turnover is relevant, as high turnover rates could be disruptive for students and increase the workload of existing faculty members. Similarly, retaining effective teachers fosters greater student learning. Beyond influencing turnover, effective principals may influence classroom assignments. For instance, Bates et al. (2022) documents large academic gains from assigning the most effective teachers to the largest classrooms.

I begin by examining how principals affect teacher turnover. I estimate

$$y_{s,t} = \beta_0 + \beta_1 \hat{\theta}_p^{eb} + X'_{s,t}\phi_2 + \alpha_t + \alpha_s + \epsilon_{s,t} \tag{1.13}$$

where $y_{s,t}$ is either the annual teacher attrition rate or classroom size, $\hat{\theta}_p^{eb}$ is mean PVA, and $X'_{s,t}$ is a set of school-level means defined in equation 1.12. For the teacher attrition rate, I use two measures. The first is the overall annual turnover rate at the school level. The second is the difference in attrition rates for teachers above and below the median *within* their school. To construct this measure, I calculate, for each year, the median TVA in mathematics at the school level. The outcome variable is then the difference in attrition rates between teachers with above- and below-median value-added[27]. As the second measure already represents a within-school ranking of teachers, the school fixed effects are omitted when I estimate β_1.

Figure 1.6 plots the binscatter analogue of equation 1.13 examining the relationship be-

[27] The corresponding exit rate is given by $\frac{\sum \mathbf{I}\{exit\}_{j,t} \times \mathbf{I}\{AboveMed\}_{j,t}}{N_{j,s,t}} - \frac{\sum \mathbf{I}\{exit\}_{j,t} \times \mathbf{I}\{1-AboveMed\}_{j,t}}{N_{j,s,t}}$, where *exit* is an indicator for whether teacher j exits at the end of year t. $AboveMed$ is an indicator for whether teacher j's value-added is greater than the school median in year t. $N_{j,s,t}$ is the number of teachers in school s in year t.

tween principal quality and attrition rates, where the point estimate and standard error correspond to an ordinary least squares (OLS) regression on the entire micro-data (e.g., at principal–year level) and not the binned averages. Principals who boost test scores are also stronger at reducing overall turnover and keeping their best teachers. A one-SD increase in principal quality reduces overall turnover by 27% and reduces the relative turnover rate for effective teachers by 6.4%[28]. This ability to not only retain the best faculty but also attract stronger staff members is consistent with the finding of Jacob and Lefgren (2005) that principals can identify teachers at the top of the distribution.

Figure 1.7 examines the teacher "poaching" rate across principal quality[29]. Consistent with effective principals reducing overall turnover, Figure 1.7 reveals that stronger principals are also less likely to lose their teachers to another principal. A one-SD increase in PVA reduces the likelihood a teacher is "poached" away by 5.7% suggesting that principals with high PVA might possess certain characteristics that make them desirable to work with.

Turning to classroom assignments, Figure 1.8a begins by examining the raw correlations between TVA and class size separately for above- and below-median value-added principals, as more effective principals may sort into schools with smaller classrooms, allowing greater scope to adjust teacher assignments. The raw data suggest that more effective principals are more likely to adjust class sizes. However, as Table 1.5 indicates, more effective principals tend to work in schools with smaller classrooms, which might give them greater scope to adjust classroom allocations. To address this concern, 1.8b plots the same relationship after I residualize class size and TVA on school–year shocks to account for changes in cohort sizes

[28]Note that since the outcome variable relates to teacher quality, value-added in math is used for both teachers and principals. Appendix Figure A.7 provides robustness using mean PVA.

[29]Poaching is defined as whether a teacher separating from a school in year t then joins a different school in year $t + 1$

over time or policies influencing the allocation of students[30].

Figure 1.8b reveals that, even after I account for selection into certain schools, higher-PVA principals are more likely to assign stronger teachers to larger classrooms. For above-median value-added principals, a one-SD increase in TVA increases class size by approximately two students. Below-median principals still manipulate classroom assignments, albeit to a lesser degree, as a one-SD increase in TVA increases class size by less than one student. This finding suggests that one key mechanism is strategic task allocation or simply greater personnel awareness, as, even after I account for school–year-level shocks, more effective principals are more likely to assign their strongest staff to the largest classrooms to generate the largest learning gains.

6.3 Principal Leadership and School Climate

Having shown that test score–raising principals are more adept at attracting and retaining effective faculty, I now investigate whether certain principals possess soft skills that make them more desirable to work with. Literature documents that working conditions are important in determining where employees choose to work (Maestas et al., 2023). For instance, Bates et al. (2022) note that teachers prefer working with principals with stronger leadership skills. If principals who are effective at increasing test scores also exhibit stronger leadership skills or can positively shape school climate, then the observed differences in recruiting and retention may reflect variation in underlying soft skills.

The NCERDC's Working Conditions Survey allows me to directly test this hypothesis. Beginning in 2002 and administered biannually, the survey asks staff members to express on

[30] Specifically, I first estimate $classsize_{j,t} = \beta_0 + \alpha_{s,t} + \alpha_g + \epsilon_{j,t}$ and obtain the residuals $\hat{\epsilon_{j,t}}$. I then estimate $\hat{\theta}_j^{EB} = \beta_0 + \alpha_{s,t} + \alpha_g + \nu_j$ and obtain the residuals $\hat{\nu}_j$. I then plot β_1 corresponding to $\hat{\epsilon_{j,t}} = \beta_0 + \beta_1 \hat{\nu}_j + \gamma_{j,t}$.

a scale of one (“strongly disagree") through five (“strongly agree") their sentiment on various questions related to overall school climate, teacher empowerment, and principal leadership[31]. While these data cannot be linked to individual teachers to gauge the overall response rate, based on documentation provided by the NCERDC, over 85% of schools participate and respond to survey questions, providing relatively comprehensive coverage of North Carolina schools and principals[32]. Table 1.8 details specific items in each of the three main categories and provides the share of respondents who agree with each statement. To unpack the relationship between principals' effectiveness in raising test scores and their soft skills, I estimate the analogue of equation 1.13.

Figure 1.9 details the association between PVA and overall school climate, principal leadership, and teacher empowerment. Across all measures, there is a robust relationship between test-score effectiveness and principal soft skills. A one-SD increase in principal effectiveness is typically associated with a one-point increase in faculty satisfaction—equivalent to the average teacher going from “unsure" to “agreeing" that principal leadership is effective. Further, a one-SD rise in test-score effects implies that nearly all teachers “strongly agree" that the school climate is effective (up from a mean of “agree').

While stronger soft skills may explain the differences in teacher retention and recruitment, a natural question is whether these principal traits can influence teacher effectiveness. I explore this in more detail in Appendix Section 1.4. Appendix Table A.2 reports regression results of TVA on observable teacher characteristics and the various aspects of soft skills and school leadership. The estimates suggest large implications regarding the role of school

[31]Survey responses correspond to the following: 1=“strongly disagree", 2=“disagree", 3=“don't know/unsure", 4=“agree", and 5=“strongly agree".

[32]In 2002, the first survey year, approximately two-thirds of all schools participated. However, by 2008, approximately 87% of all *teachers* had completed the survey (NCERDC, 2008 WCS Codebook). For the most recent data used, nearly 91% of all teachers completed the survey (NCERDC, 2018 WCS Codebook). See Appendix Figure A.8 for details on teacher participation rate and the number of schools participating.

management and leadership even when I include school fixed effects and utilize within-school variation. Moving from “unsure" to “agreeing" that school leadership is effective is associated with at least a 0.043-SD increase in teacher effectiveness, while the impacts for school climate are smaller at 0.0078 SD. These large returns to principal leadership may be driven by effective principals instilling greater human capital in their teachers or by the role of teacher–school match quality (Jackson, 2013).

Whereas prior studies have documented the importance of effective managerial practices in schools (Bloom et al. (2015), Lemos et al. (2021)) for student outcomes, Figure 1.9 indicates that management quality is not an institutional feature of a given school but rather a product of principal effectiveness. Even when I compare within the same school, differences in principal quality are heavily correlated with overall school climate, teacher empowerment, and effective leadership. This result suggests that certain aspects of school quality are attributable to differences in principal effectiveness. Appendix Figure A.9 examines the relationship between principal quality and the individual questions used to construct the outcomes in Figure 1.9.

7. Conclusion

This paper provides new estimates on the impact of principal quality on student achievement, is among the first to show that value-added models can be extended to quantify principal effects on test scores, and examines how principals influence student outcomes. Extending the variance decomposition from Araujo et al. (2016) and Bau and Das (2020), I show that a one-SD increase in principal quality leads to a 0.047-SD increase in average student performance. These estimates are nearly as large as prior work examining school quality, suggesting that estimates of school effects may largely reflect underlying differences in principal effectiveness. Principal effects are particularly relevant since principals are responsible for all students and not just those in a specific classsroom.

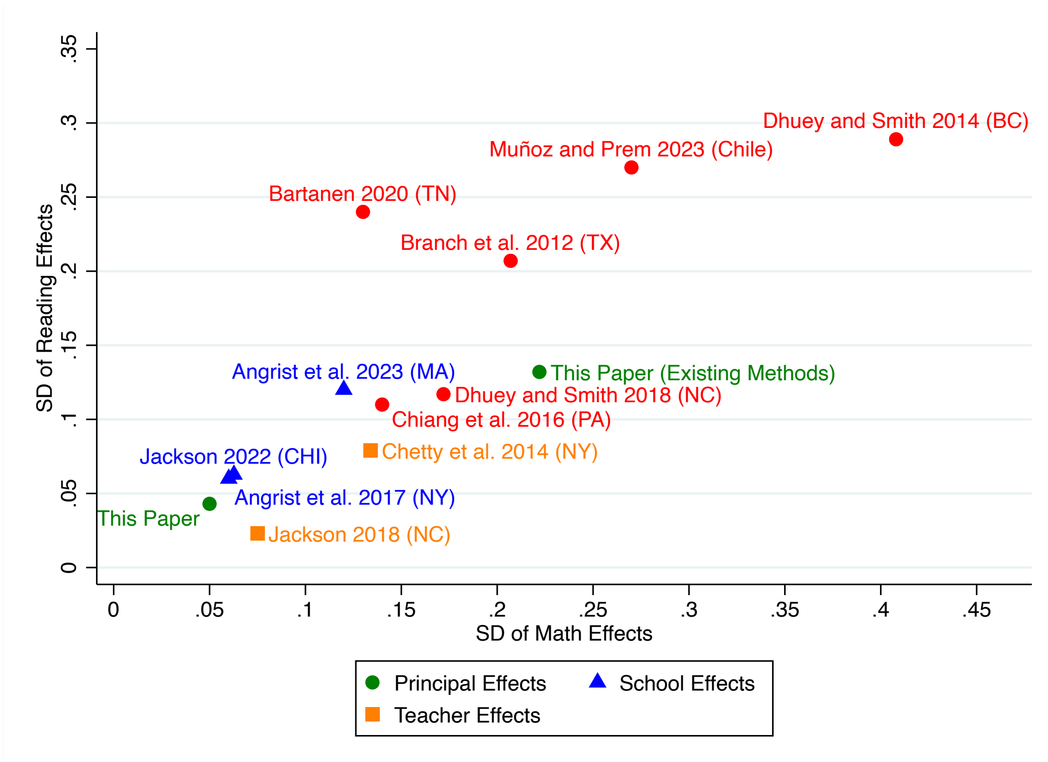
Central to principal effectiveness is the role of personnel management. I show that test score–boosting principals are better at attracting high-quality instruction and retaining their best teachers and are more likely to allocate their best teachers to the largest classrooms, which benefits more students. Differences in soft skills may explain why effective principals can positively shape their teaching staff. Higher-value-added principals are more likely to score higher on various measures of leadership, teacher empowerment, and stronger school climate—results that echo findings from a recent literature emphasizing the role of working conditions and leadership in job choice (Mas and Pallais, 2017; Bates et al., 2022; Maestas et al., 2023).

The analysis from this paper provides important directions for future research. A key finding is that effective principals are better at attracting and retaining the best teachers, but this may come at the expense of other schools. Furthermore, principals with higher value-added are more likely to work in more advantaged districts, potentially exacerbating inequities in access to effective teachers and schools and widening the gap in academic achievement across socioeconomic status. Changing the allocation of principals to schools may address this issue. Future work examining principal–school match quality may offer insights on where principals should be allocated. For instance, there could be principal match effects based on student demographics, similar to the match effects highlighted in the teacher literature (Dee, 2005; Porter and Serra, 2020; Gershenson et al., 2022), while policies such as increasing compensation may nudge principals to work in such schools (Bobba et al., 2021).

Finally, value-added captures principal effects only on test scores, ignoring their potential impacts on nonacademic outcomes. Principals are responsible for administering student discipline, and their large influence on school climate suggests that their impacts extend beyond academics. Understanding whether test score–boosting principals also improve student behavioral outcomes would shed further light on how principals affect student outcomes and the multidimensionality of principal quality.

8. Figures

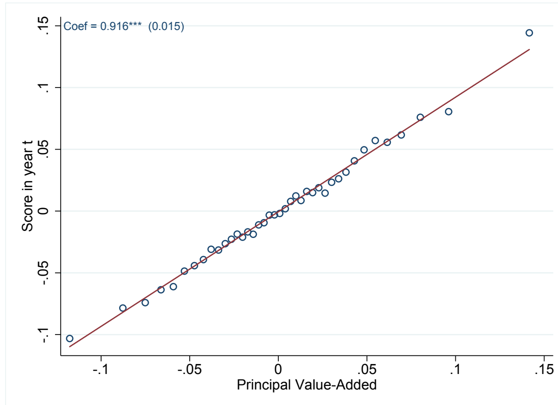
Figure 1.1: Comparison of Existing Estimates



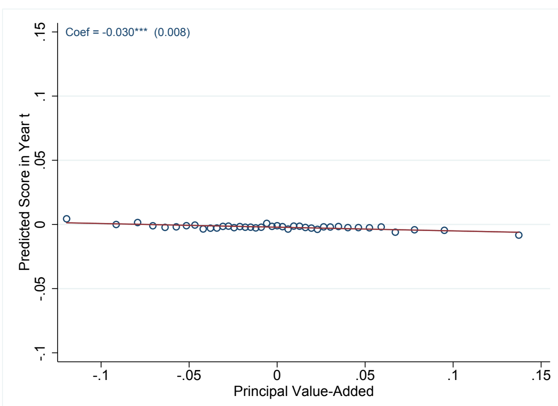
Notes: This figure plots existing estimates of school, principal, and teacher quality. School effects are denoted with blue triangles. With the exception of the red circles, which reflect my estimates, principal estimates are denoted with green circles. Teacher effects are denoted with orange squares. To obtain the red circle labeled "This Paper (Existing Estimates)", I follow the procedure discussed in Footnote 12.

Figure 1.2: PVA Predictions and Bias from Omitted Twice-Lagged Scores

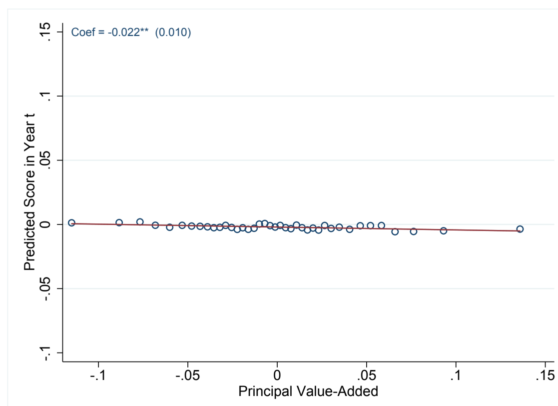
(a) Actual Score



(b) Predicted Score Using Year $t-2$ Score (All Principals)

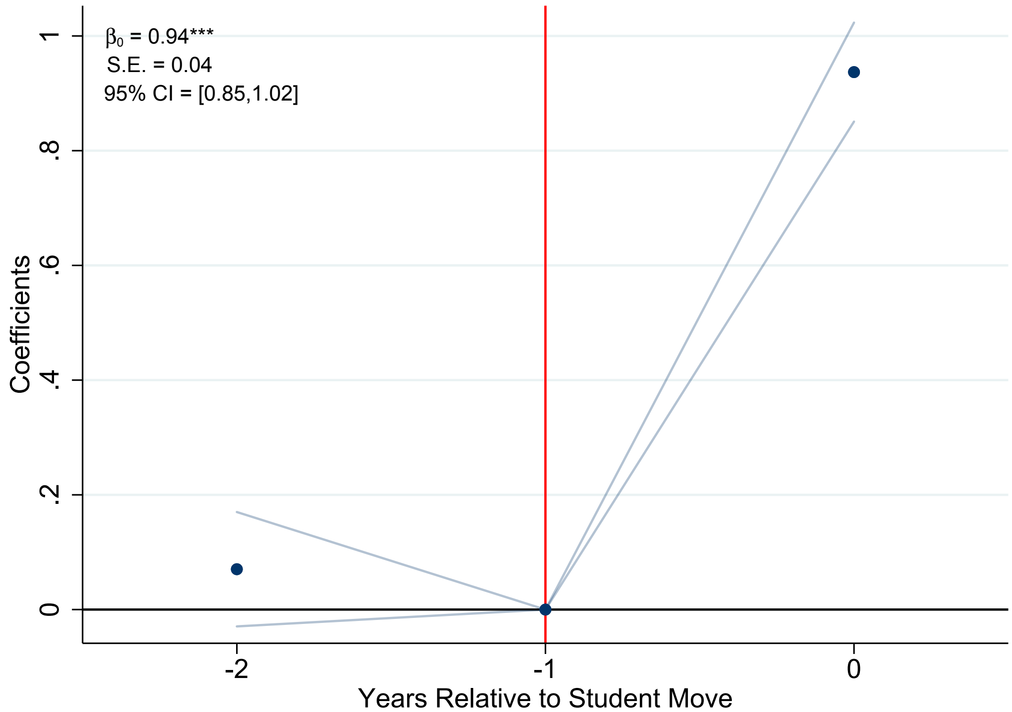


(c) Predicted Score Using Year $t-2$ Score (Principal Movers)



Notes: This figure pools all grades and subjects for the sample used to estimate PVA. Observations are at the student–subject–school–year level. Panel A plots the binscatter of actual student test scores in year t against PVA after I residualize both variables at the subject (math versus English) by school (middle versus elementary school) level. Panel B plots the relationship between predicted test scores using $t-2$ data on the full sample. Panel C plots the same analysis but restricts the sample to principals who worked in more than one school. To construct this variable, I residualize the test score outcomes in year t and $t-2$ using the set of covariates described in equation 1.6. I regress the residuals of year t against the residuals in year $t-2$ to obtain the the predicted outcome of interest. The binscatter plots the relationship between this predicted score and value-added with the coefficient corresponding to the student micro-level and with standard errors clustered at the school level. Numbers of observations are 11,417,400, 4,430,890, and 2,238,913 for panels A, B, and C, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

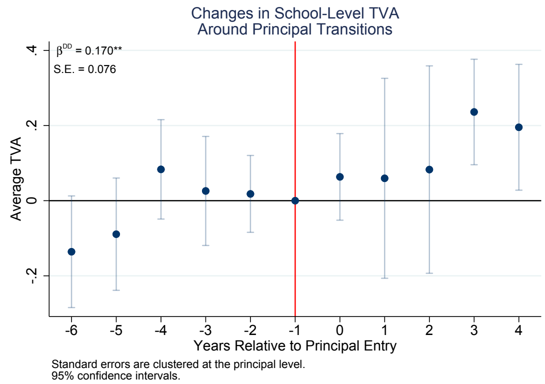
Figure 1.3: Effect of Mean TVA at $\tau = 0$ on Switchers' Test Scores



