Principal Quality and Student Outcomes: Evidence from North Carolina

Calvin Kuo[†]

October 29, 2023

[Click here for the latest version]

Abstract

Principals shape almost every aspect of schools, assigning students to classrooms and matching teachers and students with resources. This paper quantifies the impact of principal quality and documents the correlates of principal effectiveness. 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.

[†]I am grateful to my advisors Martha Bailey, Natalie Bau, Felipe Gonçalves, and Yotam Shem-Tov for their guidance and support throughout the development of this project. I've also benefited from conversations with Laura Giuliano, Daniel Haanwinckel, Meredith Phillips, Jesper Böjeryd, Domenico Fabrizi, Johnny Huynh, Manu Navjeevan, and Huihuang Zhu. I thank Kara Bonneau from the North Carolina Education Research Data Center for her help in accessing the data.
Department of Economics, UCLA. Contact email: ckuo0127@gmail.com

1 Introduction

The United States spends approximately \$870 billion dollars 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, 2020). 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) and provides the first evidence showing the estimates' validity, and documents the correlates of principal effectiveness. I use administrative classroom-level 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, to build my analysis. Unique to my setting, the data also contain survey records asking faculty members 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 principal effectiveness.

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. I follow McCaffrey et al. (2009) and use the movement of over 2,000 principal transitions across 1,900 schools to quantify principal effects and properly account for sampling error. I find that a one-standard-deviation (SD) increase in principal quality raises average student test scores by 0.047 SD. These estimates have important long-run implications. Using data from Chetty et al. (2014b), my estimates suggest that a one-SD increase in principal quality during middle school raises lifetime earnings by \$45,825 for one student. Considering that one middle school 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 (Chetty et al., 2014b).

Having quantified the extent to which principals matter for student academic performance, I then examine *who* effective principals are and what makes them successful. I calculate individual PVA and provide the first set of results validating these estimates. I begin by following 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 first show that future principal quality is not predictive of prior student achievement, 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 and reflects variation in principal effectiveness.

As a second test, I follow a procedure similar to that in Chetty et al. (2014a) and use principal entry and exit within schools to examine whether changes in principal quality predict changes in mean test scores. I show, consistent with the previous test, that PVA is forecast unbiased, as the coefficient on test score changes is not statistically different from one. Moreover, estimates from a placebo test examining whether changes in future principal quality predict lag score changes are statistically indistinguishable from zero, assuaging concerns that changes in value-added reflect students' sorting on academic gains.

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 explain very little variation in PVA, as less than 4% of the within-district variation in principal effectiveness is explained by experience, education quality, highest degree obtained, and other covariates. Interestingly, conditional on having taught in North Carolina, 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 strongest teachers. A 1-SD increase in principal effectiveness reduces the job separation rate of teachers with above-median value-added by 6.5% relative to that of teachers below the median. My results contribute to a literature that has found mixed evidence on whether principals can identify effective teachers (Hinrichs, 2021; Bates et al., 2022). My analysis is consistent with the results of Jacob and Lefgren (2005) suggesting that principals can identify effective teachers, with greater scope for such positive selection among high-value-added principals.

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 school 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 ones 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.

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

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

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 discusses estimating PVA, provides validation checks, and examines the relationship between value-added and observable principal characteristics and wages. Section 5 examines the mechanisms through which principal effects operate, including their relationship to leadership skills. Section 6 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, allowing me to examine how principals affect aspects of school climate and potential returns to effective leadership[2].

[2] Section 5.3 describes the survey data in more detail.

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 (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 provide summary statistics for my analytic sample. The analysis 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, we observe that the typical school undergoes nearly five principal transitions, consistent with the data period spanning 25 years and the typical principal serving approximately four years. Regarding student demographics, 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:

$$Y_{i,s,j,p,t} = \theta^j + \theta^s + \theta^p + \theta^{j,s,p,t} + X'_{i,t}\tau + \epsilon_{i,s,j,p,t} \tag{1}$$

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 does not account for complementarities between each of the causal θ parameters, ruling out potential match effects between teachers and principals. Equation 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:

$$Y_{i,s,j,p,t} = \sum D_{i,t}\delta_{j,s,p,t} + X'_{i,t}\tau + \mu_{i,s,j,p,t} \tag{2}$$

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. 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^2_j, σ^2_s, σ^2_p, and $\sigma^2_{j,s,p,t}$ 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.

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

[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.

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 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 teachers 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, which I control for (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 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 Assumption 1, the covariance can be rewritten as

$$\begin{aligned} cov(\delta_{j,s,p,t}, \delta_{j',s',p,t'}) &= cov(\theta_j + \theta_s + \theta_p + \theta_{j,s,p,t}, \theta'_j + \theta'_s + \theta_p + \theta_{j',s',p,t'}) && (3) \\ &= \sigma^2_{\theta_p} + cov(\theta_p, \theta_{j'}) + cov(\theta_p, \theta_{s'}) + \dots && (4) \\ &= \sigma^2_{\theta_p} && (5) \end{aligned}$$

where the last line follows by applying Assumption 2. 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 $cov(\theta_j + \theta_s + \theta_p +$

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

$\theta_{j,s,p,t}, \theta_{j'}+\theta_s+\theta_p+\theta_{j',s',p,t'}) = \sigma^2_{\theta_s}+\sigma^2_{\theta_p}$. 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].

While the conditional independence and uncorrelated effects assumptions provide identification of the variances of interest, they may be too strong. To address this concern, Appendix A2 provides an alternative decomposition that isolates the variance of principal effects in the spirit of Kline et al. (2020) and Rose et al. (2022). Under the assumption that cohort or school shocks are uncorrelated over time, this approach yields nearly identical estimates of the variance of principal effects. Additionally, since identification of $\sigma^2_{\theta_p}$ rests on principals who work across multiple schools, a natural question is what determines principal mobility. Appendix Section A5, 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 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 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 they do not separate out the impact of principals from that of schools.

[6] See Appendix A1 for full details.

While principals' effects on individual students are modest, principal quality remains economically important since they oversee entire schools[7]. 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[8]. 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[9]. 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 findings as well as results from prior work on school effects. To quantify principal effects, these papers estimate a regression similar to equation 6 but also include school fixed effects, and correct for sampling error using the estimator proposed by Krueger and Summers (1988)[10]. 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 address this issue, Rose et al. (2022) extend Kline et al. (2020) and show that the leave-one-out procedure allows one to estimate the variance of interest without estimating the sampling error and potentially misspecifying the estimator.