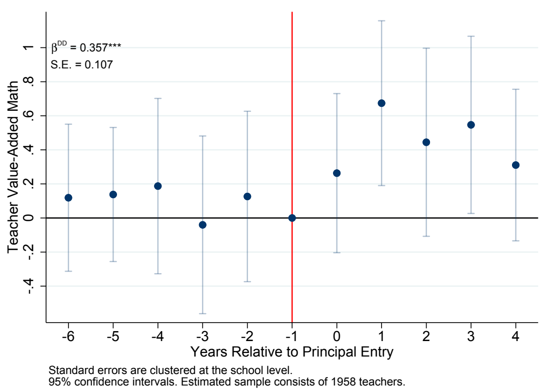
Notes: This figure plots the β coefficients from equation 1.7, which estimates the effect of a current principal's value-added on student test scores before and immediately after the switch occurs. The sample consists of students who switch schools and principals at $t = 0$. Estimates control for lagged student achievement, observable student characteristic and school means, grade fixed effects, and school-by-year fixed effects. Regression is at the student–year level. 1,242,217 student observations with 6,123 principals. Standard errors are clustered at the school level, and 95% confidence intervals are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1.4: Changes in Principal Quality on Teacher Composition

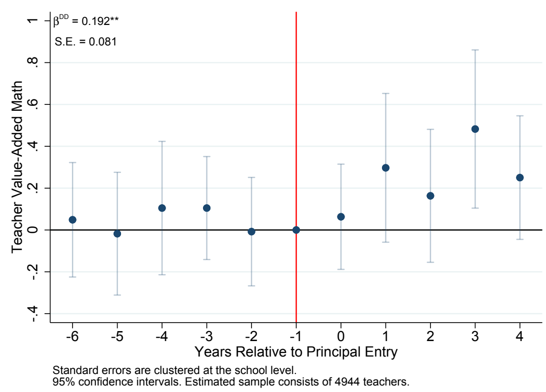
(a) Average Teacher Effectiveness



(b) Entering Teachers



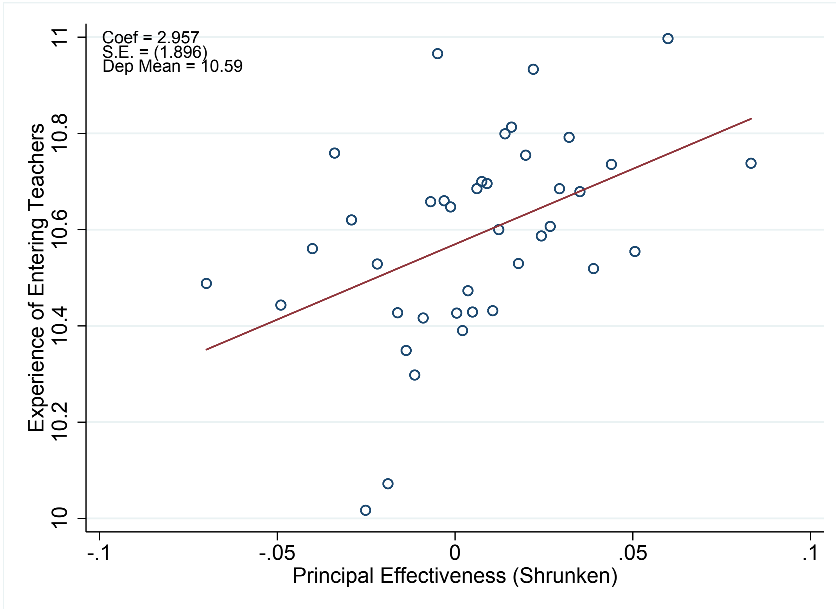
(c) Exiting Teachers



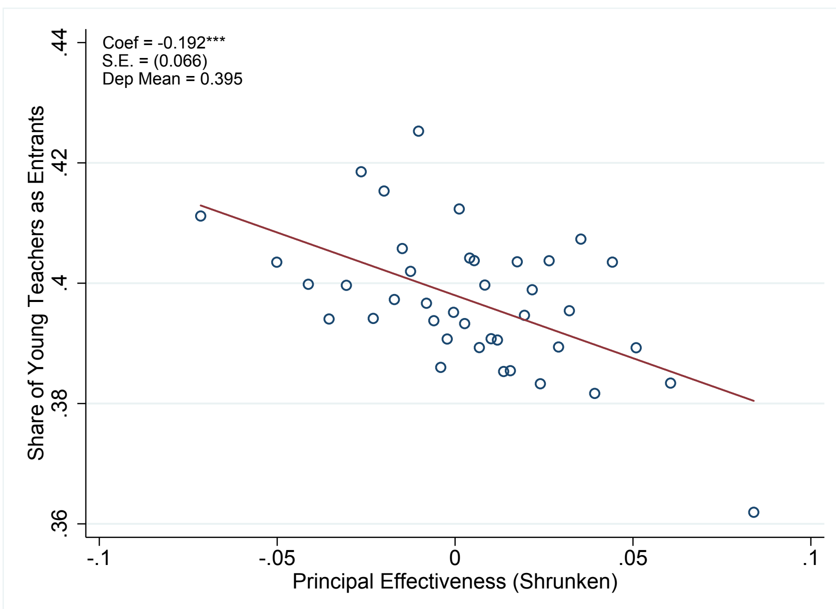
Notes: This figure plots the β coefficients from equation 1.12, which examines the effect of within-school changes in principal quality on average (school-level) and entering and exiting teacher value-added. The analysis focuses on "events", where the exiting and arriving principal are observed for four years; in total, there are 245 such events. Reported coefficients correspond to the difference-in-differences analogue of equation 1.12. All standard errors are clustered at the school level, and 95% confidence intervals are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1.5: Characteristics of Entering Teachers

(a) Experience of Entering Teachers



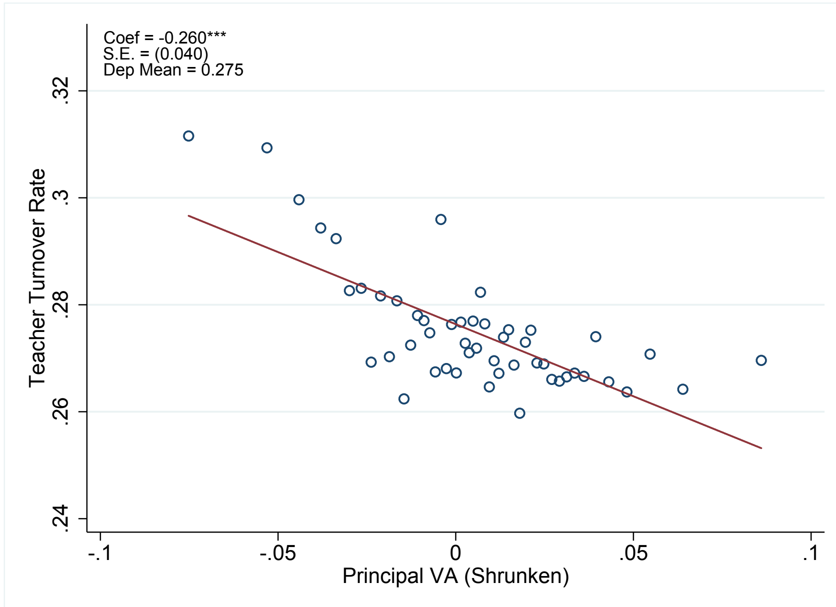
(b) Teachers with Less than 2 Years of Experience as Share of New Hires



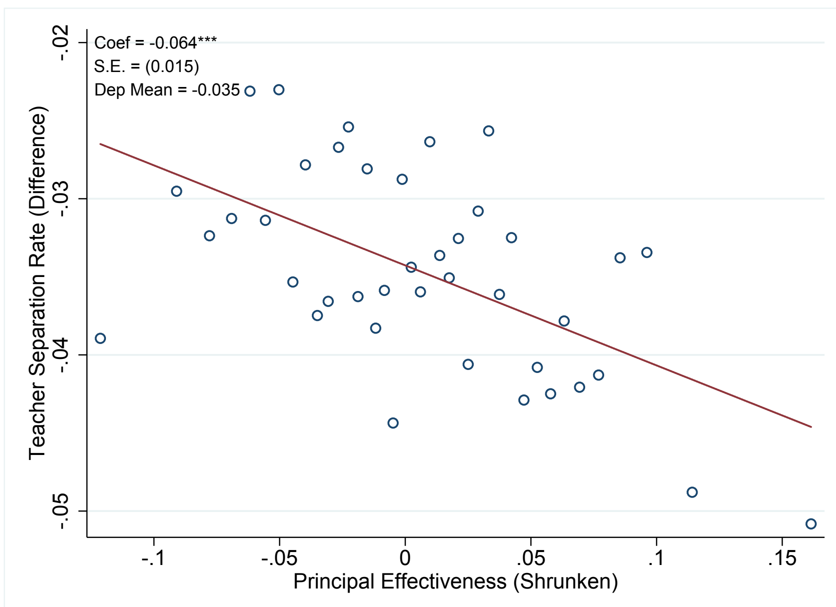
Notes: This figure shows the association between principal quality and characteristics of entering teachers. Panel A reports the results for the average experience level of entering teachers. Panel B reports the share of new entrants who have less than two years of experience. The binned scatter plots are divided into 40 equally sized bins that correspond to the conditional expectation of the outcome variable given a particular value of shrunken PVA. Both the dependent and independent variables are first residualzed against time-varying school means of student characteristics (described in equation 1.12), school fixed effects and year fixed effects. The coefficient and standard error correspond to the identical regression at the principal–year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1.6: Principal Quality and Teacher Attrition

(a) Turnover: All Teachers

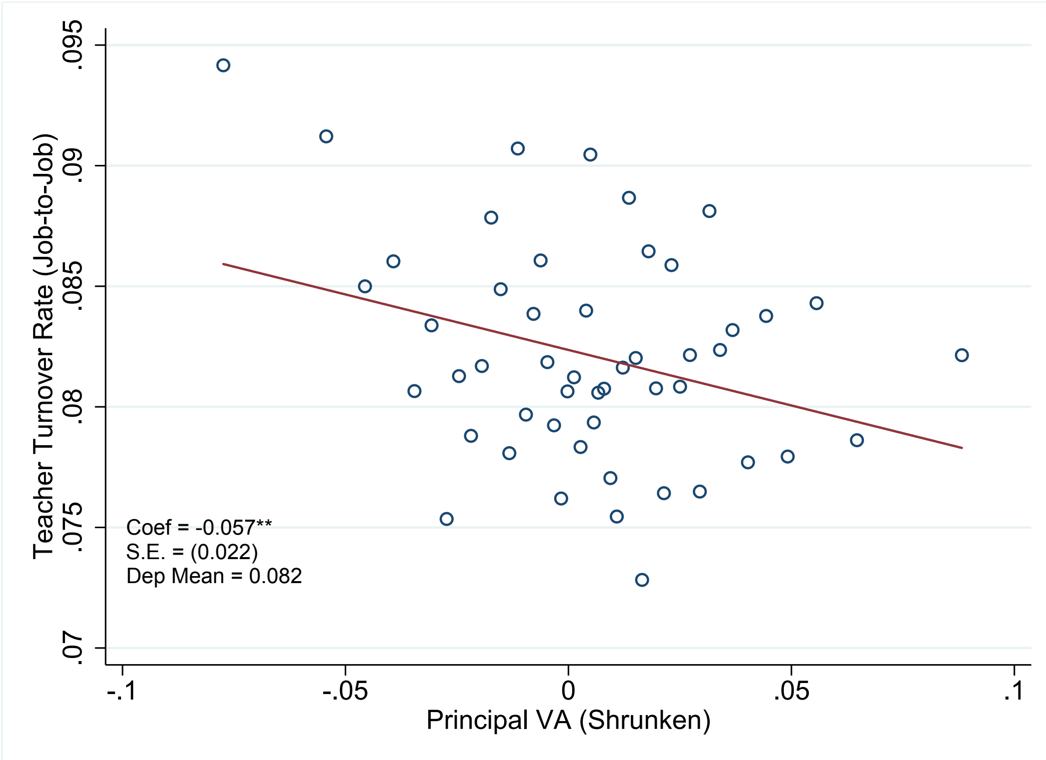


(b) Differential Turnover



Notes: This figure plots the binned scatter plot of equation 1.13. Panel A plots the turnover rate for all teachers. Panel B plots the difference in annual turnover rates for teachers with above- and below-median TVA in mathematics for a given school. The binned scatter plots are divided into 40 equally sized bins that correspond to the conditional expectation of the outcome variable given a particular value of shrunken PVA. Both the dependent and independent variables are first residualzed against time-varying school means of student characteristics (described in equation 1.12), school fixed effects and year fixed effects. Mean PVA and math PVA are used in Panels A and B, respectively. The coefficient and standard error correspond to the identical regression at the principal-year level. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

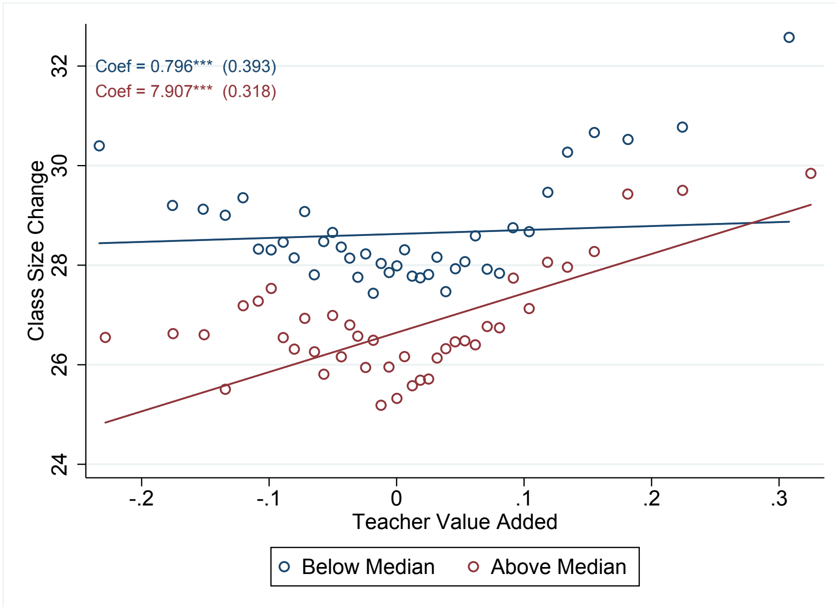
Figure 1.7: Principal Quality and Teacher Poaching



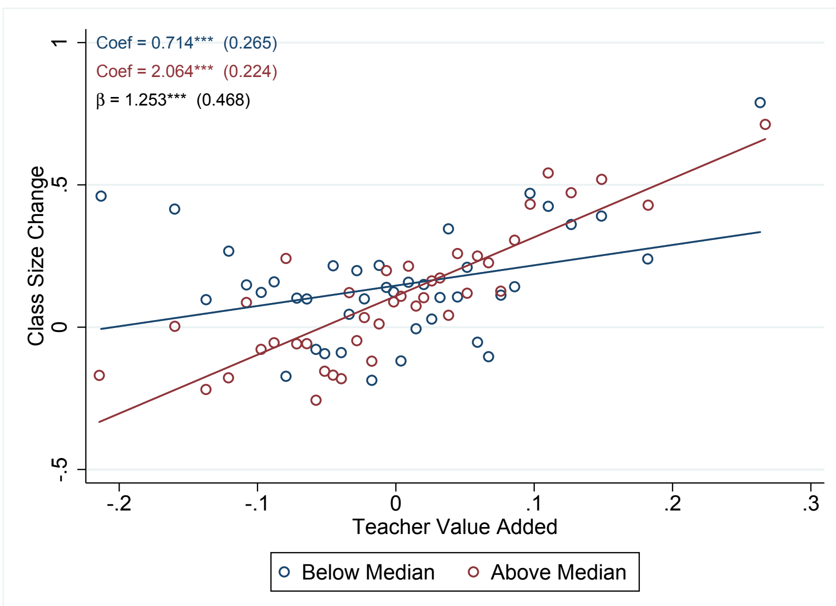
Notes: This figure plots the binned scatter plot of equation 1.13 where the outcome variable is whether a teacher exits from a school in year t and joins a different school in year $t+1$ and the dependent variable is mean PVA. Both the dependent and independent variables are first residualzed against time-varying school means of student characteristics (described in equation 1.12), school fixed effects and year fixed effects. The coefficient and standard error correspond to the identical regression at the principal-year level. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1.8: Principal Quality and Classroom Allocation

(a) Raw Correlation



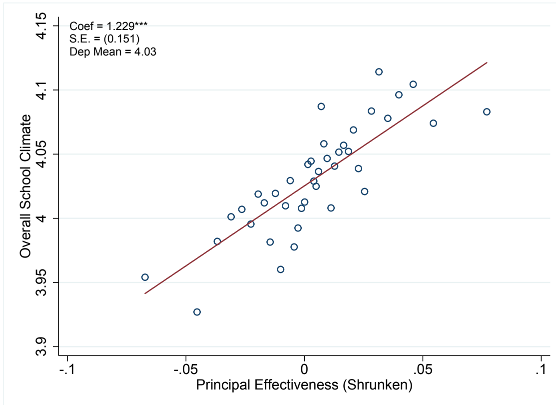
(b) Covariate Adjusted



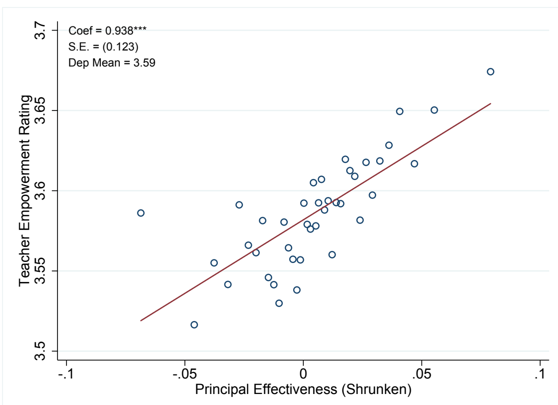
Notes: Panel A plots the raw correlation between class size and the shrunken TVA estimates separately for principals with above- and below-median value-added. Panel B plots the same association after I residualize both class size and the shrunken value-added estimates against school–year and grade fixed effects. The binned scatter plot figures are divided into 40 equally sized bins that correspond to the conditional expectation of the outcome variable given a particular value of the shrunken PVA. The β estimate in panel B corresponds to the estimate from equation $\text{ClassSize}_{j,t} = \alpha_0 + \alpha_1 \widehat{TVA}_j^{EB} + \beta \widehat{TVA}_j^{EB} \times D_p + \alpha_{s,t} + \alpha_g + \epsilon_{j,t}$, where D_p is an indicator for whether a principal has above-median value-added. Number of observations is 314,124 for both panels. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1.9: Principal Quality and School Climate

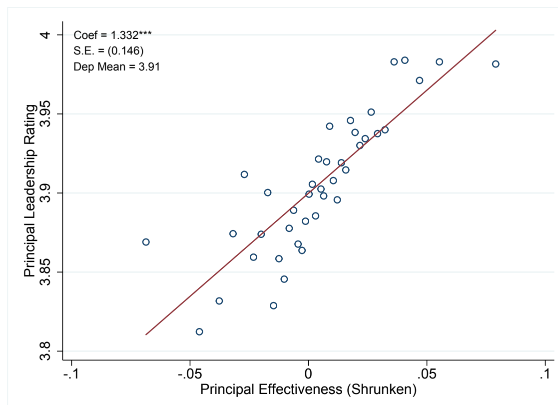
(a) Average School Climate



(b) Teacher Empowerment



(c) Effective Leadership



Notes: Each panel reports the binscatter associated with equation 1.13, with the results at the principal level. For panel A, respondents are asked to express on a scale of 1 through 5 whether they agree that their school is a good place to work, with 5="strongly agree". This regression contains 12,619 observations and 4,690 principals. For panels B and C, respondents are asked whether they agree that teachers are empowered or whether they believe that leadership is effective and strong. These regressions contain 13,645 observations and 4,954 principals. Surveys are biannual beginning in 2002 for teacher empowerment and school leadership. Survey questions asking about overall climate begin in 2004. In 2002, survey responses are expressed on a 6-point scale, with 6="strongly agree". Standard errors are clustered at the school level.