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[11]. With this approach, a one-SD increase in prin-

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

[8]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)

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

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

[11]Specifically, I first estimate $y_{i,t} = \sum_a \beta_a y_{i,t-1} \mathbf{I}\{\mathrm{g} = a\} + X'_{i,t}\tau + X'_{s,t}\tau + \theta_p + \theta_s + \epsilon_{i,t}$. 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.

cipal 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 NCERDC dataset[12]. The large difference in estimate sizes highlights the challenge of adjusting for sampling error in high-dimensional settings.

Finally, since prior work examines the role of school quality in student outcomes without considering the role of principals, I assess the degree to which existing school effects reflect differences in principal quality. To this end, I omit θ_p from equation 1 and then estimate its population analogue. Using a similar variance decomposition that compares the covariance of various classroom effects, 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 2, which suggests that over one-third of the observed variation in school effects is driven by differences in principal quality[13]. This finding may have large 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.

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 4.3 documents which observable characteristics are predictive of principal effectiveness. Finally, Section 4.4 details whether principal quality is correlated with wages.

The second term under the square root is similar to the sampling correction term proposed by Krueger and Summers (1988).

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

[13] 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.

4.1 Estimating Principal Value-Added

To estimate PVA, I follow the teacher 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:

$$y_{i,g,p,t} = \sum_a \beta_a y_{i,t-1} \mathbf{I}\{\text{grade} = a\} + X'_{i,t}\delta + X'_{j,s,t}\tau + \alpha_t + \alpha_g + \theta_p + \epsilon_{i,g,p,t} \tag{6}$$

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 2, and their corresponding school- and classroom-level means[14]. 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 value-added are unbiased, $\hat{\theta}_p$ still contains sampling error. This implies that if PVA is an explanatory variable, estimation error will attenuate its coefficient. To address this, I construct empirical Bayes estimates of value-added using the variances from Table 2, with full details described in Appendix A3. When value-added is the outcome variable, I use the principal fixed effects from estimating equation 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

[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).

[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

Having estimated PVA, I now evaluate whether these estimates are 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.

4.2.1 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 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 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 6. After calculating these residuals, I obtain fitted values from a regression of the residualized $y_{i,t}$ on the residualized $y_{i,t-2}$. Fi-

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

nally, 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 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.

4.2.2 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:

$$y_{i,s,t} = \alpha_0 + \sum_{\tau \neq -1} \beta_\tau \hat{\theta}_p^{EB} + \mathbf{X}_{i,s,t} + \alpha_g + \alpha_t + \alpha_s + \epsilon_{i,s,t} \tag{7}$$

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 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 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 $\tau = -2$.

Figure 3 provides a visual representation of these empirical tests. The coefficient at $t = 0$ is indistinguishable from 1, as the p-value under a test of unity is approximately 0.11. While limited in preperiods, Figure 3 indicates that future principal quality is uncorrelated with student performance: the coefficient for the period prior to the move is indistinguishable from zero, which strengthens the evidence for the underlying identification assumption. Estimates from this regression are reported in 3. Column 1 estimates the most parsimonious version of equation 7 by omitting, with the exception of lagged test scores, all the controls contained in $X_{i,s,t}$ and the school fixed effects. Column 2 includes the school fixed effects. Column 3 uses the full set of reported controls. Consistent with prior work on school effects, the estimates in Column 1 are larger and statistically different from one, emphasizing that school quality is an important determinant of student achievement. Unsurprisingly, the estimates for β_0 in Columns 2 and 3 become smaller and are not statistically different from 1 since differences in school quality and student sorting patterns are accounted for.

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:

$$\Delta Y_{s,t} = \beta_0 + \beta_1 \Delta Q_{s,t} + \alpha_s + \alpha_t + \epsilon_{s,t} \tag{8}$$

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 4 report the regression results from equation 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 A3 presents event-study estimates using the analogue of equation 8 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.

4.3 Correlates of Principal Value-Added

Having established that my PVA estimates are unbiased, I now examine whether observable characteristics can predict principal quality. 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 even if these associations are not causal.

Table 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, and were more likely to have been effective teachers[17]. 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 6:

$$\hat{\theta}_p = \beta_0 + \tau X_p + \hat{\theta}^{eb}_{p(j)} + \epsilon_p \tag{9}$$

Where X_p is a vector of time-invariant principal characteristics such as gender, race, highest degree obtained, and indicators for whether the school is 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].

To account for other relevant but 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.

Table 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 endogenous sorting. 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" (Benson et al., 2019) and mitigates concerns that promoting the strongest teachers dampens student learning since the pool of teachers becomes weaker. Additionally, this result stands in contrast with existing literature assessing teacher quality. A well-documented phenomenon in the education literature[19] is the difficulty of using observable characteristics to predict teacher quality; 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 stan-

[17] This refers to the subsample of principals who were previously observed teaching in the North Carolina public school system.

[18] TVA is calculated similarly to PVA but replaces the principal fixed effects with teacher fixed effects. The shrinkage parameter for the teacher fixed effects follows from Kane and Staiger (2008), Jackson (2018), and Bau and Das (2020).

[19] See Rockoff (2004), Rockoff et al. (2011), and Rothstein (2015) for more details.

dard deviation than their less experienced counterparts. This role of experience is 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 6 shed light on what attributes school district officials could value, as prior teaching experience is strongly correlated with future principal quality.

4.4 Compensation and Sorting

This subsection explores the correlation between principal compensation and value-added. Table 5 suggests that the labor market rewards more effective principals; however, differences in principal compensation might reflect differences in bonus salary compensation across school districts, with more affluent districts typically offering larger bonuses to offset higher living costs. To test whether more effective principals are rewarded with higher salaries, I regress log salaries on principal characteristics:

$$log(salary)_p = \beta_0 + \hat{\theta}_p^{eb} + \tau \mathbf{X}_p + \alpha_d + \epsilon_p \tag{10}$$

where $log(salary)_p$ is the log of the mean salary of principal p and $\mathbf{X}_p$ is the set of observable characteristics in equation 9. α_d is a district fixed effect and, in some specifications, a school fixed effect. Table 7 reports the estimation results.

Consistent 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. These results are consistent with

[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.

the fact that public-sector salaries are experience based, with little scope for rewarding of performance.

While these results are less interesting in isolation, combined with those in Column 4, they shed light on potential principal sorting patterns. Without district or school fixed effects, a one-SD increase in principal quality is associated with 17% higher average principal compensation. As this estimate becomes statistically insignificant after I include district fixed effects, Column 4 suggests that more effective principals sort into, or are potentially allocated to, less distressed areas. This finding is consistent with the results of Ba et al. (2021), who find that more effective police officers tend to work in low-crime areas, and suggests potential equity gains from incentivizing stronger principals' working in more disadvantaged areas.

5 Mechanisms

In this section, I explore what practices are espoused by effective principals and how they shape student outcomes. Having constructed individual-level measures of principal effectiveness, I then examine how within-school changes in principal quality affect student outcomes and whether effective principals possess certain skills that differentiate them from their peers.

Panels B and C of Table 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.

5.1 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:

$$\overline{TVA}_{s,t} = \beta_0 + \sum_{\tau \neq -1} \beta_\tau \Delta P + X'_{s,t}\lambda + \alpha_s + \alpha_t + \epsilon_{s,t} \tag{11}$$

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 6 by replacing the principal fixed effects with a teacher fixed effects[22].

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 is

[22]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.

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 A4 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 11 focusing on how changes in principal quality affect the average TVA in mathematics[23]:

$$\begin{aligned} \overline{TVA}_{s,t} &= \beta_0 + \sum_{\tau \neq -1} \beta_\tau \widehat{\Delta P} + X'_{s,t}\lambda + \alpha_s + \alpha_t + \epsilon_{s,t} \\ \text{where } \widehat{\Delta P} &= \hat{\theta}^{eb,-s}_{p,\text{incoming}} - \hat{\theta}^{eb}_{p,\text{outgoing}} \end{aligned} \tag{12}$$

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 4a plots the event-study coefficients. Schools gaining more effective principals see improvements in average teacher quality. Differences-in-differences estimates of 12 indicate that a one-SD increase in principal quality raises average teacher quality by 0.17 SD. 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 12 at the teacher–year level with ex ante value-added as the outcome variable. Panels 4b and 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.

[23] 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.

To address whether changes in teacher quality reflect a change in the volume of entries and exits, Appendix Figure A2 plots the corresponding event-study estimates under the same sample restrictions but with the dependent variable replaced in 12 with the number of entering and exiting teachers. Figures A1a and A1b provide evidence that entering and exiting teacher counts remain constant following principal transitions. Figure A2 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 5.2. Finally, Figure A2b 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 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 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)[24], 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 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%. Appendix Figure A5b 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[25].

This subsection provides evidence that higher-PVA principals raise aggregate teacher quality. While compositional changes largely reflect stronger teacher recruitment, as shown in Figure 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

[24] To be consistent with the specification in equation 12, Appendix Figure A5a 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.

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

teaching staff. I explore potential explanations in the next subsection.

5.2 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

$$y_{s,t} = \beta_0 + \beta_1 \hat{\theta}_p^{eb} + X'_{s,t}\phi_2 + \alpha_t + \alpha_s + \epsilon_{s,t} \tag{13}$$

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 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[26]. As the second measure already represents a within-school ranking of teachers, the school fixed effects are omitted when I estimate β_1.

Figure 6 plots the binscatter analogue of equation 13 examining the relationship between 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%[27]. 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.

[26] 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.

[27] Note that since the outcome variable relates to teacher quality, value-added in math is used for both teachers and principals. Appendix Figure A6 provides robustness using mean PVA.

Turning to classroom assignments, Figure 7a 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 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, 7b plots the same relationship after I residualize class size and TVA on school–year shocks to account for changes in cohort sizes over time or policies influencing the allocation of students[28].

Figure 7b 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. While recent work, such as Fenizia (2022), indicates that public-sector managerial quality is important for productivity, less is known about how exactly public managers shape outcomes. 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.

5.3 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 a scale of one ("strongly disagree") through five ("strongly agree") their sentiment on various

[28] 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 $\hat{\epsilon_{j,t}} = \beta_0 + \beta_1 \hat{\nu}_j + \gamma_{j,t}$.

questions related to overall school climate, teacher empowerment, and principal leadership[29]. 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[30]. Table 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 13.

Figure 8 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 A4. Appendix Table A2 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 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 8 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 empower-

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

[30]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).

ment, and effective leadership. This result suggests that certain aspects of school quality are attributable to differences in principal effectiveness. Appendix Figure A7 examines the relationship between principal quality and the individual questions used to construct the outcomes in Figure 8.

6 Conclusion

This paper reassess the role of principal quality in student outcomes and is among the first to show that value-added models estimating principal effectiveness are forecast unbiased. 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. While my results are smaller than prior estimates of teacher and school quality, principal effects are particularly relevant since principals are responsible for all students and not just those in a specific classroom.

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. 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.

References

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.

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.

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.

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.

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.

Biasi, Barbara, 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.

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.

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.

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.

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.

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'économique*, 2014, *47* (2), 634–663.

_ **and** _ , "How school principals influence student learning," *Empirical Economics*, 2018, *54* (2), 851–882.

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.

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, "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 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.

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

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.

Lemos, Renata, Karthik Muralidharan, and Daniela Scur, "Personnel management and school productivity: Evidence from india," Technical Report, National Bureau of Economic Research 2021.

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.

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.

_ , ""Fast Facts: Expenditures"," 2020.

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.