9. Tables

Table 1.1: Summary Statistics: AY 1996–2019

	All Principals		Principals Appearing in Multiple Schools	
	Mean (1)	SD (2)	Mean (3)	SD (4)
Principal Characteristics				
Female	0.59	0.49	0.60	0.49
Experience	4.18	3.82	5.20	4.15
Age	48.55	7.74	48.04	7.52
White	0.74	0.44	0.74	0.44
Annual Compensation (000s)	67.37	19.43	66.72	19.96
Top 10 or Ivy	0.00	0.06	0.00	0.05
School Ranked in *USN&WR*	0.15	0.35	0.15	0.36
Graduated from NC Institution	0.81	0.39	0.82	0.38
BA and Above	1.00	0.00	1.00	0.00
MA and Above	1.00	0.06	1.00	0.04
Doctorate	0.16	0.37	0.18	0.39
N Principals	6,549	0.00	2,184	0.00
School Characteristics				
Enrollment	261.03	186.23	266.13	189.83
Male	0.51	0.05	0.51	0.05
Eligible for FRP Lunch	0.52	0.22	0.52	0.23
Limited English	0.05	0.07	0.05	0.07
Black or Hispanic	0.38	0.27	0.39	0.27
Standardized Math Scores	-0.04	0.37	-0.03	0.38
Standardized Reading Scores	-0.04	0.35	-0.04	0.35
Average Class Size	31.33	17.70	31.32	17.79
N Schools	2,098	0.00	1,918	0.00
N student–subject–years	11,624,281			
N students	2,547,802			
N teachers	70,173			

Notes: This table presents principal- and school-level summary statistics for 1995–2019. Principal characteristics are separated into two groups. Columns 1 and 2 present statistics for all principals. Columns 3 and 4 focus on principals who worked in at least two schools. These statistics are collapsed from the principal–year (school–year) to the principal (school) level.

Table 1.2: Effects of 1-SD Improvement in Classroom, School, Principal, and Teacher Effects

	Math	Reading	Average
Classroom ($\sigma^2_{\theta_{j,s,p,t}}$)	0.175	0.150	0.162
School ($\sigma^2_{\theta_s}$)	0.057	0.037	0.047
Principal ($\sigma^2_{\theta_p}$)	0.050	0.043	0.047
Teacher ($\sigma^2_{\theta_j}$)	0.135	0.072	0.103

Notes: This table reports the effect of a one-SD increase in classroom, school, principal, or teacher quality on students' subject-level test scores and the average effect. Test scores are measured in standard deviations and are standardized at the grade–year–subject-level. To arrive at these numbers, I follow the procedure outlined in Section 3.1.

Table 1.3: Out-of-Sample Validation:
Future Principal Value-Added and Student Test Scores

	(1) Mean Score	(2) Mean Score	(3) Mean Score
$\geq$ 2 Years Before Switch	0.073	0.079	0.314***
	(0.055)	(0.056)	(0.053)
Year of Switch	0.937***	0.910***	1.214***
	(0.046)	(0.048)	(0.058)
School FE	Yes	Yes	No
Controls	Yes	No	No
P-value for Coeff = 1	0.172	0.059	0.0001
Observations	1,242,217	1,242,217	1,242,218

Notes: Each column reports the coefficients from a regression of PVA at time $\tau = 0$ on student test-scores two years prior to the student transition and the year of the transition. Controls refer to the set of observable student characteristics and the corresponding classroom- and school-level means as defined in 1.6. Standard errors in parentheses are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.4: Out-of-Sample Validation:
Changes in Principal Quality and Changes in Mean Scores

	(1) Math	(2) Reading	(3) Mean	(4) Math Placebo	(5) Reading Placebo	(6) Mean Placebo
Math PVA	0.649***			0.0763		
	(0.192)			(0.154)		
Reading PVA		0.830***			0.105	
		(0.309)			(0.303)	
Mean PVA			0.824***			0.192
			(0.238)			(0.218)
School FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
P-value for Coeff = 1	0.0685	0.583	0.459			
Lag Score Change	Yes	Yes	Yes	No	No	No
Observations	1,719	1,719	1,719	1,719	1,719	1,741

Notes: Each column reports the coefficients from a regression of changes in mean school test scores on changes in PVA. Change in value-added is the difference between the incoming and exiting principals' value-added at time t and is estimated excluding data from years t and $t-1$, respectively. Mean test scores are the average school performance in math and reading, while mean PVA is the average effectiveness across math and reading. Standard errors in parentheses are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.5: Faculty and School Characteristics:
Above- and Below-Median PVA

	Above Median		Below Median		
	Mean	SD	Mean	SD	P-value
Principal Characteristics					
Annual Salary	66.545	15.49	65.444	15.33	0.00
Female	0.615	0.49	0.543	0.50	0.00
Age	48.117	7.80	47.705	7.62	0.06
Principal Experience	3.233	2.36	2.850	2.24	0.00
White	0.747	0.43	0.704	0.46	0.00
Top 10 or Ivy	0.007	0.08	0.003	0.05	0.04
School Ranked in *USN&WR*	0.103	0.30	0.071	0.26	0.00
BA and Above	1.000	0.00	1.000	0.00	.
MA and Above	0.997	0.06	0.994	0.08	0.55
Doctorate	0.158	0.36	0.157	0.36	0.09
Class Size	31.094	15.44	34.836	17.84	0.00
TVA Math	0.043	0.27	-0.020	0.24	0.00
TVA Reading	0.026	0.23	-0.005	0.21	0.00
School Characteristics					
Male	0.505	0.03	0.508	0.04	0.01
Eligible for FRPL	0.516	0.21	0.564	0.19	0.00
Black or Hispanic	0.400	0.26	0.397	0.26	0.60
Limited English Proficiency	0.062	0.07	0.048	0.06	0.00
Teacher Characteristics					
Female	0.893	0.08	0.868	0.10	0.00
Age	41.707	3.19	41.970	3.07	0.00
Experience	12.063	2.81	12.196	2.84	0.36
Rookie Teacher	0.067	0.04	0.070	0.05	0.02
TVA Math (School)	0.032	0.06	-0.035	0.06	0.00
TVA Read (School)	0.029	0.05	-0.020	0.06	0.00
Turnover Rate	0.276	0.14	0.305	0.16	0.00
Turnover Rate: Over 65	0.003	0.02	0.004	0.02	0.22
Avg TVA Math Exiters	0.026	0.13	-0.067	0.13	0.00
Avg TVA Reading Exiters	0.025	0.12	-0.038	0.11	0.00
Avg TVA Math Entry	0.023	0.15	-0.055	0.16	0.00
Avg TVA Reading Entry	0.022	0.13	-0.031	0.12	0.00
N Principals	3,257 (636)		3,292 (689)		

Notes: This table reports summary statistics for principals with above- and below-median value-added (across math and reading). Principal (school) statistics reflect averages across all principal (school) years. Teacher characteristics are first calculated at the school level then collapsed to the principal level. Annual salary is indexed to 2019. School Ranked in *USN&WR* is an indicator for whether an individual's highest degree–granting institution was ranked in *US News & World Report*'s "Historical Liberal Arts College and University Rankings" as of 2023 (accessed via http://andyreiter.com/datasets/text). Rookie teacher indicates whether a newly hired teacher had never taught before. Parentheses show numbers of teachers who became principals.

Table 1.6: Relationship between Principal Characteristics and Mean Principal Value-Added

	(1)	(2)	(3)	(4)	(5)	(6)
Age	-0.000069	-0.00045	0.00015	-0.000095	0.00017	-0.00093
	(0.00017)	(0.00037)	(0.00016)	(0.00038)	(0.00017)	(0.00070)
At Least 4 Years of Exp	0.014***	0.0019	0.012***	0.00094	0.0091***	-0.013
	(0.0024)	(0.0056)	(0.0022)	(0.0053)	(0.0024)	(0.012)
School Ranked in *USN&WR*	0.018***	0.014	0.0070*	-0.00057	0.0033	0.014
	(0.0040)	(0.010)	(0.0040)	(0.012)	(0.0042)	(0.021)
Female	0.012***	-0.0035	0.0089***	0.00066	0.0016	-0.011
	(0.0024)	(0.0055)	(0.0023)	(0.0051)	(0.0025)	(0.0100)
White	0.014***	0.014**	0.012***	0.012**	0.0065**	0.033**
	(0.0030)	(0.0062)	(0.0029)	(0.0062)	(0.0032)	(0.014)
Doctorate	0.00063	0.0047	0.0016	0.0046	0.0085***	0.012
	(0.0033)	(0.0085)	(0.0031)	(0.0083)	(0.0031)	(0.015)
Graduated from NC University	0.00069	0.015*	0.0049	0.026***	0.0064*	0.0036
	(0.0034)	(0.0086)	(0.0034)	(0.0094)	(0.0035)	(0.021)
Previously Observed Teaching	-0.0038		-0.0018		-0.0030	
	(0.0031)		(0.0030)		(0.0030)	
Mean TVA		0.22***		0.14***		0.14**
		(0.046)		(0.046)		(0.065)
Fixed Effects	None	None	District	District	School	School
Observations	6,471	1,270	6,471	1,259	6,213	535
R^2-within	0.017	0.031	0.012	0.024	0.0092	0.062
F	11.3	4.37	8.64	2.92	4.49	2.15

Notes: This table reports estimates of the association between mean PVA and principal characteristics. Columns 1, 3 and 5 are estimated in the sample of principals containing the full set of observable characteristics. Columns 2, 4, and 6 are estimated in the subsample of principals who previously worked as teachers in North Carolina. I use mean TVA, instead of math TVA, since over 95% of principals have value-added in both math and reading. Standard errors in parentheses are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.7: Relationship between Mean PVA and Log Salary

	(1)	(2)	(3)	(4)
Age	0.0093***	0.0093***	0.0093***	0.0097***
	(0.00070)	(0.00069)	(0.00069)	(0.00097)
At Least 4 Years of Exp	0.064***	0.061***	0.060***	0.056***
	(0.0079)	(0.0076)	(0.0077)	(0.011)
School Ranked in *USN&WR*	-0.0095	0.0098	0.0096	0.0017
	(0.017)	(0.017)	(0.017)	(0.024)
Female	-0.022***	-0.024***	-0.024***	-0.026**
	(0.0083)	(0.0083)	(0.0083)	(0.012)
White	0.024**	0.018*	0.017*	0.0071
	(0.0099)	(0.010)	(0.010)	(0.016)
Doctorate	-0.00040	-0.0034	-0.0034	-0.0013
	(0.012)	(0.012)	(0.012)	(0.018)
Graduated from NC University	-0.020*	0.0033	0.0032	-0.0076
	(0.010)	(0.010)	(0.010)	(0.016)
Mean PVA	0.17*		0.049	0.091
	(0.10)		(0.11)	(0.17)
Fixed Effects	None	District	District	School
Observations	6,423	6,423	6,423	6,423
R^2 (Adjusted)	0.064	0.11	0.11	0.052
F	50.8	50.5	46.1	21.3

Notes: This table reports estimates of the association between mean principal salary and principal characteristics. Mean PVA is the average value-added across math and reading. Age and experience are the average values associated with a given principal. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.8: Working Conditions Survey Responses:
Above- and Below-Median PVA

	Above Median		Below Median			
	Mean	SD	Mean	SD	P-value	N-Principals
Overall (Aggregate) Ratings						
Overall School Climate	4.062	0.43	3.942	0.46	0.00	4,855
Principal Leadership	3.947	0.38	3.828	0.39	0.00	5,142
Teacher Empowerment	3.612	0.39	3.544	0.38	0.00	5,142
Principal Leadership						
Mutual vision among staff	3.992	0.47	3.829	0.50	0.00	5,040
Supports teachers	3.908	0.49	3.781	0.51	0.00	5,142
Effective leadership	3.887	0.49	3.758	0.50	0.00	5,142
Trust teachers	3.881	0.40	3.860	0.40	0.00	4,855
Cares about leadership	3.717	0.46	3.608	0.47	0.00	5,142
Mutual respect	3.697	0.55	3.549	0.56	0.00	5,142
Teacher Empowerment						
Teachers held to high standards	4.393	0.32	4.244	0.32	0.00	5,142
Consistent evaluation	4.029	0.40	3.942	0.40	0.00	4,855
Teachers set curriculum	3.938	0.41	3.895	0.40	0.00	1,625
Principals empower teachers	3.842	0.42	3.747	0.42	0.00	5,142
Opinions matter	3.493	0.45	3.407	0.44	0.00	5,142

Notes: This table reports the average rating that a principal receives on questions related to overall school climate, principal leadership, and teacher empowerment. With the exception of overall school climate, each category has a variety of subquestions. For all questions, teachers are asked to express on a scale of 1 through 5 how much they agree with various statements, with 1=“strongly disagree" and 5=“strongly agree". An exception is the 2002 survey, which utilizes a 1–6 scale.

Chapter 2

The Impact of Tenure Removal on Teacher's Labor Supply Responses

1. Introduction

Non-monetary aspects of a job, have become increasingly important in attracting qualified applicants (Mas and Pallais, 2020; Le Barbanchon et al., 2021; Maestas et al., 2023; Angelici and Profeta, 2024). A relevant, but controversial, example is tenure, or career status, which provides teachers with greater job protections and safeguards against dismissals unrelated to productivity. Proponents of tenure argue that it allows schools to attract talented teachers through compensating differentials. Critics contend that steep job protections lower effort and serve as a costly barrier to removing ineffective teachers.

These opposing views have manifested themselves in public discord. Over the past three decades, over 200 tenure laws, typically aimed at weakening job protections, have been passed (Ng, 2021). These laws have important implications on the structure of teacher quality, but quantifying the impact of tenure elimination is difficult since strong teacher unions limit the ability to overhaul teacher contracts. Most of the existing empirical work focuses on temporary or small changes to tenure protection[1] making it difficult to extrapolate these results. Career decisions and labor supply responses differ when a policy is permanent or significantly affects expected future compensation[2] (Kraft, 2015; Carruthers et al., 2018; Ng, 2021; Anderson et al., 2022).

This paper fills in that gap. In 2013, the North Carolina legislature passed Senate Bill 402 which abruptly closed tenure paths for some existing teachers and eliminated tenure for all new hires. I leverage this policy to study how large changes in job security affect the composition of new teachers and teacher effort. My analysis reveals that new teachers entering after the policy were approximately 4% and 2% of a standard-deviation less effective on student math and reading test scores compared to their existing counterparts, even after accounting for differences in experience. Importantly, I find that these gaps grow over time suggesting serious long-term effects on the quality of future teachers. Regarding the role of job protections and worker effort, I find limited evidence that increasing job protections decreases effort as teacher value-added (TVA) parallels the pre-tenure period and follows the same trajectory as teachers who were ineligible for tenure but entered before the policy.

To answer these questions, I draw from rich administrative data from the North Carolina Education Research Data Center (NCERDC) which contains records on over two million students and 33,000 teachers in North Carolina spanning from 1995 to 2019. These classroom records allow me to construct estimates of teacher value-added and control for differences in teacher experiences (Chetty et al., 2014a; Bau and Das, 2020; Bates et al., 2022). I then use detailed information on teacher characteristics to examine whether changes in value-added

[1] For instance, Ng (2021) studies how extending the teacher probationary period by an additional year affects teacher effort and selection.

[2] Weakening job protections can impact future earnings, as the probability of being employed drastically decreases, thereby affecting expected future earnings.

are driven by changes in observable teacher characteristics or reflect a shift in the selection to teaching.

A central finding is that eliminating tenure harmed the quality of incoming teachers with gaps growing over time. I argue that this decline largely reflects changes in teacher sorting patterns as teachers entering after Senate Bill 402 are similar in terms of teaching background to existing teachers. Relative to existing teachers, those entering after 2013 graduated from similar quality universities, as measured by the natural log of their Alma mater's rejection rate and the bottom quartile of incoming SAT scores. However, post reform teachers were less likely to have graduated from a North Carolina institution which is consistent with the policy gaining more attention at the local level. Additionally, post reform teachers were nearly 6 percentage points more likely to be Black or Hispanic but equally likely to be female.

The second set of analyses examines how job protections affect worker effort. To this end, I construct time-varying estimates of TVA and analyze how it evolves before and after tenure receipt. In both math and reading, I find limited evidence that teacher effort decreases upon achieving career status, as TVA remains similar to the year before tenure. Using cohorts who arrived before the policy reform, and thus were unaffected by potential selection issues, I find limited evidence that the plateauing effects reflect teachers shirking or strategically changing their efforts. The TVA for both tenure-eligible and ineligible cohorts follows a nearly identical trajectory before and after tenure receipt, assuaging concerns of strategic behavior.

This paper contributes to the literature in three ways. First, it provides one of the first sets of results examining how eliminating tenure affects the composition of incoming teachers. Since eliminating tenure is rare due to the strength of teacher unions, prior work focuses on the labor market effects of policies that temporarily weaken teacher protections or extend the teacher probationary period (Jacob, 2013; Goldhaber et al., 2016; Loeb et al., 2015; Ng, 2021). In the study closest to mine, Carruthers et al. (2018) examine the short-run impacts of tenure removal in Florida and do not find evidence of teacher compositional effects in the two years after the policy. However, career decisions are shaped by expected future earnings, and later cohorts can more flexibly adjust to the policy (Lazear, 2003, Rothstein, 2015). Therefore, examining the policy's effect in the immediate short term likely masks changes in labor supply decisions. My analysis fills this gap and shows that the effects of tenure removal are sharpest in the medium to long run.

Second, this paper contributes to the literature examining the role of tenure and teacher effort. In a recent study, Ng (2021) examines how extending the probationary period by an additional year affects teacher effort in New Jersey. My results are broadly consistent as both find limited evidence that tenure reduces teacher effort in the long run, though he does find small transitory effects in math. In a related study, Boboshko (2021) examines how the implementation of tenure impacts student academic achievement across the United States. While he cannot disentangle changes in student outcomes due to teacher selection or effort, he argues that the adverse effects of tenure occur when there is a teacher surplus which may explain these findings. In my setting, I find that after Senate Bill 402, teachers were less likely to graduate from North Carolina institutions suggesting that North Carolina students might have been more aware of the policy or the difficulty of the teacher labor market further emphasizing the role of information in occupational choice.

Finally, this paper contributes to the value of non-monetary benefits in the workplace. I focus on the role of decreased job protections and examine how this affects worker productivity. Across a variety of settings, existing work has found mixed evidence on the relationship between job protections and productivity (Bjuggren, 2018; Kaur et al., 2021; Besley and Burgess, 2004; Autor et al., 2007; Boboshko, 2021; Dinerstein and Opper, 2022). My analysis extends prior work by examining this question in the United States and considering a much larger set of workers and its effect on selection. Broadly, my results suggest that changes in productivity or outcomes reflect changes in the composition of future workers rather than declines in effort by existing ones.

The remainder of the paper is organized as follows. Section 2. describes the policy and its ramifications. Section 3. details the NCERDC data. Section 4. examines how eliminating tenure affects teacher selection. Section 5. explores how tenure receipt affects teacher effort levels. Section 6. concludes.

2. Policy Background

Prior to May 2013, teachers in North Carolina were eligible for career status, or "tenure", after four years of service at a given school district. Career status effectively provided teachers with greater job protections as they could not be "cannot be dismissed except for proper cause" and ended their probationary period; however, career status was at the district level meaning that a teacher's tenure clock would reset if they moved to a different district.[3] These job protections effectively made it very costly for administrators to dismiss teachers increasing overall job stability.

In May 2013, North Carolina passed Senate Bill 402 which eliminated tenure paths for teachers entering the profession after August 2013 and those who had yet to achieve career status. This meant that teachers who entered the profession between academic years 2011

[3]See Appendix 2.1 for a description of career status from the North Carolina Department of Public Instruction.

through 2013 lost their ability to obtain tenure[4]. Beyond just eliminating career status pathways, Senate Bill 402 also proposed eliminating teacher career status entirely. When first passed, the Bill sought to phase out tenure by the 2018 school year with teachers placed on temporary contracts at the discretion of each school district[5]. However, this proposal was never fully enacted as in June 2015, the North Carolina Court of Appeals ruled that policymakers could not retroactively remove career status for teachers who had already obtained it.

Figure 2.2 provides a visual representation of the affected and unaffected cohorts this paper studies. Since four years of teaching experience was required for career status, teachers entering the profession between the 2011 and 2013 academic years could not obtain career status since they had, at most, three years of experience when the policy was implemented. The 2014 cohort marked the first set of new teachers who could not obtain career status. Given that the policy could not retroactively remove career status for tenured teachers, this meant that the last cohort to achieve this status were teachers entering in 2010 as they would have accrued four years of experience.

3. Data and Sample Description

3.1 Data Description

I use administrative microdata from the North Carolina Education Research Data Center to examine how eliminating teacher tenure affects the quality and composition of new teachers and whether tenure receipt reduces teacher effort. The data contain detailed information on the near universe of students enrolled in North Carolina public schools from 1995 to 2019,

[4]Academic year t refers to the calendar year time beginning August or September of year t through June of year $t+1$.

[5]The Senate Bill also eliminated pay increases for teachers with a master's degree. Specifically, those who entered a master's program after August 2013 were ineligible for pay increases while existing teachers with advanced degrees were unaffected.

including their background information as well as detailed records on teachers and their classroom assignments.

Like other administrative datasets, the NCERDC dataset contains student-level variables, including end-of-grade test scores in math and reading and demographic information such as race and ethnicity, gender and indicators for whether a student is eligible for free or reduced-price lunch, academically gifted status, and whether a student repeated a grade or course. The data also allow me to link teachers to students to form classrooms, construct classroom-level controls, and estimate teacher value-added. Relevant staff-level data include demographics, experience, and where teachers completed their formal education.

3.2 Sample

The key variables for my analysis are experience and when an individual joined the North Carolina teaching force as this allows me to identify whether a teacher was, or would have been, eligible for tenure. While the NCERDC provides an experience variable, Jackson (2018) notes this may be subject to mis-recording and that this experience measure grants teachers experience even if they do not assume a teaching role in a year[6]. To address this, I create a new experience variable that only scales with each year of teaching. I use the NCERDC Pay files, which contain payroll information on all individuals employed by the North Carolina Department of Public Instruction, to see if an individual was employed as a teacher and scale experience accordingly. Note that my created variable should be weakly smaller than the provided measure since I only scale with teaching experience. The correlation between the provided experience variable and the self-created is about 91% suggesting that my measure broadly tracks the raw data.

[6]For instance, a teacher may take a non-teaching or assistant role and still accrue experience under the NCERDC provided variable.

To identify whether teachers entered before the policy reform, I define a teacher's "cohort" as the first year a teacher is observed in the Pay file dataset, regardless of whether or not she was employed as a teacher. Since the data begin in 1995, I cannot identify a true cohort year for a subset of teachers. Specifically, I cannot identify a teacher's cohort year if she had multiple years of experience at the start of the 1995 academic year. For such individuals, I assign them to the catch-all cohort that arrived prior to 1995. Figure 2.1 plots the distribution of teachers over time using the constructed cohort year measure. Perhaps unsurprisingly, fewer teachers enter the profession in the years following Senate Bill 402 providing preliminary evidence of the policy's impact on the teacher's labor market.

My analysis focuses on teachers and students in public elementary and middle schools (grades 4-8). To identify classroom assignments, I follow Jackson (2018) and Rose et al. (2022) and use the NCERDC provided "Course Membership" files, which directly link teachers to students for the academic years 2007–2019. For the remaining years from 1995 to 2006, I follow Rothstein (2017) and use the provided "End-of-Grade" files, which directly link students to end-of-year testing proctors as proxies for the student's true teacher.

After matching teachers to students, I impose additional sample restrictions. I limit my sample to students with valid test scores in the prior year (prior-year test scores are key controls in my construction of value-added). To mitigate potential mismatches of teachers to students, I limit the sample to classrooms with 15–100 students (Rose et al., 2022). Finally, I drop observations in which teachers teach at multiple schools or in multiple grades in a year as their effects since their students are only partially exposed to their effects or may capture the impacts of substitute teachers. After I impose these restrictions, my final sample consists of over two million students and nearly 34,000 teachers.

3.3 Sample Description

Table 2.1 reports summary statistics corresponding to the full analytical sample. The first half of the panel reports student characteristics, and the second half reports teacher characteristics. Roughly 37% of students in the sample are Black (about 25%) or Hispanic (about 12%) and about half the students are eligible for Free or Reduced Price Lunch (FRP Lunch). Test scores are normalized to be mean zero with a standard-deviation of one; however average test scores are slightly larger at 0.04 and 0.03 for math and reading scores, respectively, due to the exclusion of students without valid prior test scores (Rose et al., 2022). Additionally, about five percent of students are classified as Limited English proficient, and about 14% percent are classified as academically gifted. Broadly, students in North Carolina resemble their national counterparts, though fewer students here as classified as limited English proficient, suggesting strong external validity.

Regarding teacher characteristics, North Carolina instructors are predominately female - even more so than the national average of 77% (NCES, 2020a), and about 14% of teachers are Black or Hispanic. Using the NCERDC provided measure, teachers have over ten years of experience on average. When using my constructed measure of teaching experience, this measure falls slightly to just over nine years. In terms of Alma mater quality, the average rejection rate in the sample is about 37 percent (moderately selective) and the bottom quartile math SAT score is 480.

4. Tenure Reform and Teachers' Labor Supply Response

I divide the task of quantifying the impacts of eliminating tenure protections on teacher labor supply decisions into two parts. First, Section 4.1 describes and constructs my measure of teacher quality: teacher value-added. Second, Section 4.2 then examines how value-added and observable teacher characteristics evolve around the policy implementation to examine how the policy shaped the incoming teacher composition.

4.1 Quantifying Teacher Quality

This section constructs a measure of teacher-effects on student test scores to better understand how removing tenure might impact student academic performance. To this end, I follow the literature and construct estimates of value-added as my measure of teacher quality (Chetty et al., 2014a; Jackson, 2018; Bau and Das, 2020; Rose et al., 2022). Each teacher has a subject specific value-added. I estimate the equation below for math and reading separately:

$$y_{i,g,j,t} = \beta_0 + \sum_{a} \beta_a y_{i,t-1} \mathbf{I}\{\text{grade} = a\} + X'_{i,t}\delta + X'_{j,s,t}\tau + \text{Exp}_{j,t}\lambda + \alpha_t + \alpha_g + \theta_j + \epsilon_{i,g,j,t} \quad (2.1)$$

where the subscripts i denote a student, g a grade, j a teacher, and t a year. The outcome variable $y_{i,g,j,t}$ is a student's contemporaneous test score and is standardized at the grade-year-subject level.

θ_j is the key parameter of interest capturing the averaging test score gain from being assigned to teacher j conditional a variety of controls. Teacher value-added corresponds to estimates of θ_j, a time-unvarying teacher fixed-effect. Prior work by Rockoff (2004), Wiswall (2013), Bau and Das (2020) find that inexperienced teachers, in particular those with less than three years of experience, tend to have smaller test-score impacts[7]. This implies that the value-added of teachers unaffected by the policy will look mechanically better because they have more years of experience relative to the affected cohorts. To address this, $\text{Exp}_{j,t}$ is a vector of indicator variables for teacher experience based on my constructed measure. I include dummies for discrete experience values between zero through four, and an indicator

[7] Appendix Figure A.10 also shows that such patterns emerge in this setting.

for at least five years of teaching experience in a given academic year. With these indicators, one may then interpret $\hat{\theta}_j$ as teacher effects on student test scores net of experience.

The identifying assumption to estimate teacher value-added is the controls sufficiently account for student sorting into classrooms. To ensure this assumption holds, I include lagged student test scores, $y_{i,t-1}$, interacted with grade indicators, which prior work has shown to address teacher sorting across schools Chetty et al. (2014a) and within-school sorting of students to classrooms Jackson (2018). Furthermore, $X'_{i,t}$ is a vector of student characteristics such as gender, age, Free or Reduced Price Lunch eligibility, indicators for academically gifted, and whether a student repeated a subject to account for the possibility that prior student achievement fails to fully capture observable student quality. To parse out the role of observed and unobserved school and classroom effects driving estimates of θ_j, I follow Altonji and Mansfield (2018) and include the vector $X'_{j,s,t}$ which contains classroom and school-level means of the same student characteristics[8]. If values are missing for a particular student, then the variable is equal to zero and an indicator denoting missing is included. Following the literature, α_t and α_g are year and grade fixed effects that account for common shocks across all cohorts and potential differences in testing rigor across grades.

4.2 Tenure Reform and Entering Teacher Quality

After obtaining estimates of teacher value-added, net of experience, I next examine how tenure removal impacts the quality and observable characteristics of incoming teachers. Absent increases in teacher compensation, removing teacher tenure protections should adversely affect the quality of incoming teachers since the outside option of teaching becomes relatively more appealing. To test this hypothesis, I examine how teacher value-added and observable

[8] I also provide alternative specifications that include school fixed-effects and identify θ_j using the subset of teachers who switch schools. The results are nearly identical to using the group-level means approach.

characteristics evolve around the policy's implementation. I estimate:

$$y_j = \beta_0 + \sum_{c \neq 2013} \mathbf{I}\{\text{cohort} = \text{c}\}\beta_{j,c} + \epsilon_j \tag{2.2}$$

where y_j are teacher characteristics such as their value-added, separately, in math and reading, an indicator for whether she graduated from a North Carolina institution, the natural log of her Alma mater's rejection rate, the natural log of the the 25th percentile SAT math score for her Alma matter, and indicators for female and underrepresented groups (Black or Hispanic). $\mathbf{I}\{\text{cohort} = \text{c}\}$ indicates whether teacher j entered the profession in year c using the cohort definition described in Section 3.1. Point estimates are relative to the 2013 cohort, the year before the policy's implementation, and cohorts entering before 2008 are grouped into a single indicator for simplicity.

Figure 2.3 provides visual evidence of the policy's effect on the value-added of teachers across cohorts where the blue lines estimate TVA using equation 2.1. Relative to cohorts unaffected by the policy, those entering after 2013 were less impactful on math, statistically significant at an α-level of 5% level for most cohorts, and reading, with the declines in teacher quality growing over time. Collapsing the analysis to a simple binary comparison of cohorts affected versus unaffected by the policy indicates that tenure reform reduced teacher effectiveness by 3.7% and 1.7% standard-deviations in math and reading[9]. To ensure that $\hat{\theta}_j$ is not capturing differences in school quality, the red lines provide point estimates that construct TVA with a school fixed effect. Across both subjects, the point estimates are nearly identical to the main specification indicating that the main controls sufficiently account for observed and unobserved school heterogeneity.

[9] The β estimates correspond to a simple regression of $y_{j,c} = \beta_0 + \beta\text{Treat}_{j,c} + \epsilon_{j,c}$ where the outcome variable is TVA in math or reading and Treat is an indicator for whether a teacher was part of the 2014 cohort or later. The sample is limited to cohorts graduating from 2007 and after.

Since the impact of Senate Bill 402 might take time to manifest, Figure 2.3 also illustrates the dynamic effects of the policy. While post-reform teachers were less effective than their counterparts, the effects are strongest for the most recent graduates. The widening gaps are unsurprising as later cohorts of prospective teachers had more time to respond to the policy and could change their majors as the new returns to teaching had fallen. This finding is consistent with Wiswall and Zafar (2015) who find that beliefs about one's ability are a key determinant of college major choice and Xia (2016) who finds that major choices respond to perceived expected earnings, implying that making a job less stable drives out potential teachers. The policy's permanent and *full* elimination of tenure likely explains why my results differ from prior work. Most noticeably, Ng (2021) finds that extending the pre-period tenure period from three to four years in New Jersey did not affect the value-added of incoming teachers. As the New Jersey policy did not eliminate tenure, it is reasonable to expect larger effects in this setting as prospective teachers face a much tighter set of constraints.

The change in teacher effectiveness may be driven by changes in teacher composition. For instance, if more selective universities produce more effective teachers, then a decline in teacher effectiveness may result from teachers graduating from less prestigious universities. To examine whether compositional shifts drive my analysis, I estimate equation 2.2 with various teacher characteristics as the outcomes of interest. Figure 2.4 shows the evolution of teacher characteristics following tenure removal. Relative to cohorts who arrived before the policy, those entering after 2013 were less likely to have graduated from a North Carolina institution, suggesting they were more aware of the policy and local teacher labor market conditions. Table 2.2 reports the point estimates from a collapsed regression for the treated and untreated groups[10]. It finds that, relative to cohorts unaffected by the policy, new teachers entering the profession after 2013 were 2.5 percentage points less likely to graduate from an in-state university.

[10]See Footnote for 9 for details.

Despite geographic changes in *where* teachers graduated, Panels 2.4b and 2.4c provide mixed evidence that post reform teachers were more likely to come from less-selective universities. Those entering after 2013 graduated from universities with comparable admissions rates and standardized testing scores, indicating that the observed decline in teacher value-added effects likely reflects changes in teacher selection patterns rather than differences in observable ability or human capital investments.

Finally, in terms of demographics, post reform teachers were similar in terms of gender but were more likely to be either Black or Hispanic. Table 2.2 indicates that post reform teachers were 5.6 percentage points more likely to be Black or Hispanic, up from a baseline mean of 14 percent, indicating a large shift in the demographics of teachers. Broadly, Figure 2.4 finds that while post reform teachers were more likely to come from underrepresented backgrounds, they were not more likely to come from more selective universities. These results suggest that the observed decline in teacher effectiveness is unlikely driven by shifts in observable teacher characteristics but instead reflects changes in prospective teacher labor supply decisions or, simply, *who* becomes a teacher.

5. Tenure Receipt and Teacher Effort

A consequence of Senate Bill 402 was the elimination of tenure paths for some recently arrived teachers. Prior to the reform, teachers were eligible for career status, at the district-level, after four years of working in the same school district. This meant teachers starting their careers between the academic years 2011 through 2013 were ineligible for greater job protections but were not adversely selected into the profession since they arrived before the policy.

To study whether tenure receipt reduces teacher efforts, I modify equation 2.1 and allow teacher value-added to vary over time[11]. I then use these estimates to examine how TVA evolves in the years surrounding tenure receipt. I estimate the following regression for the cohorts of teachers eligible for tenure (those entering before 2011):

$$\hat{\theta}_{j,t} = \alpha_0 + \sum_{e \neq 3} \beta_e \mathbf{I}\{\text{years} = e\} + \alpha_t + \epsilon_{j,t} \tag{2.3}$$

where $\hat{\theta}_{j,t}$ is the estimated teacher value-added for teacher j in year t. β_e captures the difference in teacher value-added relative to her first tenured year[12]. α_t is a year fixed-effect to account for common shocks that may affect teacher productivity in a given year.

Figure 2.5a reports estimates of β_e for TVA, separately, for math and reading. Across both subjects, teacher test score impacts remain stable upon tenure receipt as the estimates in the years after tenure are indistinguishable from those in the year prior to tenure receipt. While not statistically significant, the point estimates are positive limiting the scope for teacher shirking. These results stand in contrast to the critics of tenure who argue that teachers reduce their efforts once they are tenured. If anything, point estimates corresponding to three or more years after tenure, indicate that teachers potentially increase their effort later on or improve.

One explanation behind the limited TVA change is that teachers have discovered the optimal amount of effort to exert and are simply targeting a given output. In other words,

[11]Specifically, I estimate equation 2.1 but estimate TVA as a teacher-year fixed effect $\theta_{j,t}$ and do not include school fixed-effects or experience dummies.

[12]Specifically, I include experience indicators as this is analogous to years around tenure receipt. E.g, as experience is measured at the beginning of the academic year, years = 1 corresponds to a teacher with four years of experience at the beginning of the year

tenured teachers know more about the school's standards and adjust to meet this threshold. To test this hypothesis, I use the cohorts entering between 2011 and 2013 as a control group and examine the evolution of their TVA. Since these cohorts were ineligible for tenure, behavioral responses after their fourth year are unlikely given the threat of dismissal.

Panels 2.5b and 2.5c examine the evolution of teacher value-added on math and reading for tenure eligible and ineligible teachers. Across both subjects, teacher value-added evolves similarly for both groups in the years following tenure. Teachers ineligible for teachers display the same plateauing effect after four years suggesting that the results in Panel 2.5a are unlikely due to teachers adjusting their effort levels, but instead broadly suggest that tenure does not lead to teachers systematically shirking.

6. Conclusion

This paper provides the first set of estimates examining how fully removing tenure affects teachers' labor supply decisions and examines how tenure receipt affects teacher effort. I find that relative to cohorts who entered before the policy's announcement, teachers entering after the reform were nearly 4% and 2% of a standard-deviation less impactful on student math and reading achievement. These behavioral responses are not trivial. The magnitude is similar to changes in teacher effort when teacher salaries are no longer experience-based but instead are correlated with performance (Brown and Andrabi, 2020, Biasi, 2021), which further underscores the role of expected compensation and selection in shaping teacher quality.

These compositional effects have significant medium to long-term implications, as the analysis indicates that declines in teacher effectiveness were most pronounced for later cohorts. As future students have more time to adapt to the policy, the long-term effects likely exceed the estimated declines in teacher effectiveness. This has important consequences, as teachers enhance a variety of cognitive and behavioral skills and can serve as potential role models (Chetty et al., 2014b, Jackson, 2018, Rose et al., 2022, Dee, 2005, Gershenson et al., 2022).

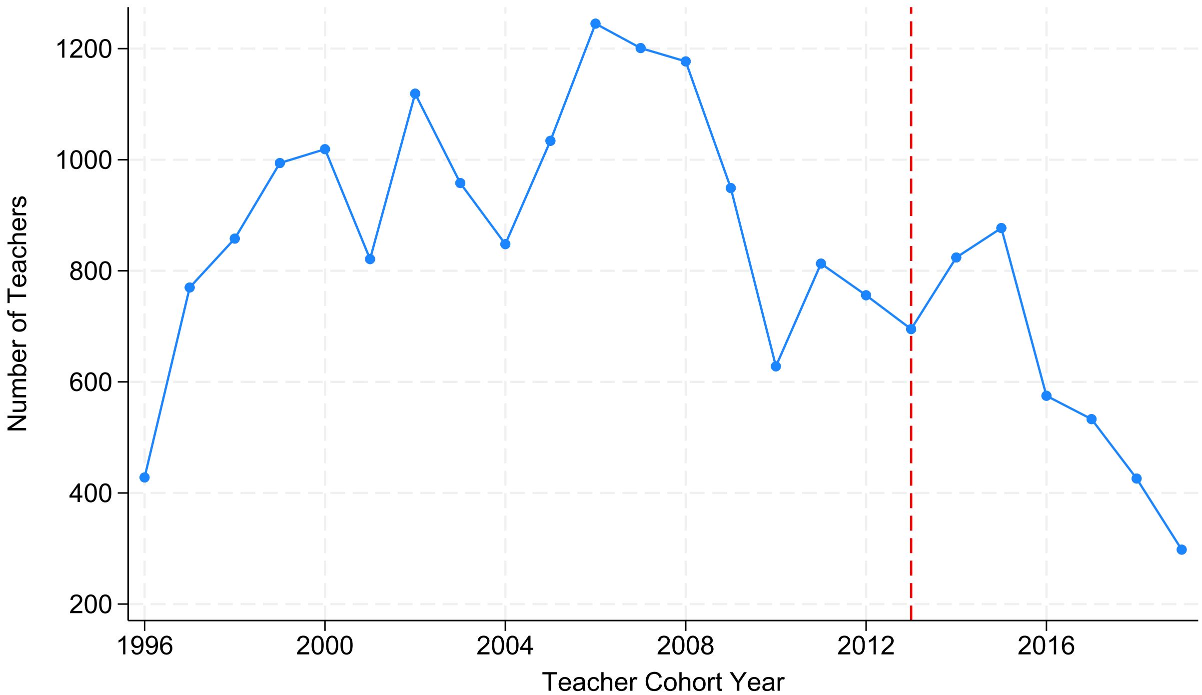
While reducing job protections can harm the pool of prospective teachers, tenure receipt itself is not associated with a decrease in teacher effort. Teacher value-added for cohorts affected and unaffected by Senate Bill 402 evolves similarly, mitigating concerns that teachers are shirking or targeting particular effort levels. The lack of behavioral effects aligns with the common narrative that soft skills or personality traits are key drivers of teacher quality, which are likely more innate and less likely to systematically change (Klein and Others, 2010).