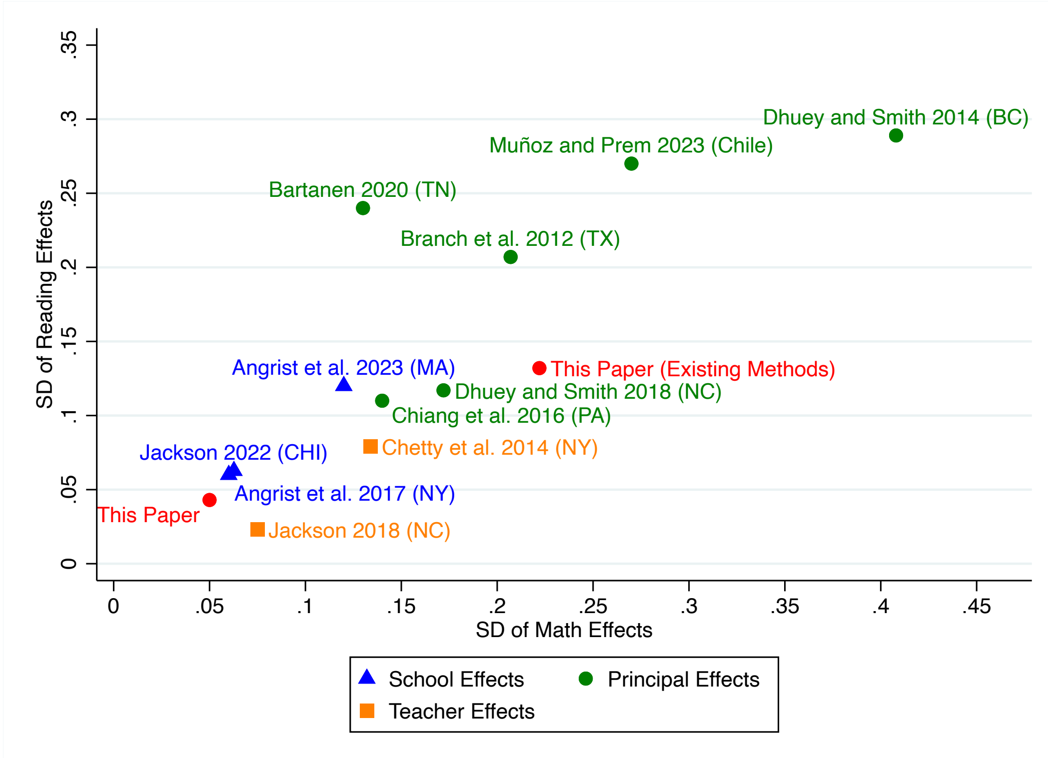
Rothstein, Jesse, "Teacher quality in educational production: Tracking, decay, and student achievement," *The Quarterly Journal of Economics*, 2010, *125* (1), 175–214.

_ , "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.

7 Figures and Tables

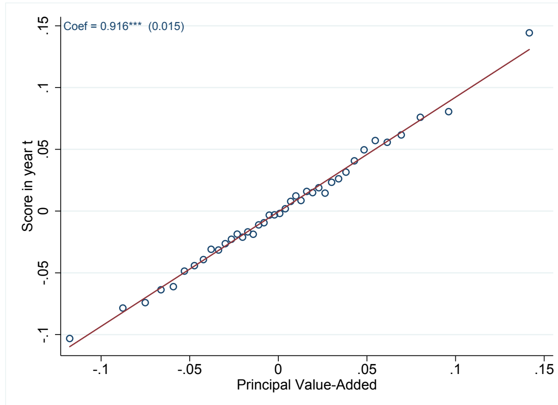
Figure 1: Comparison of Existing Estimates



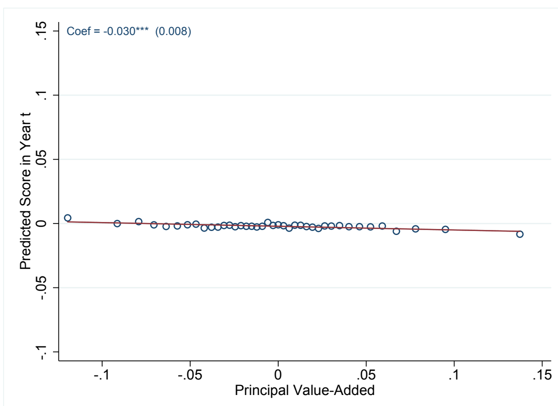
Notes: This figure plots existing estimates of school, principal, and teacher quality. School effects are denoted with blue triangles. With the exception of the red circles, which reflect my estimates, principal estimates are denoted with green circles. Teacher effects are denoted with orange squares. To obtain the red circles labeled "This Paper (Existing Estimates)", I follow the procedure discussed in Footnote 11.

Figure 2: PVA Predictions and Bias from Omitted Twice-Lagged Scores

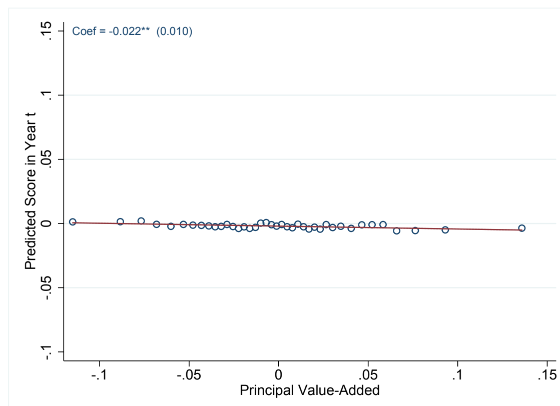
(a) Actual Score



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

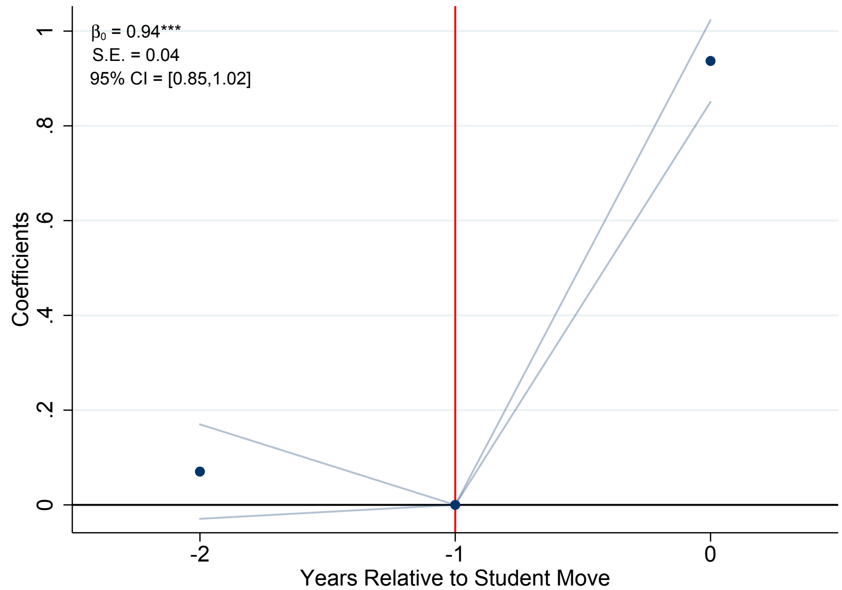


(c) Predicted Score Using Year $t-2$ Score (Principal Movers)