The analysis in this paper provides important insights into how policy can influence labor supply decisions and worker efforts. A central finding is that workers value job protections, but increasing stability does not induce changes in effort. This indicates that eliminating tenure, on the grounds of preventing shirking, is unjustified and can harm the pool of prospective workers. Consistent with theoretical simulations (Rothstein, 2015), eliminating tenure without increasing compensation or providing other non-monetary benefits hurts average teacher quality by dissuading effective applicants from joining the profession. Given this, a natural next step to improving teacher quality is raising teacher salaries or providing alternative work benefits to better attract more teachers.

Finally, future work exploring heterogeneous effects of tenure elimination across districts of varying socioeconomic status may shed new light on teacher sorting patterns. As disadvantaged school districts have more difficulty attracting stronger teachers (Jackson, 2009, Biasi et al., 2021), tenure may provide such districts with greater leverage in attracting teachers or serve more value as teachers are protected in tougher environments.

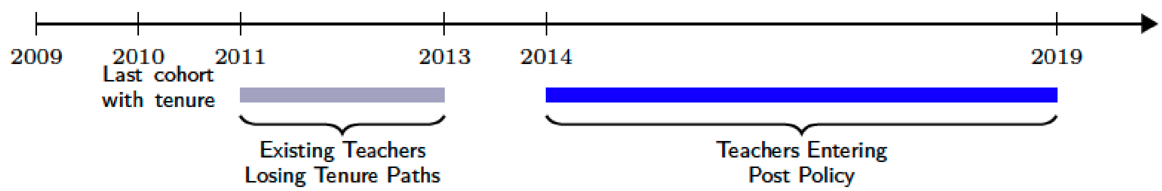
7. Figures

Figure 2.1: Entering Cohort Sizes Over Time



Notes: This figure plots total number of teachers for each cohort year. A teacher's cohort is defined as the first time she appears in the NCERDC's Pay files which contains records on all individuals employed by the North Carolina Department of Public Instruction.

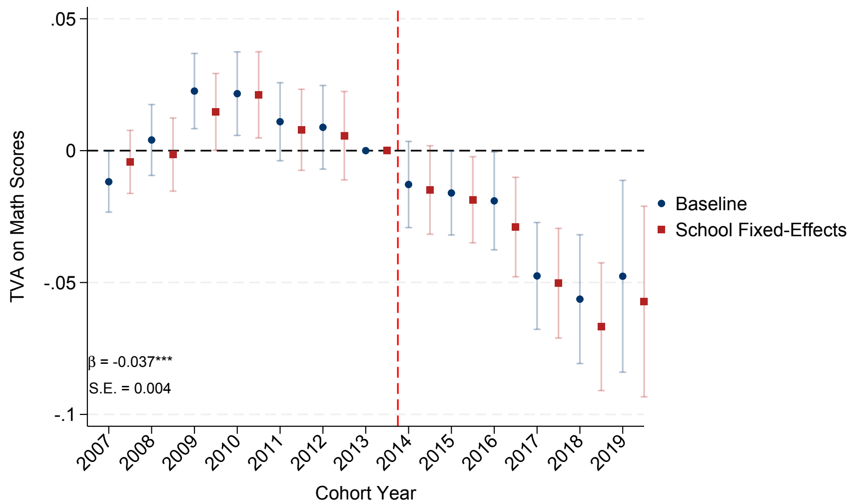
Figure 2.2: Policy Timeline



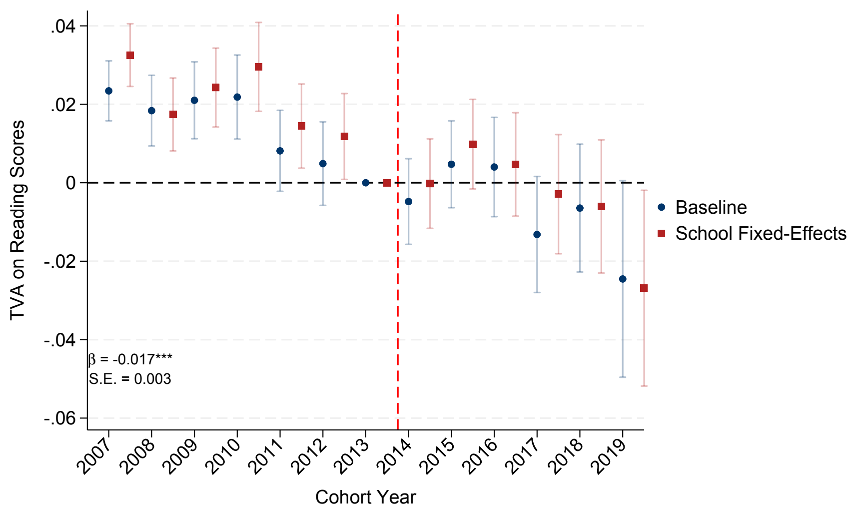
Notes: This figure provides a visual illustration of which cohorts lost their tenure paths due to Senate Bill 402. Prior to the policy, teachers were eligible for career status after four years of experience. As the policy was passed in the summer of 2013, this meant that the last cohort eligible for tenure were those entering in 2010 as they just accrued four years of experience. The 2011 through 2013 cohorts therefore lost their tenure paths. Those entering after the summer of 2013 also could not achieve tenure.

Figure 2.3: Teacher Value-Added by Teacher Cohort Year

(a) Math TVA



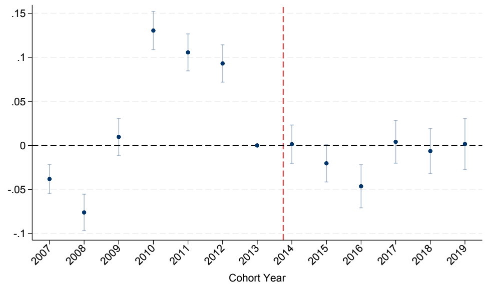
(b) Reading TVA



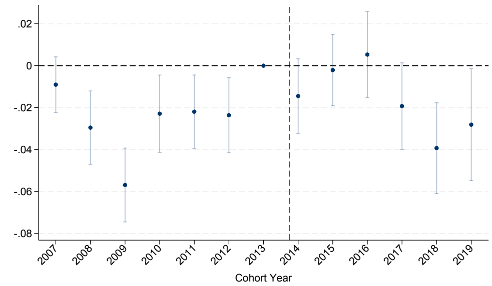
Notes: This figure plots the $\beta_{j,c}$ coefficients from equation 2.2. The Baseline estimates, reported in blue, are those from equation 2.2, whereas the School Fixed-Effects, reported in red, are estimated similarly but include the school fixed effect. Robust standard errors and 95% confidence intervals are reported. The single β and standard error estimates are from the simple regression $y_{j,c} = \beta_0 + \beta \text{Treat}_{j,c} + \epsilon_{j,c}$ where the outcome variable is TVA in math or reading and Treat is an indicator for whether a teacher was part of the 2014 cohort or later. The sample is limited to cohorts graduating from 2007 and after.

Figure 2.4: Tenure Reform and Teacher Characteristics

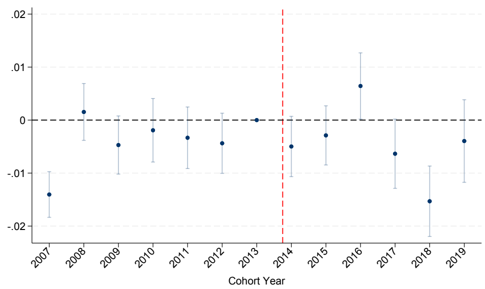
(a) Graduated from NC University



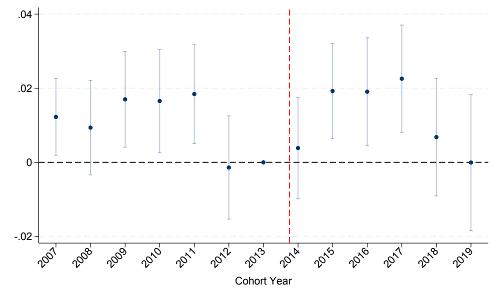
(b) Log(Alma Mater Rejection Rate)



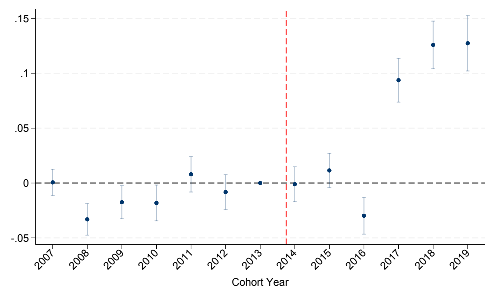
(c) Log(25th Percentile SAT Math Score)



(d) Female



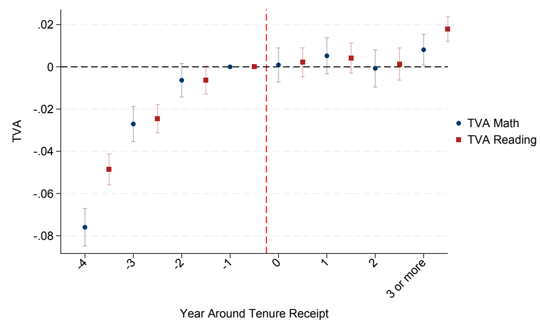
(e) Black or Hispanic



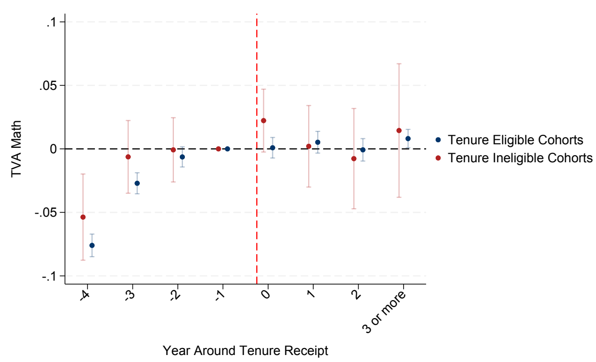
Notes: This figure plots the $\beta_{j,c}$ coefficients from equation 2.2. Robust standard errors and 95% confidence intervals are reported.

Figure 2.5: Tenure Receipt and Teacher Effort

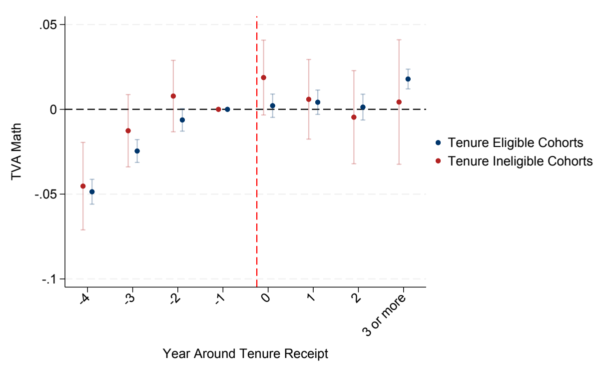
(a) Cohorts Eligible for Tenure



(b) Math



(c) Reading



Notes: This figure plots the β_e coefficients from equation 2.3. Panel A reports the coefficients on the sample of teachers who entered the profession prior to 2010. In panels B and C, the same coefficients are reported in blue (Tenure Eligible Cohorts). For those panels, the red estimates correspond to the coefficients on the subsample of teachers who entered between 2011 and 2013. Robust standard errors and 95% confidence intervals are reported.

8. Tables

Table 2.1: Summary Statistics

	Mean	Standard Deviation
Student Characteristics		
Male	0.50	0.50
Black or Hispanic	0.37	0.48
Eligible for FRP Lunch	0.48	0.50
Limited English	0.05	0.21
Academically Gifted	0.14	0.35
Repeated a Subject	0.02	0.14
Standardized Math Score	0.04	0.98
Standardized Reading Score	0.03	0.98
N Students	2,100,975	
Teacher Characteristics		
Female	0.91	0.28
Experience (Teaching)	9.42	9.39
Experience (Raw)	10.89	8.79
Black or Hispanic	0.14	0.35
Percent of applicants rejected in 2010	36.60	13.67
SAT math score 25th percentile in 2010	480.88	62.30
N Teachers	33,553	

Notes: This table reports summary statistics for the full analytical sample as described in Section 3.2. The variable “Experience (Raw)" refers to the NCERDC provided experience variabvle whereas the “Experience (Teaching)" corresponds to the manually created variable that scales with each year an individual has a teacher position in the NCERDC Pay files. Math SAT information and selectivity comes from Biasi (2021) replication files.

Table 2.2: Tenure Reform and Teacher Characteristics

	(1) Graduated from NC University	(2) Log(Alma Matter Rejection Rate)	(3) Log(25th Percentile Math Score)	(4) Female	(5) Black or Hispanic
Treated	-0.0254***	0.0169***	-0.00203*	0.00151	0.0560***
	(-5.68)	(4.40)	(-1.72)	(0.57)	(16.40)
Observations	49582	47241	46601	49582	49582

Notes: This table reports the β coefficient corresponding to the regression $y_{j,c} = \beta_0 + \beta \text{Treat}_{j,c} + \epsilon_{j,c}$ where the outcome variables are a variety of teacher characteristics and Treat is an indicator for whether a teacher was part of the 2014 cohort or later. The sample is limited to cohorts graduating from 2007 and after. Robust standard errors are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

Appendix and Supplementary Material

1. Appendix to "Principal Quality and Student Outcomes: Evidence from North Carolina"

1.1 Sampling Bias Correction

To obtain a closed-form solution for the sampling error, I follow the approach discussed in Bau and Das (2020). From Section 3., the estimated variance of classroom effects has the form $Var(\hat{\delta}_{j,s,p,t}) = \sigma_j^2 + \sigma_s^2 + \sigma_p^2 + \sigma_{j,s,p,t}^2 + \phi$, where ϕ is the variance of the sampling error. The estimated classroom effects reflect true classroom variation and sampling error: $\hat{\delta}_{j,s,p,t} = \delta_{j,s,p,t} + \frac{1}{N_{j,s,p,t}} \sum_{i=1}^{N_{j,s,p,t}} \mu_{i,s,j,p,t}$, where $N_{j,s,p,t}$ is the number of students in classroom (j, s, p, t). Assume that $var(\mu_{i,s,j,p,t})$ is homoskedastic with variance given by σ_μ^2. Then,

$$
\begin{aligned}
Var(\hat{\delta}_{j,s,p,t}) &= cov(\hat{\delta}_{j,s,p,t}), \hat{\delta}_{j,s,p,t}) && (3.1)\\
&= cov(\delta_{j,s,p,t} + \frac{1}{N_{j,s,p,t}} \sum_{i=1}^{N_{j,s,p,t}} \mu_{i,s,j,p,t}, \delta_{j,s,p,t} + \frac{1}{N_{j,s,p,t}} \sum_{i=1}^{N_{j,s,p,t}} \mu_{i,s,j,p,t}) && (3.2)
\end{aligned}
$$

As noted in Bau and Das (2020), equation 3.2 is equivalent to

$$
Var(\hat{\delta}_{j,s,p,t}) = E[\delta_{j,s,p,t}]^2 + 2E[\delta_{j,s,p,t} \sum_{i=1}^{N_{j,s,p,t}} \frac{\mu_{i,s,j,p,t}}{N_{j,s,p,t}^2}] + E[\frac{1}{N_{j,s,p,t}} \sum_{i=1}^{N_{j,s,p,t}} \mu_{i,s,j,p,t} \times \frac{1}{N_{j,s,p,t}} \sum_{i=1}^{N_{j,s,p,t}} \mu_{i,s,j,p,t}]
$$

Recognizing that $E[\delta_{j,s,p,t} =]E[\mu_{i,s,j,p,t}] = 0$ by construction and that $\mu_{i,s,j,p,t}$ and $\delta_{j,s,p,t}$ are uncorrelated, it follows that

$$
\begin{aligned}
Var(\hat{\delta}_{j,s,p,t}) &= \sigma_{j,s,p,t}^2 + \frac{1}{N_{j,s,p,t}^2} E[\sum_{i=1}^{N_{j,s,p,t}} \mu_{i,s,j,p,t} \times \sum_{i=1}^{N_{j,s,p,t}} \mu_{i,s,j,p,t}]\\
&= \sigma_{j,s,p,t}^2 + \underbrace{\frac{1}{N_{j,s,p,t}} \sigma_\mu^2}_{\phi}
\end{aligned}
$$

where one can estimate ϕ by computing the average residual squared after estimating equation 1.2.

A second approach to examine the variance of principal effects directly examines the equation used to estimate principal value-added, equation 1.6.

$$y_{i,g,t} = \beta_0 + \sum_a \beta_a y_{i,t-1} \mathbf{I}\{\text{grade} = a\} + \theta_p + X'_{i,t}\delta + \alpha_t + \alpha_g + \epsilon_{i,g,t} \tag{3.3}$$

where, for ease of notation, $X'_{i,t}$ is a vector of student controls, as described in equation 1.6 as well as their corresponding school and classroom means. All other parameters are defined as before with the identifying assumptions of θ_p are discussed in Section 4.1.

While $\hat{\theta}_p$ is unbiased when drawn from a random sample, the estimated variances exceed the true variance, as our estimates combine true principal effects with sampling error. To eliminate sampling error bias, I follow Kane et al. (2008), Chetty et al. (2014a), Kline et al. (2020) and Rose et al. (2022), and estimate σ_p^2 by constructing principal–year-level mean residuals from equation 3.3.

$$\bar{Y}_{p,t} = \frac{1}{n_{p,t}} \sum y_{i,t} - X'_{i,t}\hat{\delta} \tag{3.4}$$

$$= \theta_p + \underbrace{\frac{1}{n_{p,t}} \sum \epsilon_{i,g,p,t} + X'_{i,t}(\delta - \hat{\delta})}_{\psi_{p,t}} \tag{3.5}$$

where $n_{p,t}$ is the number of students whom a principal oversees in a given year. Suppose that $\psi_{p,t}$ is uncorrelated across time, implying that student- or cohort-level shocks are uncorrelated over time. Then, the covariance across mean principal–year-level residuals allow one to recover $\hat{\theta}_p$.

Assumption 3. *Uncorrelated principal–year estimation error:* $cov(\psi_{p,t}, \psi_{p,t'}) = 0, \forall t \neq t'$

Assumption 3 implies that the sorting due to unobservable student characteristics is uncorrelated across years for each principal, ruling out that students with “large" unobservable characteristics systematically sort to particular types of principals or, alternatively, that sampling error is uncorrelated across years for a given principal. This assumptions is supported as Section 4.2 finds limited scope for student sorting on unobservable characteristics. Under Assumption 3, the covariance across the mean principal–year-level residuals is equivalent to

$$cov(\bar{Y}_{p,t}, \bar{Y}_{p,t'}) = cov(\theta_p + \psi_{p,t}, \theta_p + \psi_{p,t'}) = \sigma_p^2$$

where, conditional on controls, principal effects are uncorrelated with student unobservables. This procedure allows one to remove the bias from sampling error by omitting the product of the residuals from the same year (Kline et al., 2020, Rose et al., 2022).

To map equation 3.3 back into the framework discussed in Section 3., one can think of the causal model as corresponding to the following equation:

$$Y_{i,s,j,p,t} = \theta^p + X'_{i,t}\tau + \nu_{i,s,j,p,t}$$

where the previous school, teacher, and classroom shocks are now contained in $\nu_{i,s,j,p,t}$. Here, the requirements of Conditional Independence and Uncorrelated Effects are relaxed.

Table A.1 provides the estimates of the variance of principal effects from the alternative decomposition. Using the entire sample of principals, the first row of Table A.1 indicates that a one-SD increase in principal quality raises average student achievement by 0.068 SD. This estimate is nearly a 50% increase from Table 1.2, indicating the difficulty of quantifying principal quality in settings with limited turnover, since the variance of principal effects, likely also contain the impact of school quality.

The second row restricts the decomposition to only principals observed at multiple schools to isolate the potential school effects on the variances of principals. The estimates from this restricted sample are nearly identical to those in Table 1.2, with slightly smaller estimates on the impact of principal quality on student reading exams. Smaller estimates are unsurprising, given that the original decomposition allows for more granular drivers of student achievement (e.g., teacher, classroom, and school effects) beyond just school and classroom-level means of student covariates which should pick up differences in teacher quality that may incorrectly be attributed to principals. The fact that the estimated variances are comparable provides some assurance that the underlying assumptions in Section 3. are credible. Furthermore, estimates from this more parsimonious specification reinforce the finding that prior estimates of school quality largely reflect differences in principal quality as the second row of Table A.1 nets out impact of school effects.

Table A.1: Effects of 1-SD Improvement in Principal Effects

	Math	Reading	Average
All Principals: $cov(\bar{Y}_{p,t}.\bar{Y}_{p,t'})$	0.079	0.057	0.068
Principal Movers: $cov(\bar{Y}_{p,t,s}, \bar{Y}_{p,t',s'})$	0.050	0.040	0.045
Original Estimates	0.050	0.043	0.047

Notes: This table reports the effect of a one-SD higher PVA on students' subject-level test scores and the average effect. Test scores are measured in standard deviations and are standardized at the grade–year–subject level.

To construct empirical Bayes estimates of PVA, I note that, from Section 3.1, student achievement is characterized as:

$$Y_{i,s,j,p,t} = \theta^j + \theta^s + \theta^p + \theta^{j,s,p,t} + X'_{i,t}\tau + \epsilon_{i,s,j,p,t}$$

where θ^j is a teacher shock, θ^s is a school shock, and θ^p is a principal shock, $\theta^{j,s,p,t}$. $X'_{i,t}$ captures observable student heterogeneity, and $\epsilon_{i,s,j,p,t}$ is the idiosyncratic student-specific shock. The variances of these shocks are denoted by σ^2_j, σ^2_s , σ^2_p , $\sigma^2_{j,s,p,t}$, and σ^2_ϵ and are assumed to be independent and homoskedastic.

The object of interest is the expected test score that a student will achieve under a given principal:

$$\gamma_p = \theta^p + \sum_{j \in p} \frac{N_j}{N_p} \theta^j \tag{3.6}$$

where $j \in p$ refers to the set of teachers who have worked for a principal, N_j is the number of students taught by teacher j, and N_p is the number of students taught by principal p. The variance of γ_p is given by

$$Var(\gamma_p) = E[(\theta^p + \sum_{j \in p} \frac{N_j}{N_p} \theta^j)^2] \tag{3.7}$$

since $E[\gamma_p] = 0$ by construction. Under the assumption that θ^p and θ^j are uncorrelated,

$Var(\gamma_p)$ can be rewritten as

$$\begin{aligned} Var(\gamma_p) &= E[(\theta^p)^2] + E[(\sum_{j\in p}\frac{N_j}{N_p}\theta^j)^2] \\ &= \sigma_p^2 + E[\frac{\sum_{j\in p} N_j^2}{N_p^2}\sigma_j^2] \end{aligned} \tag{3.8}$$

Let δ_p be the estimate for γ_p. Then, δ_p is given by:

$$\gamma_p = \theta^p + \frac{1}{N_p}\sum_{i\in p}(\theta^j + \theta^s + \theta^{j,s,p,t} + \epsilon_{i,s,j,p,t}) \tag{3.9}$$

where the $i \in p$ denotes the set of students taught by principal p. Applying the same logic from equation 3.8, the variance of δ_p can be characterized as

$$\begin{aligned} Var(\gamma) &= E[(\theta^p + \frac{1}{N_p}\sum_{i\in p}(\theta^j + \theta^s + \theta^{j,s,p,t} + \epsilon_{i,s,j,p,t})]^2 \\ &= \sigma_p^2 + E[\frac{\sum_{j\in p} N_j^2}{N_p^2}\sigma_j^2] + E[\frac{\sum_{s\in p} N_s^2}{N_p^2}\sigma_s^2] + E[\frac{\sum_{j,s,p,t\in p} N_{j,s,p,t}^2}{N_p^2}\sigma_{j,s,p,t}^2] + E[\frac{1}{N_p}\sigma_\epsilon^2] \end{aligned} \tag{3.10}$$

where N_s is the number of students in school s associated with principal p and $N_{j,s,p,t}$ is the number of students in a given classroom denoted by (j, s, p, t).

The empirical Bayes shrinkage factor is then given by the total signal-to-noise ratio:

$$\lambda_p = \frac{\sigma_p^2 + \frac{\sum_{j\in p} N_j^2}{N_p^2}\sigma_j^2}{\sigma_p^2 + \frac{\sum_{j\in p} N_j^2}{N_p^2}\sigma_j^2 + \frac{\sum_{s\in p} N_s^2}{N_p^2}\sigma_s^2 + \frac{\sum_{j,s,p,t\in p} N_{j,s,p,t}^2}{N_p^2}\sigma_{j,s,p,t}^2 + \frac{1}{N_p}\sigma_\epsilon^2} \tag{3.11}$$

The shrinkage factor is then evaluated with the variance estimates reported in Table 1.2.

This section explores the relationship between general principal soft skills and TVA. I estimate individual TVA and examine whether the soft skills described in Section 6.3 explain the observed variation in TVA. Following the teacher literature, I estimate

$$y_{i,j,g,t} = \beta_0 + \sum_a \beta_a y_{i,t-1} \mathbf{I}\{\text{grade} = a\} + \theta_j + X'_{i,t}\delta + X'_{j,s,t}\tau + \alpha_t + \alpha_g + \epsilon_{i,j,g,t} \tag{3.12}$$

where TVA is the estimate of θ_j, the teacher fixed effect, and the controls are identical to those in equation 1.6.

I then use the estimates of θ_j to examine whether observable teacher characteristics and principal leadership predict TVA, where the intuition mirrors that of the approach in Section 5.1:

$$\hat{\theta}_j = \beta_0 + \tau X_j + \epsilon_j \tag{3.13}$$

X_j is a vector of time-invariant teacher controls such as the average experience and age that I observe for a given teacher, gender, race, indicators for whether a teacher graduated from a North Carolina institution and the average principal leadership, teacher empowerment, and overall school climate rating associated with a teacher. Since the school climate questions are not available until 2002, I replace these measures of principal soft skills with a 0 and include an indicator for missing. Table A.2 reports the regression results.

Table A.2: Relationship between Teacher Characteristics and Teacher Value-Added

	(1)	(2)	(3)
Experience	0.0071***	0.0071***	0.0069***
	(0.00041)	(0.00040)	(0.00041)
Experience squared	-0.00014***	-0.00014***	-0.00014***
	(0.000012)	(0.000012)	(0.000012)
Female	0.026***	0.025***	0.012***
	(0.0024)	(0.0024)	(0.0024)
White	0.026***	0.025***	0.021***
	(0.0027)	(0.0028)	(0.0029)
Top 10 University or Ivy	0.021	0.016	0.026*
	(0.015)	(0.015)	(0.015)
Aggregate School Leadership	0.048***	0.052***	0.043***
	(0.0060)	(0.0061)	(0.0065)
Overall School Climate	0.019***	0.017***	0.0078*
	(0.0038)	(0.0038)	(0.0041)
Aggregate Teacher Empowerment	-0.0089	-0.0077	-0.0078
	(0.0056)	(0.0056)	(0.0059)
Observations	63,062	63,062	63,062
Fixed Effects	None	District	School
Adjusted R^2	0.022	0.040	0.071
F	88.3	94.1	61.0

Notes: This table reports estimates of the association between TVA and observable teacher characteristics and school climate measures. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This section examines the determinants of principal mobility with a focus on general principal turnover, which combines job-to-job transitions with job separations, and job-to-job transitions in isolation. I estimate the following regression:

$$\mathbf{I}\{\text{exit}_{p,t}\} = \beta_0 + \sum_{\tau=1}^{3} \beta_{k,t-\tau} Y_{s,t-\tau} + \lambda \mathbf{X}_p + \gamma \mathbf{X}_{s,t} + \alpha_s + \alpha_t + \epsilon_{p,t} \tag{3.14}$$

where $\mathbf{I}\{\text{exit}_{p,t}\}$ is an indicator for general teacher turnover or a job transition and $Y_{s,t-\tau}$ is lagged (up to three years) average school test scores in both math and reading. $\mathbf{X}_p$ is a vector of principal characteristics such as experience, highest degree, institutional quality of degree-granting institution, etc., and $\mathbf{X}_{s,t}$ is a vector of school-level means of observable student characteristics (e.g., share of boys, eligibility for free or reduced-price lunch, academically gifted) and, in some specifications, average scores on school climate surveys (see Section 6.3 for more details). α_s and α_t are school and year fixed effects. The parameters of interest are the β_k's since principals are partly evaluated on school academic performance.

Table A.3 reports the regression estimates. Column 1 reports statistically significant estimates for the effect of lagged math and reading scores on general principal turnover; however, the two effects are of opposite sign. In general, there is weak evidence suggesting that prior school performance is predictive of general turnover or principal job transitions, assuaging concerns that the observed principal effects, and their potential positive impacts, are attributable to school “takeovers" in which a principal is specifically hired to raise student test scores and reform the overall school.

Table A.3: School Achievement and Principal Turnover

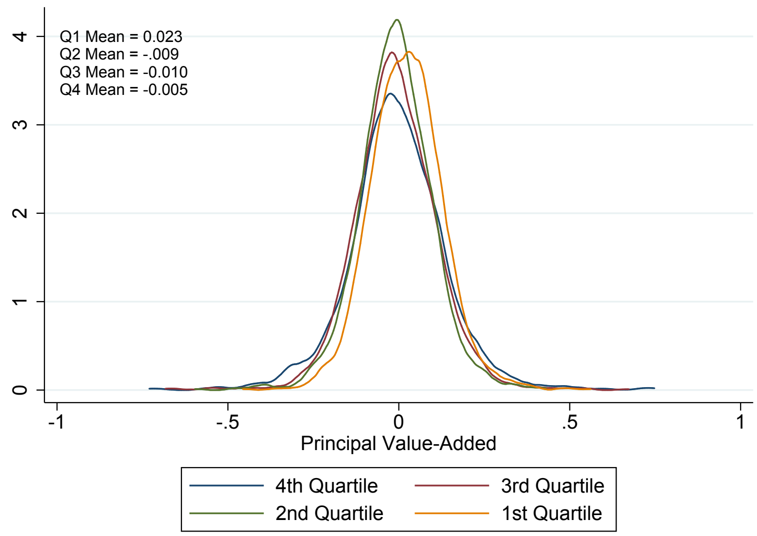
	(1) General Turnover	(2) General Turnover	(3) Job Transition	(4) Job Transition
Lag Math Scores	-0.0423**	-0.0168	0.00236	0.0110
	(0.0169)	(0.0251)	(0.0107)	(0.0157)
Lag Reading Scores	0.0344*	0.0361	0.0155	0.0140
	(0.0200)	(0.0303)	(0.0126)	(0.0191)
2-Year-Lag Math Scores	0.0108	0.0269	-0.0122	0.00240
	(0.0177)	(0.0276)	(0.0113)	(0.0176)
2-Year-Lag Reading Scores	-0.00444	-0.0522	0.000249	-0.0244
	(0.0209)	(0.0329)	(0.0132)	(0.0207)
3-Year-Lag Math Scores	-0.000683	-0.00720	0.00506	0.00147
	(0.0170)	(0.0266)	(0.0110)	(0.0171)
3-Year-Lag Reading Scores	0.0104	0.0128	0.00278	0.00117
	(0.0199)	(0.0324)	(0.0128)	(0.0206)
Principal Controls	Yes	Yes	Yes	Yes
Fixed Effects	School, Year	School, Year	School, Year	School, Year
School Climate Controls	None	Leadership Surveys	None	Leadership Surveys
F-stat	65.74	22.06	8.210	6.673
Observations	35,467	16,312	35,467	16,312

Notes: This table reports the determinants of principal turnover as well as job-to-job transitions. Standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

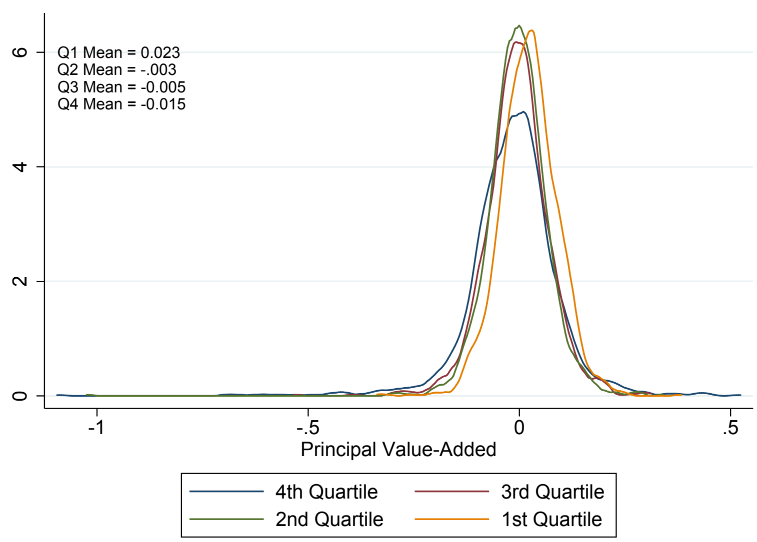
1.6 Additional Figures

Figure A.1: Distribution of Principal Value-Added by FRPL Status

(a) PVA Math



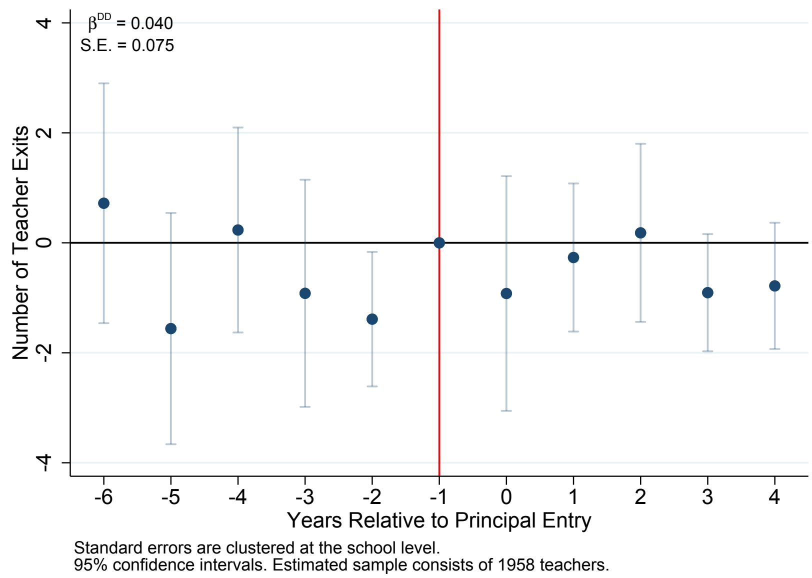
(b) PVA Reading



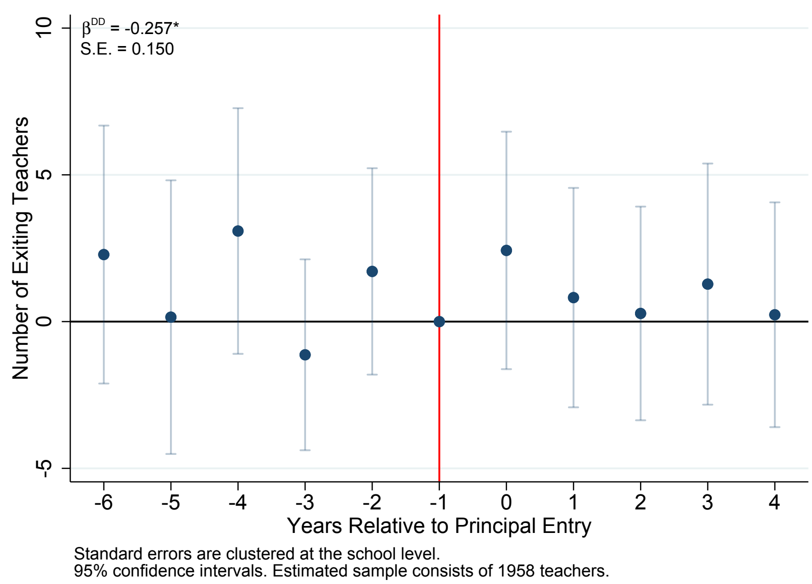
Notes: This figure plots the distribution of principal value-added for math and reading based on the share of students who are eligible for Free or Reduced Price Lunch status (FRPL) at the school-level. Schools are grouped into quartiles based on the average share of students eligible for FRPL across all school-years. The first quartile corresponds to the schools with the fewest share of FRPL while the fourth quartile corresponds to the schools with the largest share.

Figure A.2: Principal Quality and Number of Entering and Exiting Teachers

(a) Number of Entering Teachers



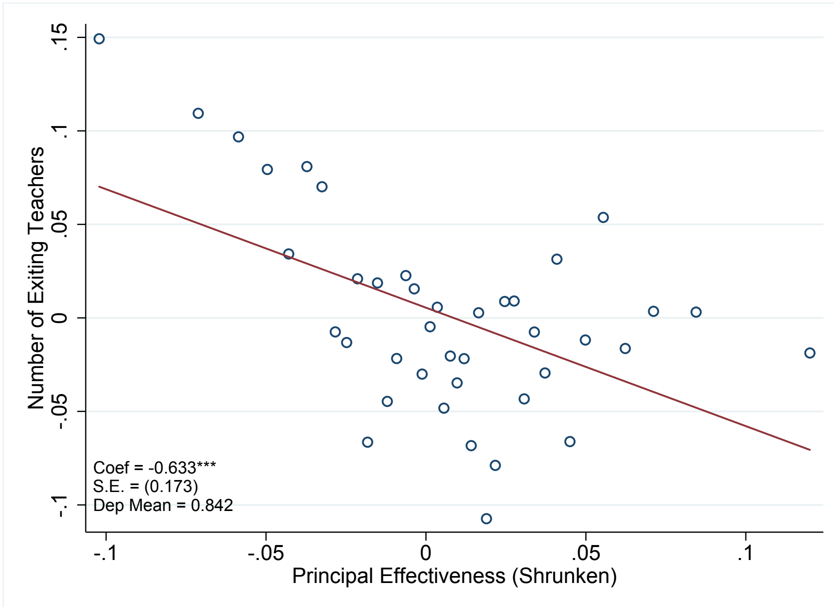
(b) Number of Exiting Teachers



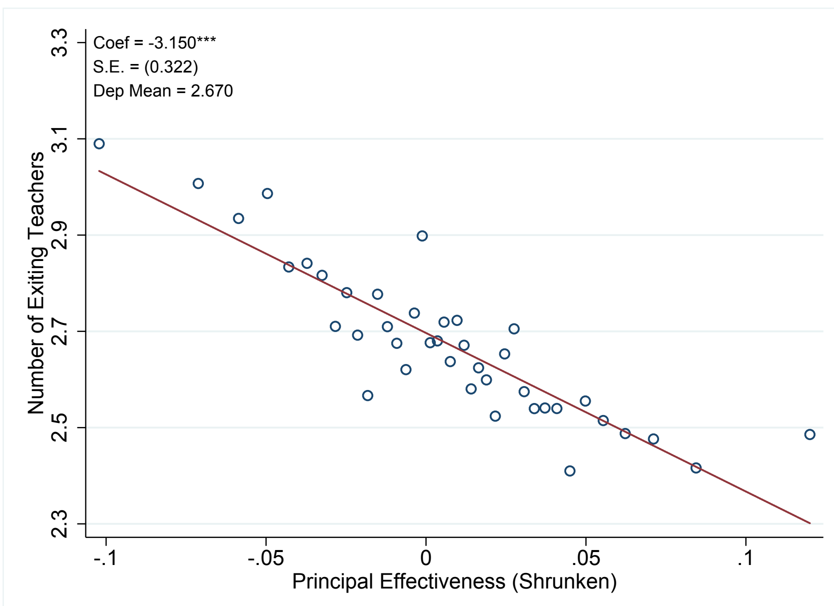
Notes: This figure plots the β coefficients from equation 1.12, which examines the effect of within-school changes in principal quality on the total of number of existing teachers entering and all teachers exiting. The analysis focuses on "events" where the exiting and arriving principals are observed for four years; in total, there are 245 such events. Reported coefficients correspond to the difference-in-differences analogue of equation 1.12. All standard errors are clustered at the principal level, and 95% confidence intervals are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.3: Principal Quality and Number of Entering and Exiting Teachers