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

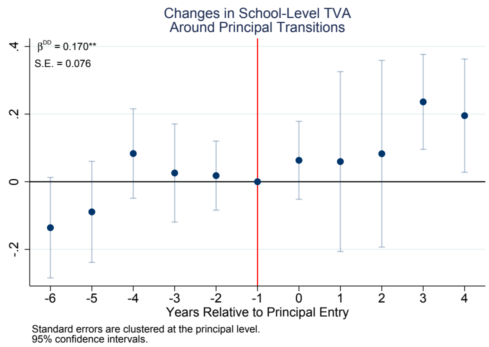
Figure 3: Effect of Mean TVA at $\tau = 0$ on Switchers' Test Scores



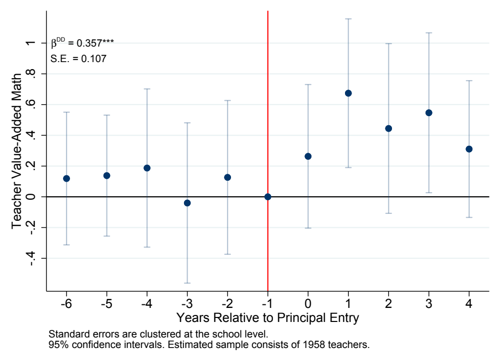
Notes: This figure plots the β coefficients from equation 7, which estimates the effect of a current principal's value-added on student test scores before and immediately after the switch occurs. The sample consists of students who switch schools and principals at $t = 0$. Estimates control for lagged 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 4: Changes in Principal Quality on Teacher Composition

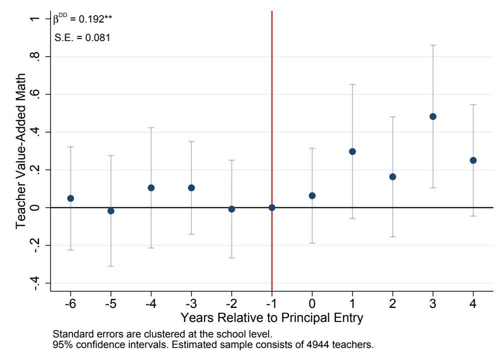
(a) Average Teacher Effectiveness



(b) Entering Teachers



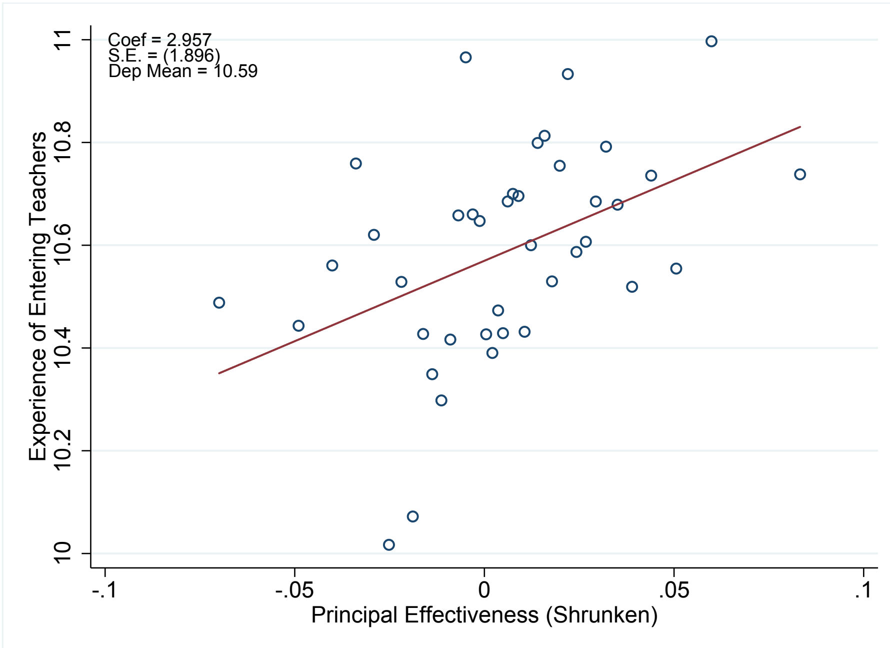
(c) Exiting Teachers



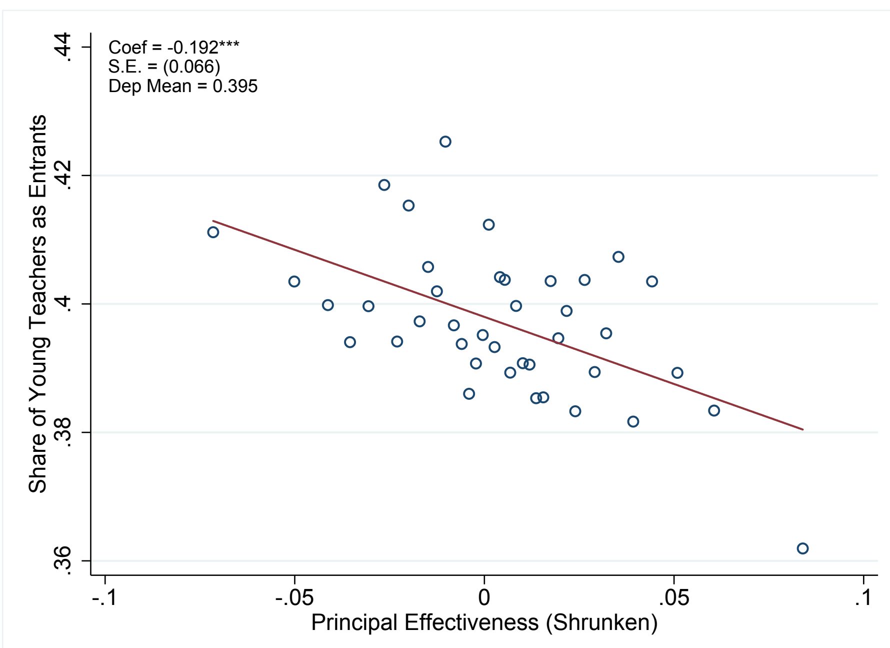
Notes: This figure plots the β coefficients from equation 12, which examines the effect of within-school changes in principal quality on average (school-level) and entering and exiting teacher value-added. The analysis focuses on "events", where the exiting and arriving principal are observed for four years; in total, there are 245 such events. Reported coefficients correspond to the difference-in-differences analogue of equation 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 5: Characteristics of Entering Teachers

(a) Experience of Entering Teachers



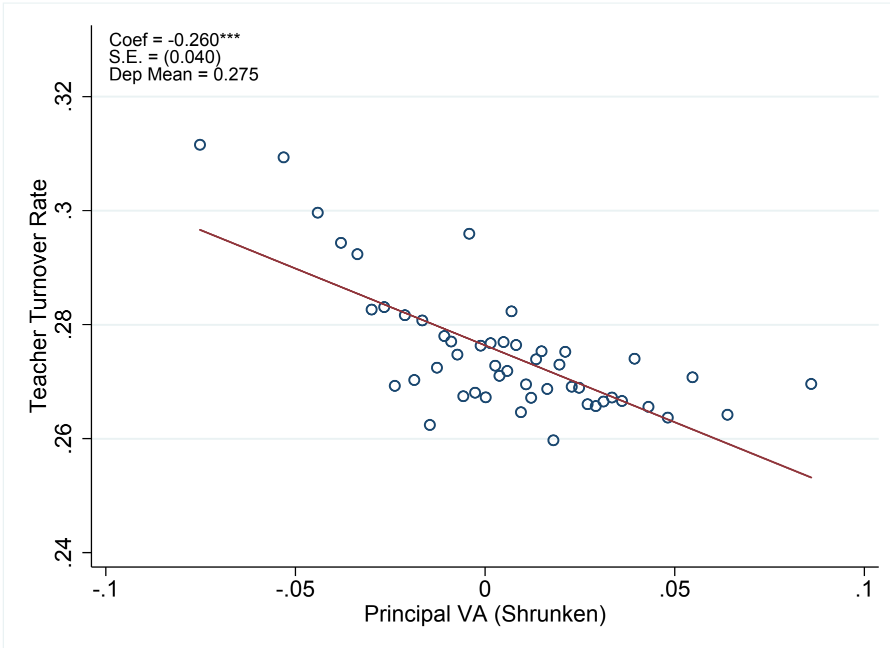
(b) Teachers with Less than 2 Years of Experience as Share of New Hires



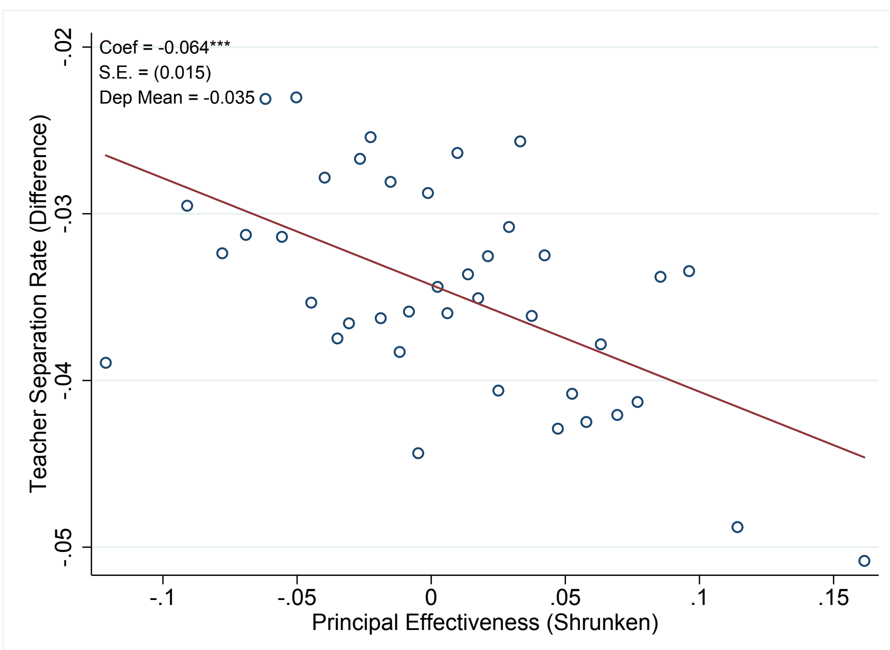
Notes: This figure shows the association between principal quality and characteristics of entering teachers. Panel A reports the results for the average experience level of entering teachers. Panel B reports the share of new entrants who have less than two years of experience. The binned scatter plots 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 12), school fixed effects and year fixed effects. The coefficient and standard error correspond to the identical regression at the school–year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 6: Principal Quality and Teacher Attrition