(a) Number of Entering Teachers

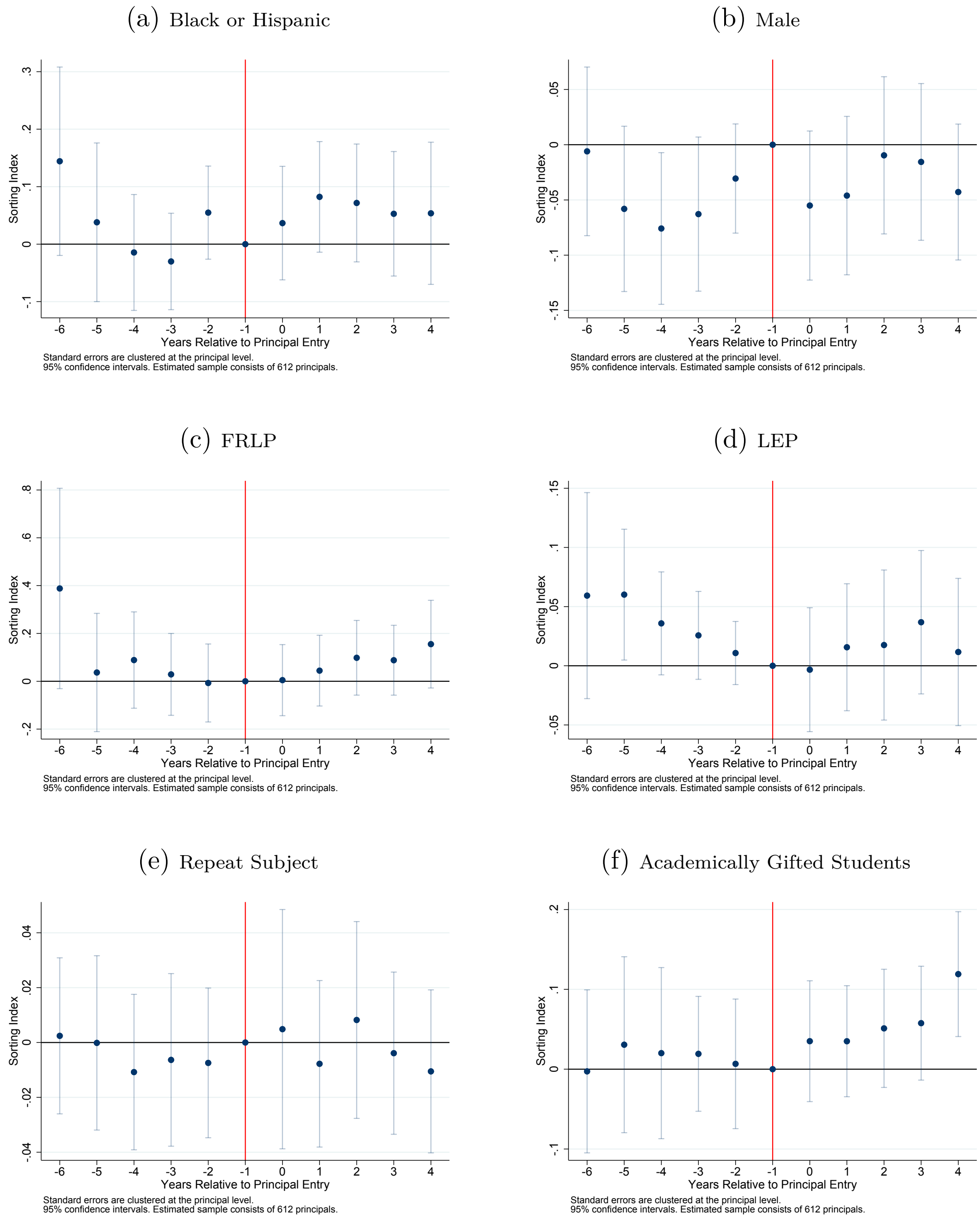


(b) Number of Exiting Teachers



Notes: This figure plots the estimates from a binned scatter plot regression using the full sample of principals: $y_{s,t} = \beta_0 + \beta_1 \hat{\theta}_p^{eb} + X'_{s,t}\phi_2 + \alpha_t + \alpha_s + \epsilon_{s,t}$ where $y_{s,t}$ is either the number of entering existing teachers or the total number of exiting teachers. $\hat{\theta}_p^{eb}$ are empirical Bayes shrunken estimates of PVA in math. $X'_{s,t}$ is a vector of school-level controls as defined in equation 1.12. The binned scatter plots are divided into 40 equally sized bins that correspond to the conditional expectation of the outcome variable given a particular value of shrunken PVA. The regression coefficient corresponds to the OLS regression on principal–year-level data. All standard errors are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

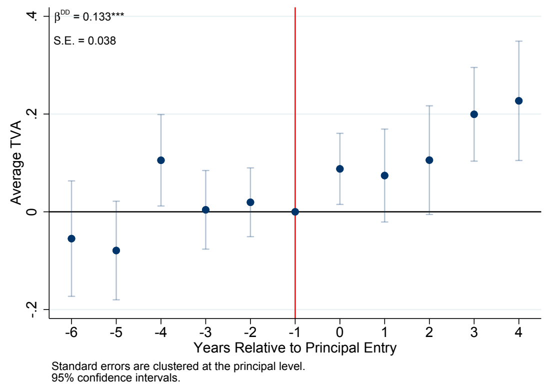
Figure A.4: Principal Quality and Student Composition



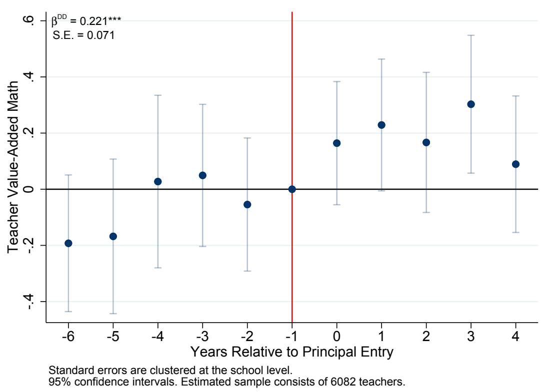
Notes: This figure plots the β coefficients from equation 1.8 where the outcome variables are the share of students who are Black or Hispanic, are male, are eligible for free or reduced-price lunch (FRLP), have limited English proficiency (LEP), have repeated a subject from a previous year, or are classified as academically gifted. Both entering and exiting principals are required to have been observed for at least four years. Standard errors are clustered at the principal level, and 95% confidence intervals are reported.

Figure A.5: Changes in Principal Quality on Teacher Composition

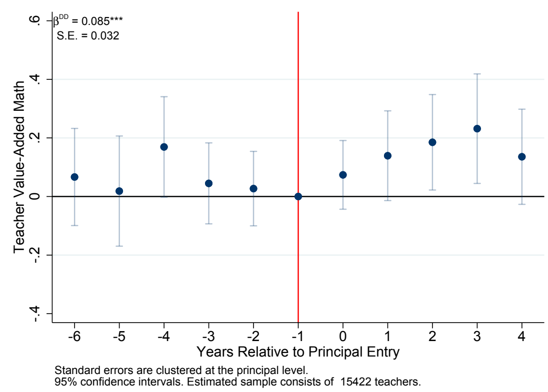
(a) Average Teacher Effectiveness



(b) Entering Teachers



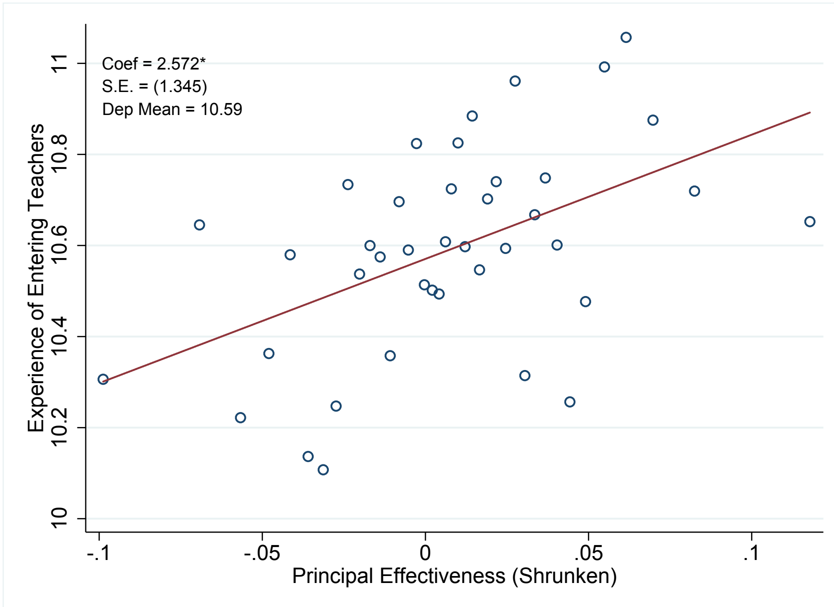
(c) Exiting Teachers



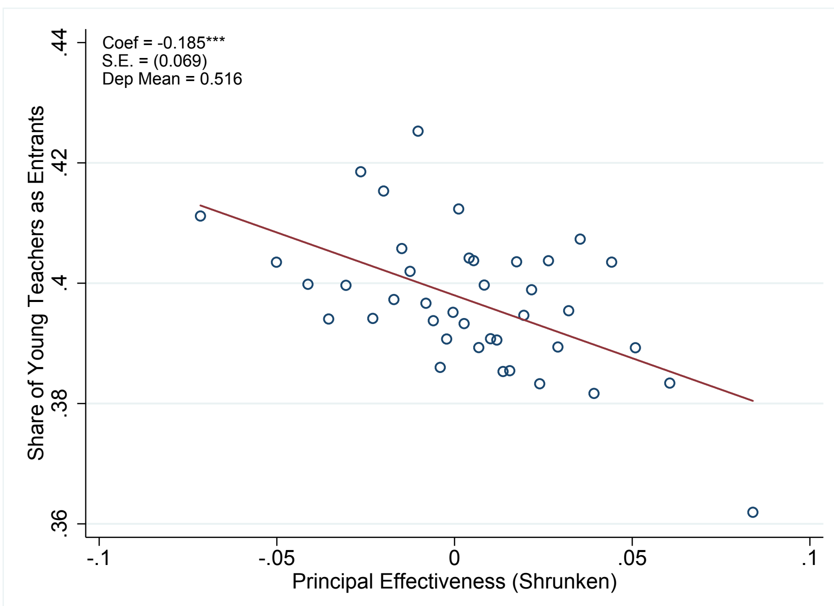
Notes: This figure plots the β coefficients from equation 1.12, which examines the effect of within-school changes to principal quality on average (school-level) and entering and exiting teacher value-added. No restrictions are placed on the number of observed years for the entering and exiting principal. Reported coefficients correspond to the difference-in-differences analogue of equation 1.12. All standard errors are clustered at the principal level, and 95% confidence intervals are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.6: Characteristics of Entering Teachers

(a) Experience of Entering Teachers

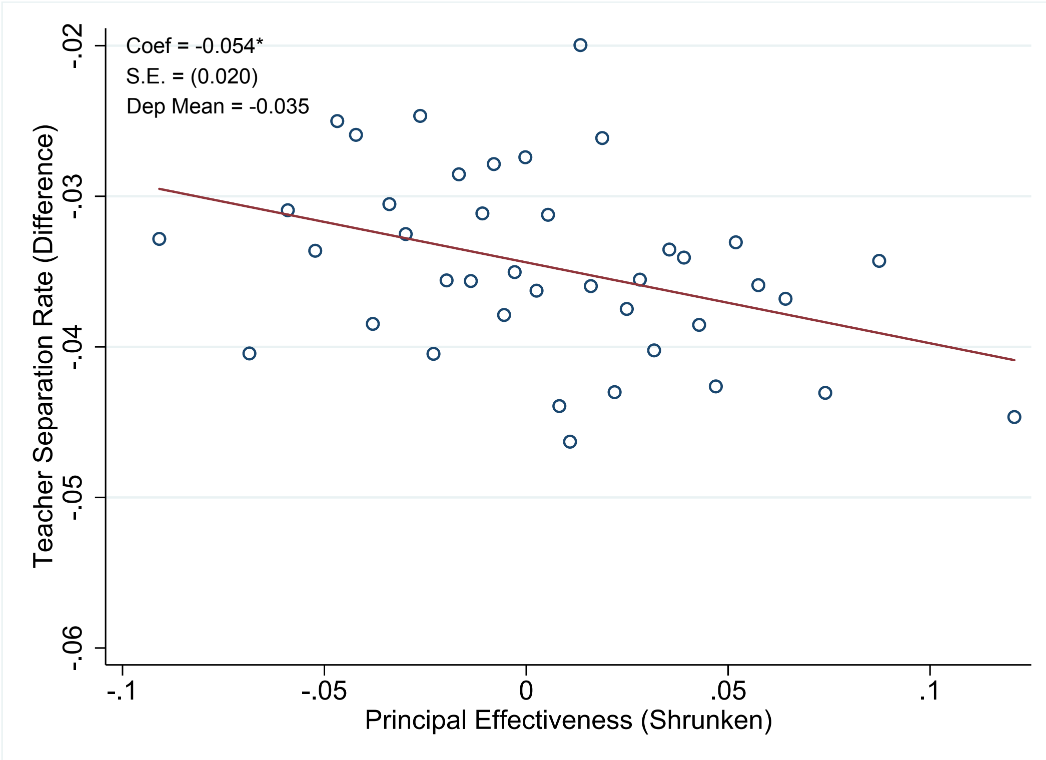


(b) Teachers with Less than 4 Years of Experience as Share of New Hires



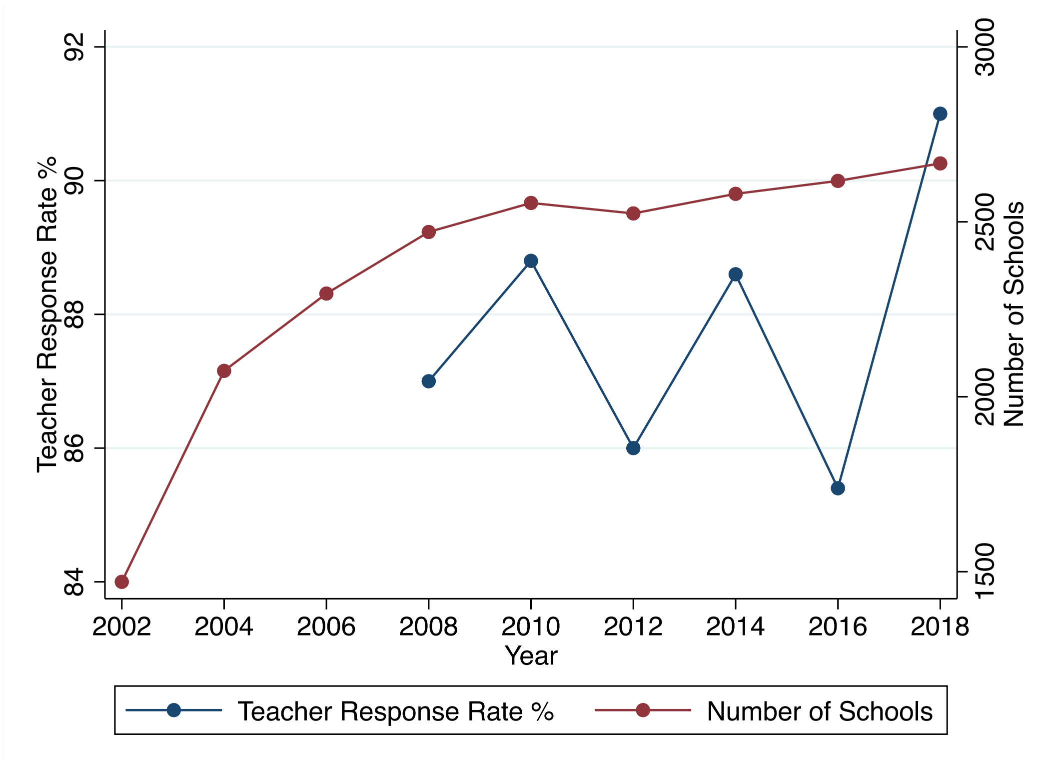
Notes: This figure shows the association between principal quality and characteristics of entering teachers. Panel A reports the results for the average experience level of entering teachers. Panel B reports the share of new entrants who have less than four years of experience. The binned scatter plots are divided into 40 equally sized bins that correspond to the conditional expectation of the outcome variable given a particular value of shrunken PVA. Both the dependent and independent variables are first residualzed against time-varying school means of student characteristics (described in equation 1.12), school fixed effects and year fixed effects. Math PVA and mean PVA are used in panels A and B, respectively. The coefficient and standard error correspond to the identical regression at the principal-year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.7: Principal Quality and Teacher Attrition



Notes: This figure plots the binned scatter plot of equation 1.13 where the outcome variable is the difference in attrition rate between teachers above and below median value-added in mathematics and the main dependent variable is mean PVA. The binned scatter plots are divided into 40 equally sized bins that correspond to the conditional expectation of the outcome variable given a particular value of shrunken PVA. Both the dependent and independent variables are first residualzed against time-varying school means of student characteristics (described in equation 1.12), school fixed effects and year fixed effects.The coefficient and standard error correspond to the identical regression at the principal-year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

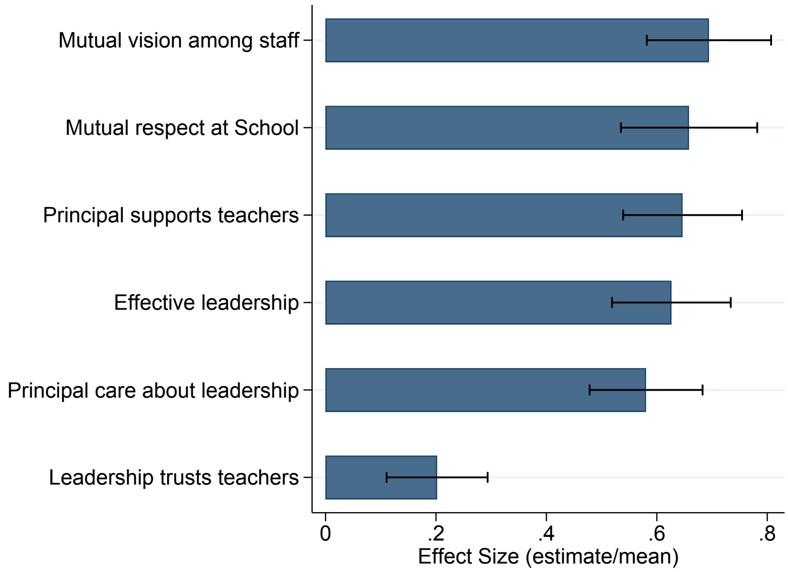
Figure A.8: WCS Survey Rates and School Responses



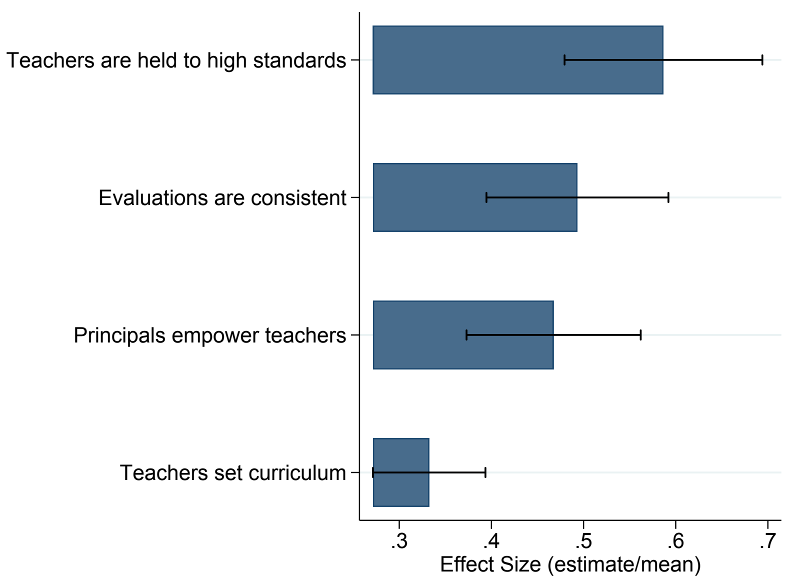
Notes: This figure plots the share of all teachers responding to the WCS survey questions over time as well as the the total number of schools participating - including non-public schools (e.g., charter) as well as high schools. Teacher response rates are provided by the NCERDC and are first disclosed in 2008.

Figure A.9: Detailed WCS Questions

(a) Teacher Empowerment



(b) School Leadership



Notes: This figure shows the size of the effect of principal quality on the responses to individual questions corresponding to teacher empowerment and school leadership. To obtain the effect size, I estimate an analogue of equation 1.13 in which I replace the outcome variable with the responses to the individual survey questions. I then scale the corresponding estimates by dividing by the outcome mean. As with equation 1.13, the outcome variable corresponds to the degree to which respondents agree with the given outcome variable.

Table A.4: Faculty and School Characteristics:
By School FRPL Status

	Q1 Mean	Q2 Mean	Q3 Mean	Q4 Mean
Principal Characteristics				
PVA Math	0.023	-0.009	-0.010	-0.005
PVA Read	0.023	-0.003	-0.005	-0.015
Annual Salary	68.871	66.689	64.447	62.925
Female	0.572	0.576	0.585	0.635
Age	48.090	47.863	46.883	47.477
Principal Experience	3.322	3.366	3.081	2.777
White	0.851	0.826	0.731	0.470
Class Size	32.030	35.166	35.799	31.665
Top 10 or Ivy	0.005	0.006	0.001	0.003
School Ranked in *USN&WR*	0.142	0.074	0.056	0.064
BA and Above	1.000	1.000	1.000	1.000
MA and Above	0.997	0.994	0.995	0.994
Doctorate	0.183	0.140	0.144	0.166
TVA Math	0.030	0.005	0.006	0.006
TVA Reading	0.026	0.007	0.007	0.006
School Characteristics				
Black or Hispanic	0.221	0.316	0.423	0.656
Limited English	0.034	0.049	0.068	0.084
Teacher Age	41.558	41.855	42.028	41.864
Teacher Experience	12.097	12.314	12.286	11.743
Female Teacher	0.123	0.123	0.111	0.123
Rookie Teacher	0.057	0.063	0.068	0.086
New Teacher	0.077	0.085	0.092	0.118
Turnover Rate	0.253	0.269	0.287	0.364
TVA Math (School)	0.024	-0.003	-0.011	-0.018
TVA Read (School)	0.029	0.005	-0.002	-0.013

Notes: This table reports summary statistics for schools separated by quartiles by the share of students eligible for Free or Reduced Price Lunch. Quartiles are calculated based on the average share of eligible students across all school-years. Principal (school) statistics reflect averages across all principal (school) years. Annual salary is indexed to 2019. School Ranked in *USN&WR* is an indicator for whether an individual's highest degree–granting institution was ranked in *US News & World Report*'s "Historical Liberal Arts College and University Rankings" as of 2023 (accessed via http://andyreiter.com/datasets/text). Rookie teacher indicates whether a newly hired teacher had never taught before.

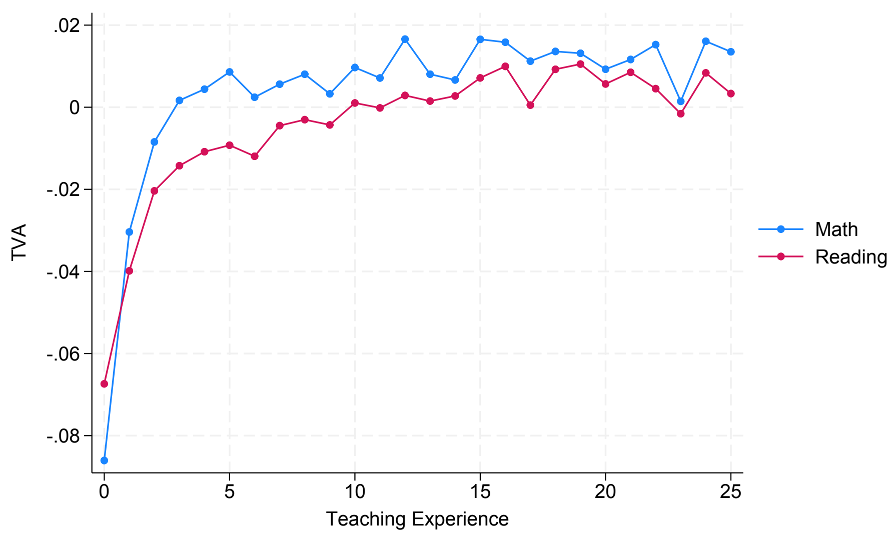
2. Appendix to "The Impact of Tenure Removal on Teacher's Labor Supply Responses"

2.1 Appendix: Career Status Description

The following excerpt is from a 2011 publication from the North Carolina Department of Public Instruction titled "Working in North Carolina's Public Schools" describing career status:

> *"Career status (tenure) provides job security by assuring that a tenured teacher cannot be dismissed except for proper cause. It also provides for due process before a tenured teacher is dismissed. Probationary teachers are certified individuals who have not obtained career status (tenure) and whose major responsibility is to teach or to supervise teaching. Administrators and teachers with emergency permits, alternative entry licenses, temporary permits, or lateral entry licenses are not classified as probationary teachers. During the term of their contract, probationary teachers have the same due process protections as career teachers. Probationary teachers who are employed by the same North Carolina public school system for four consecutive years are eligible for career status. Teachers who have obtained career status in any North Carolina public school system need not serve another probationary period of more than one year."*

Figure A.10: Teacher Value-Added by Experience



Notes: This figure plots the relationship between teacher value-added, separately for math and reading, and teacher experience. I begin by estimating a modified version of equation 2.1 using a time varying teacher fixed-effect $\theta_{j,t}$ and omitting experience dummies. I then collapse estimates of $\theta_{j,t}$ and plot the average $\hat{\theta}_{j,t}$ for each year of teacher experience.

Bibliography

Aaronson, Daniel, Lisa Barrow, and William Sander, "Teachers and Student Achievement in the Chicago Public High Schools," *Journal of Labor Economics*, 2007, *25* (1), 95–135.

Agostinelli, Francesco, Margaux Luflade, Paolo Martellini et al., "On the Spatial Determinants of Educational Access," Technical Report 2021.

Altonji, Joseph G and Richard K Mansfield, "Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects," *American Economic Review*, 2018, *108* (10), 2902–2946.

Anderson, Kaitlin P, Joshua M Cowen, and Katharine O Strunk, "The Impact of Teacher Labor Market Reforms on Student Achievement: Evidence from Michigan," *Education Finance and Policy*, 2022, *17* (4), 666–692.

Andrabi, Tahir, Natalie Bau, Jishnu Das, and Asim Ijaz Khwaja, "Heterogeneity in School Value-Added and the Private Premium," Technical Report, National Bureau of Economic Research 2022.

Angelici, Marta and Paola Profeta, "Smart Working: Work Flexibility Without Constraints," *Management Science*, 2024, *70* (3), 1680–1705.

Angrist, Joshua D, Peter D Hull, Parag A Pathak, and Christopher R Walters, "Leveraging Lotteries for School Value-Added: Testing and Estimation," *The Quarterly Journal of Economics*, 2017, *132* (2), 871–919.

Angrist, Joshua, Peter Hull, Parag A Pathak, and Christopher Walters, "Credible School Value-Added With Undersubscribed School Lotteries," *Review of Economics and Statistics*, 2021, pp. 1–46.

Araujo, M Caridad, Pedro Carneiro, Yyannú Cruz-Aguayo, and Norbert Schady, "Teacher Quality and Learning Outcomes in Kindergarten," *The Quarterly Journal of Economics*, 2016, *131* (3), 1415–1453.

Autor, David H, William R Kerr, and Adriana D Kugler, "Does Employment Protection Reduce Productivity? Evidence from US States," *The Economic Journal*, 2007, *117* (521), F189–F217.

Ba, Bocar, Patrick Bayer, Nayoung Rim, Roman Rivera, and Modibo Sidibé, "Police Officer Assignment and Neighborhood Crime," Technical Report, National Bureau of Economic Research 2021.

Barbanchon, Thomas Le, Roland Rathelot, and Alexandra Roulet, "Gender Differences in Job Search: Trading Off Commute Against Wage," *The Quarterly Journal of Economics*, 2021, *136* (1), 381–426.

Bartanen, Brendan, "Principal Quality and Student Attendance," *Educational Researcher*, 2020, *49* (2), 101–113.

_, Aliza N Husain, and David D Liebowitz, "Rethinking Principal Effects on Student Outcomes," Technical Report, Ed Working Paper: 22-621. Retrieved from Annenberg Institute at Brown ... 2022.

Bates, Michael D, Michael Dinerstein, Andrew C Johnston, and Isaac Sorkin, "Teacher Labor Market Equilibrium and Student Achievement," Technical Report, National Bureau of Economic Research 2022.

Bau, Natalie and Jishnu Das, "Teacher Value Added in a Low-Income Country," *American Economic Journal: Economic Policy*, 2020, *12* (1), 62–96.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan, "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, 2007, *115* (4), 588–638.

Beaman, Lori, Raghabendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova, "Powerful Women: Does Exposure Reduce Bias?," *The Quarterly Journal of Economics*, 2009, *124* (4), 1497–1540.

Benson, Alan, Danielle Li, and Kelly Shue, "Promotions and the Peter Principle," *The Quarterly Journal of Economics*, 2019, *134* (4), 2085–2134.

Besley, Timothy and Robin Burgess, "Can Labor Regulation Hinder Economic Performance? Evidence from India," *The Quarterly Journal of Economics*, 2004, *119* (1), 91–134.

Biasi, Barbara, "The Labor Market for Teachers Under Different Pay Schemes," *American Economic Journal: Economic Policy*, 2021, *13* (3), 63–102.

_ , Chao Fu, and John Stromme, "Equilibrium in the Market for Public School Teachers: District Wage Strategies and Teacher Comparative Advantage," Technical Report, National Bureau of Economic Research 2021.

Bjuggren, Carl Magnus, "Employment Protection and Labor Productivity," *Journal of Public Economics*, 2018, *157*, 138–157.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen, "Does Management Matter in Schools?," *The Economic Journal*, 2015, *125* (584), 647–674.

Bobba, Matteo, Tim Ederer, Gianmarco Leon-Ciliotta, Christopher Neilson, and Marco G Nieddu, "Teacher Compensation and Structural Inequality: Evidence From Centralized Teacher School Choice in Perú," Technical Report, National Bureau of Economic Research 2021.

Boboshko, Nikolai, "The Effect of Tenure Laws on Students: Evidence from the Implementation of Tenure Systems in the 20th Century," 2021.

Branch, Gregory F, Eric A Hanushek, and Steven G Rivkin, "Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals," Technical Report, National Bureau of Economic Research 2012.

Brown, Christina and Tahir Andrabi, "Inducing Positive Sorting Through Performance Pay: Experimental Evidence from Pakistani Schools," *University of California at Berkeley Working Paper*, 2020.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline, "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 2018, *36* (S1), S13–S70.

Carruthers, Celeste, David Figlio, and Tim Sass, "Did Tenure Reform in Florida Affect Student Test Scores," *Evidence Speaks Reports*, 2018, *2* (52), 1–14.

Chetty, Raj, John N Friedman, and Jonah E Rockoff, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 2014, *104* (9), 2593–2632.

_ , _ , and _ , "Measuring the Impacts of Teachers II: Teacher Value-added and Student Outcomes in Adulthood," *American Economic Review*, 2014, *104* (9), 2633–79.

Chiang, Hanley, Stephen Lipscomb, and Brian Gill, "Is School Value Added Indicative of Principal Quality?," *Education Finance and Policy*, 2016, *11* (3), 283–309.

Das, Jishnu, Alaka Holla, Aakash Mohpal, and Karthik Muralidharan, "Quality and Accountability in Health Care Delivery: Audit-Study Evidence From Primary Care in India," *American Economic Review*, 2016, *106* (12), 3745–3777.

Dee, Thomas S, "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?," *American Economic Review*, 2005, *95* (2), 158–165.

Dhuey, Elizabeth and Justin Smith, "How Important Are School Principals in the Production of Student Achievement?," *Canadian Journal of Economics/Revue canadienne d'économie*, 2014, *47* (2), 634–663.

_ **and** _ , "How School Principals Influence Student Learning," *Empirical Economics*, 2018, *54* (2), 851–882.

Dinerstein, Michael and Isaac M Opper, "Screening with Multitasking: Theory and Empirical Evidence from Teacher Tenure Reform," Technical Report, National Bureau of Economic Research 2022.

Fenizia, Alessandra, "Managers and Productivity in the Public Sector," *Econometrica*, 2022, *90* (3), 1063–1084.

Gershenson, Seth, Cassandra MD Hart, Joshua Hyman, Constance A Lindsay, and Nicholas W Papageorge, "The Long-Run Impacts of Same-Race Teachers," *American Economic Journal: Economic Policy*, 2022, *14* (4), 300–342.

Goldhaber, Dan, Michael Hansen, and Joe Walch, "Time to Tenure: Does Tenure Reform Affect Teacher Absence Behavior and Mobility? Working Paper 172.," *National Center for Analysis of Longitudinal Data in Education Research (CALDER)*, 2016.

Grissom, Jason A, Anna J Egalite, Constance A Lindsay et al., "How Principals Affect Students and Schools," *Wallace Foundation*, 2021, *2* (1), 30–41.

_ , Demetra Kalogrides, and Susanna Loeb, "Using Student Test Scores to Measure Principal Performance," *Educational Evaluation and Policy Analysis*, 2015, *37* (1), 3–28.

Harris, Douglas N and Tim R Sass, "Value-added Models and the Measurement of Teacher Quality," *Unpublished Manuscript*, 2006.

Hinrichs, Peter, "What Kind of Teachers Are Schools Looking For? Evidence From a Randomized Field Experiment," *Journal of Economic Behavior & Organization*, 2021, *186*, 395–411.

Jackson, C Kirabo, "Student Demographics, Teacher Sorting, and Teacher Quality: Evidence from the End of School Desegregation," *Journal of Labor Economics*, 2009, *27* (2), 213–256.

_ , "Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence From Teachers," *Review of Economics and Statistics*, 2013, *95* (4), 1096–1116.

_ , "What Do Test Scores Miss? The Importance of Teacher Effects on Non–test Score Outcomes," *Journal of Political Economy*, 2018, *126* (5), 2072–2107.

_ , Shanette C Porter, John Q Easton, Alyssa Blanchard, and Sebastián Kiguel, "School Effects on Socioemotional Development, School-Based Arrests, and Educational Attainment," *American Economic Review: Insights*, 2020, *2* (4), 491–508.

_ , _ , _ , and Sebastián Kiguel, "Who Benefits From Attending Effective High Schools?," 2022.

Jacob, Brian A, "The Effect of Employment Protection on Teacher Effort," *Journal of Labor Economics*, 2013, *31* (4), 727–761.

Jacob, Brian and Lars Lefgren, "Principals as Agents: Subjective Performance Measurement in Education," 2005.

Jr, Roland G Fryer and Steven D Levitt, "Understanding the Black-White Test Score Gap in the First Two Years of School," *Review of Economics and Statistics*, 2004, *86* (2), 447–464.

Kane, Thomas J and Douglas O Staiger, "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," Technical Report, National Bureau of Economic Research 2008.

_ , Jonah E Rockoff, and Douglas O Staiger, "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City," *Economics of Education Review*, 2008, *27* (6), 615–631.

Kaur, Supreet, Sendhil Mullainathan, Suanna Oh, and Frank Schilbach, "Do Financial Concerns Make Workers Less Productive?," Technical Report, National Bureau of Economic Research 2021.

Klein, Joel and Others, "How to Fix Our Schools: A Manifesto by Joel Klein, Michelle Rhee and Other Education Leaders," *The Washington Post*, 2010, *10.*

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten, "Leave-out Estimation of Variance Components," *Econometrica*, 2020, *88* (5), 1859–1898.

Kraft, Matthew A, "Teacher Layoffs, Teacher Quality, and Student Achievement: Evidence from a Discretionary Layoff Policy," *Education Finance and Policy*, 2015, *10* (4), 467–507.

Krueger, Alan B and Lawrence H Summers, "Efficiency Wages and the Inter-industry Wage Structure," *Econometrica: Journal of the Econometric Society*, 1988, pp. 259–293.

Lazear, Edward P, “Teacher Incentives,” *Swedish Economic Policy Review*, 2003, *10* (2), 179–214.

Lemos, Renata, Karthik Muralidharan, and Daniela Scur, “Personnel Management and School Productivity: Evidence From India,” Technical Report, National Bureau of Economic Research 2021.

Loeb, Susanna, Luke C Miller, and James Wyckoff, “Performance Screens for School Improvement: The Case of Teacher Tenure Reform in New York City,” *Educational Researcher*, 2015, *44* (4), 199–212.

Maestas, Nicole, Kathleen J Mullen, David Powell, Till Von Wachter, and Jeffrey B Wenger, “The Value of Working Conditions in the United States and Implications for the Structure of Wages,” *American Economic Review*, 2023, *113* (7), 2007–2047.

Mas, Alexandre and Amanda Pallais, “Valuing Alternative Work Arrangements,” *American Economic Review*, 2017, *107* (12), 3722–3759.

_ **and** _ , “Alternative Work Arrangements,” *Annual Review of Economics*, 2020, *12*, 631–658.

McCaffrey, Daniel F, Tim R Sass, JR Lockwood, and Kata Mihaly, “The Intertemporal Variability of Teacher Effect Estimates,” *Education Finance and Policy*, 2009, *4* (4), 572–606.

Mulhern, Christine, “Beyond Teachers: Estimating Individual Guidance Counselors’ Effects on Educational Attainment,” *Unpublished Manuscript, RAND Corporation*, 2020.

NCES, “"Numbers and Types of Public Elementary and Secondary Schools From the Common Core of Data: School Year 2009-10",” 2010.

_ , “Characteristics of Public School Teachers,” 2020.

— , "Fast Facts: Expenditures," 2020.

Ng, Kevin, "The Effects of Teacher Tenure on Productivity and Selection," Technical Report, Working Paper 2021.

Peter, Laurence J, Raymond Hull et al., *The Peter Principle*, Vol. 4, Souvenir Press London, 1969.

Porter, Catherine and Danila Serra, "Gender Differences in the Choice of Major: The Importance of Female Role Models," *American Economic Journal: Applied Economics*, 2020, *12* (3), 226–254.

Rockoff, Jonah E, "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *American Economic Review*, 2004, *94* (2), 247–252.

— **, Brian A Jacob, Thomas J Kane, and Douglas O Staiger**, "Can You Recognize an Effective Teacher When You Recruit One?," *Education Finance and Policy*, 2011, *6* (1), 43–74.

Rose, Evan K, Jonathan T Schellenberg, and Yotam Shem-Tov, "The Effects of Teacher Quality on Adult Criminal Justice Contact," Technical Report, National Bureau of Economic Research 2022.

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

— , "Teacher Quality Policy When Supply Matters," *American Economic Review*, 2015, *105* (1), 100–130.

— , "Measuring the Impacts of Teachers: Comment," *American Economic Review*, 2017, *107* (6), 1656–1684.

Wiswall, Matthew, "The Dynamics of Teacher Quality," *Journal of Public Economics*, 2013, *100*, 61–78.

_ **and Basit Zafar**, "Determinants of College Major Choice: Identification Using an Information Experiment," *The Review of Economic Studies*, 2015, *82* (2), 791–824.

Xia, Xiaoyu, "Forming Wage Expectations Through Learning: Evidence From College Major Choices," *Journal of Economic Behavior & Organization*, 2016, *132*, 176–196.