(a) Turnover: All Teachers



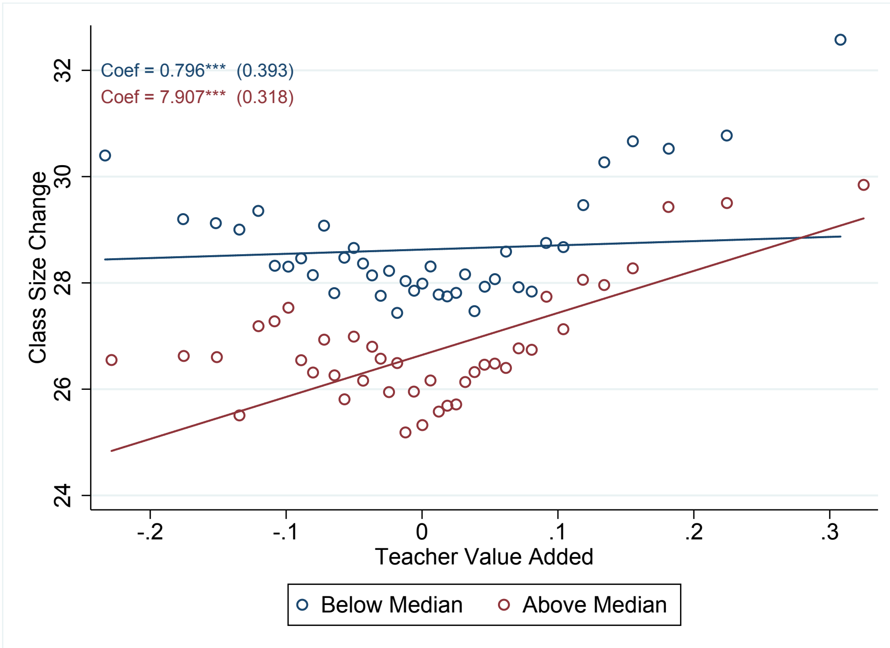
(b) Differential Turnover



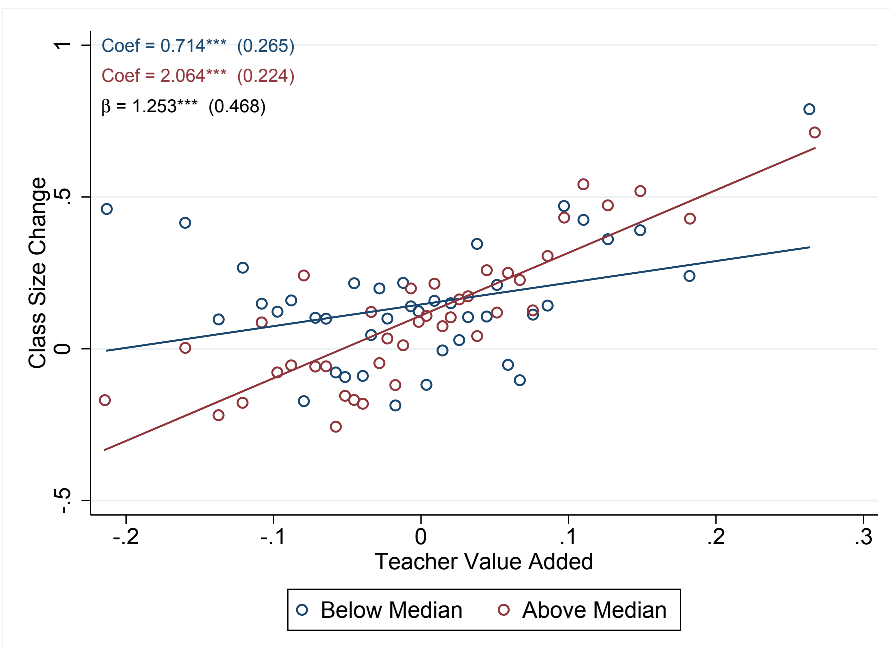
Notes: This figure plots the binned scatter plot of equation 13. Panel A plots the turnover rate for all teachers. Panel B plots the difference in annual turnover rates for teachers with above- and below-median TVA in mathematics for a given school. The binned scatter plots are divided into 40 equally 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 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 school-year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 7: Principal Quality and Classroom Allocation

(a) Raw Correlation



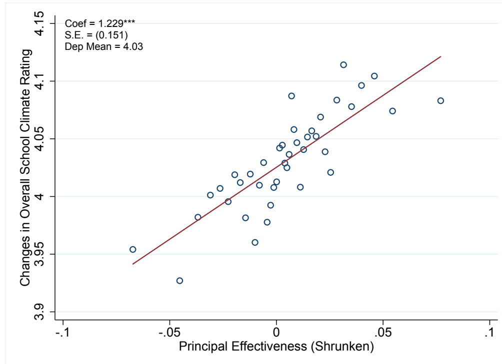
(b) Covariate Adjusted



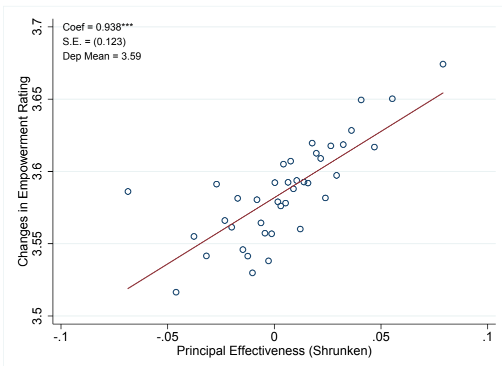
Notes: Panel A plots the raw correlation between class size and the shrunken TVA estimates separately for principals with above- and below-median value-added. Panel B plots the same association after I residualize both class size and the shrunken value-added estimates against school–year and grade fixed 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 8: Principal Quality and School Climate

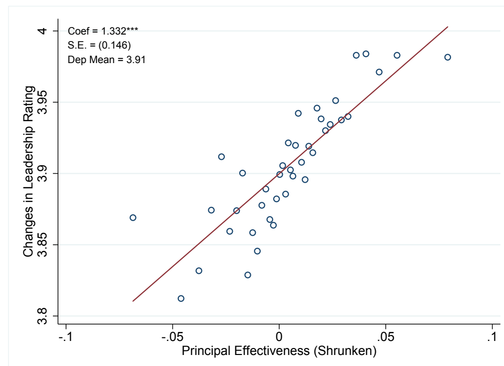
(a) Average School Climate



(b) Teacher Empowerment



(c) Effective Leadership



Notes: Each panel reports the binscatter associated with equation 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.

Table 1: Summary Statistics: AY 1996–2019

	All Principals		Principals Appearing in Multiple Schools	
	Mean	SD	Mean	SD
	(1)	(2)	(3)	(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 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 shock to 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 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	0.314***	0.079	0.073
	(0.053)	(0.056)	(0.055)
Year of Switch	1.214***	0.910***	0.937***
	(0.058)	(0.048)	(0.046)
School FE	No	Yes	Yes
Controls	No	No	Yes
P-value for Coeff = 1	0.0001	0.059	0.172
Observations	1,242,218	1,242,217	1,242,217

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 6. Standard errors in parentheses are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Out-of-Sample Validation:
Changes in Principal Quality and Changes in Mean Scores

	(1) Math	(2) Reading	(3) Mean	(4) Math	(5) Reading	(6) Mean
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 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 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 7: Relationship between Mean PVA and Log Salary

	(1)	(2)	(3)	(4)
Age	0.0091***	0.0091***	0.0095***	0.0091***
	(0.00066)	(0.00066)	(0.00092)	(0.00067)
At Least 4 Years of Exp	0.060***	0.060***	0.055***	0.062***
	(0.0075)	(0.0076)	(0.011)	(0.0079)
School Ranked in *USN&WR*	0.0084	0.0082	0.0017	-0.0076
	(0.017)	(0.017)	(0.024)	(0.017)
Female	-0.023***	-0.023***	-0.026**	-0.022***
	(0.0083)	(0.0082)	(0.012)	(0.0083)
White	0.017*	0.017*	0.0065	0.022**
	(0.010)	(0.010)	(0.016)	(0.0098)
Doctorate	0.0025	0.0026	0.0052	0.0071
	(0.011)	(0.011)	(0.017)	(0.012)
Mean PVA		0.045	0.081	0.17*
		(0.11)	(0.17)	(0.10)
Fixed Effects	District	District	School	None
Observations	6,423	6,423	6,423	6,423
R^2 (Adjusted)	0.11	0.11	0.052	0.064
F	66.7	58.8	27.0	62.7

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 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.

A1 Appendix: 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,

$$Var(\hat{\delta}_{j,s,p,t}) = cov(\hat{\delta}_{j,s,p,t}), \hat{\delta}_{j,s,p,t}) \tag{14}$$

$$= cov(\delta_{j,s,p,t} + \frac{1}{N_{j,s,p,t}} \sum_{i=1}^{N_{j,s,p,t}} \mu_{i,s,j,p,t}, \delta_{j,s,p,t} + \frac{1}{N_{j,s,p,t}} \sum_{i=1}^{N_{j,s,p,t}} \mu_{i,s,j,p,t}) \tag{15}$$

As noted in Bau and Das (2020), equation 15 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 2.

A2 Appendix: Alternative Decomposition

As a secondary approach to examine the variance of principal effects, I remove the school and teacher effects and model student achievement as strictly a function of observed student heterogeneity, principal shocks, and time-varying unobserved student heterogeneity. I estimate:

$$y_{i,g,p,t} = \sum_p \theta_p D_{i,p,t} + X'_{i,t}\delta + X'_{s.t}\gamma + \alpha_t + \alpha_g + \epsilon_{i,g,p,t} \tag{16}$$

where $D_{i,p,t}$ is an indicator for whether student i had principal p in year t. $X'_{i,t}$ is a vector of student heterogeneity, defined in equation 6, and $X'_{s.t}$ is a vector of school-level means of observable student characteristics. While $\hat{\theta}_p$ is unbiased when I draw from a random sample, the estimated variances likely exceed the true variance, as our estimates combine true principal effects with sampling error. Thus, I follow Rose et al. (2022) and Kline et al. (2020), and estimate σ_p^2 by constructing principal–year-level mean residuals from equation 16.

$$\bar{Y}_{p,t} = \frac{1}{n_{p,t}} \sum y_{i,t} - X'_{i,t}\hat{\delta} \tag{17}$$

$$= \theta_p + \underbrace{\frac{1}{n_{p,t}} \sum \epsilon_{i,g,p,t} + X'_{i,t}(\delta - \hat{\delta})}_{\psi_{p,t}} \tag{18}$$

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 sorting due to unobservable characteristics is uncorrelated across time for each principal. Under this assumption, then,

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

Table A1 provides the estimates of the variance of principal effects using the alternative decomposition. Using the entire sample of principals, the first row of Table A1 indicates that a one-SD increase in principal quality raises average student achievement by 0.068 SD. This estimate is approximately a 50% increase from Table 2, indicating the difficulty of quantifying principal quality in settings with limited turnover, as potential time-invariant forces are lumped in with principal quality.

The second row restricts the decomposition to only principals observed at multiple schools to isolate the potential school effects on the principal variances. The estimates from this restricted sample are nearly identical to those in Table 2. That the estimates from both approaches match is not surprising, as both methodologies estimate σ_p^2 from variation in test scores from principal movers.

Table A1: 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
Principals Movers: $cov(\bar{Y}_{p,t,s}, \bar{Y}_{p,t',s'})$	0.050	0.041	0.046
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.

A3 Appendix: Empirical Bayes Estimates of PVA

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:

$$\gamma_p = \theta^p + \sum_{j\in p} \frac{N_j}{N_p}\theta^j \tag{19}$$

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

$$Var(\gamma_p) = E[(\theta^p + \sum_{j\in p} \frac{N_j}{N_p}\theta^j)^2] \tag{20}$$

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

$$\begin{aligned} Var(\gamma_p) &= E[(\theta^p)^2] + E[(\sum_{j\in p} \frac{N_j}{N_p}\theta^j)^2] \\ &= \sigma^2_p + E[\frac{\sum_{j\in p} N_j^2}{N_p^2}\sigma^2_j] \end{aligned} \tag{21}$$

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

$$\gamma_p = \theta^p + \frac{1}{N_p}\sum_{i\in p}(\theta^j + \theta^s + \theta^{j,s,p,t} + \epsilon_{i,s,j,p,t}) \tag{22}$$

where the $i \in p$ denotes the set of students taught by principal p. Applying the same logic from equation 21, the variance of δ_p can be characterized as

$$\begin{aligned} Var(\gamma) &= E[(\theta^p + \frac{1}{N_p}\sum_{i\in p}(\theta^j + \theta^s + \theta^{j,s,p,t} + \epsilon_{i,s,j,p,t})]^2 \\ &= \sigma_p^2 + E[\frac{\sum_{j\in p} N_j^2}{N_p^2}\sigma_j^2] + E[\frac{\sum_{s\in p} N_s^2}{N_p^2}\sigma_s^2] + E[\frac{\sum_{j,s,p,t\in p} N_{j,s,p,t}^2}{N_p^2}\sigma_{j,s,p,t}^2] + E[\frac{1}{N_p}\sigma_\epsilon^2] \end{aligned} \tag{23}$$

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:

$$\lambda_p = \frac{\sigma_p^2 + \frac{\sum_{j\in p} N_j^2}{N_p^2}\sigma_j^2}{\sigma_p^2 + \frac{\sum_{j\in p} N_j^2}{N_p^2}\sigma_j^2 + \frac{\sum_{s\in p} N_s^2}{N_p^2}\sigma_s^2 + \frac{\sum_{j,s,p,t\in p} N_{j,s,p,t}^2}{N_p^2}\sigma_{j,s,p,t}^2 + \frac{1}{N_p}\sigma_\epsilon^2} \tag{24}$$

The shrinkage factor is then evaluated with the variance estimates reported in Table 2.

A4 Appendix: Workplace Environment and Teacher Quality

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 5.3 explain the observed variation in TVA. Following the teacher literature, I estimate

$$y_{i,j,g,t} = \beta_0 + \sum_a \beta_a y_{i,t-1}\mathbf{I}\{\text{grade} = a\} + \theta_j + X'_{i,t}\delta + X'_{j,s,t}\tau + \alpha_t + \alpha_g + \epsilon_{i,j,g,t} \tag{25}$$

where TVA is the estimate of θ_j, the teacher fixed effect, and the controls are identical to those in equation 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 4.3:

$$\hat{\theta}_j = \beta_0 + \tau X_j + \epsilon_j \tag{26}$$

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 missingness. Table A2 reports the regression results.

Table A2: TVA Correlates

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

A5 Appendix: Determinants of Principal Mobility

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:

$$\mathbf{I}\{\text{exit}_{p,t}\} = \beta_0 + \sum_{\tau=1}^{3} \beta_{k,t-\tau} Y_{s,t-\tau} + \lambda \mathbf{X}_p + \gamma \mathbf{X}_{s,t} + \alpha_s + \alpha_t + \epsilon_{p,t} \tag{27}$$

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 5.3 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 A3 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 A3: 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

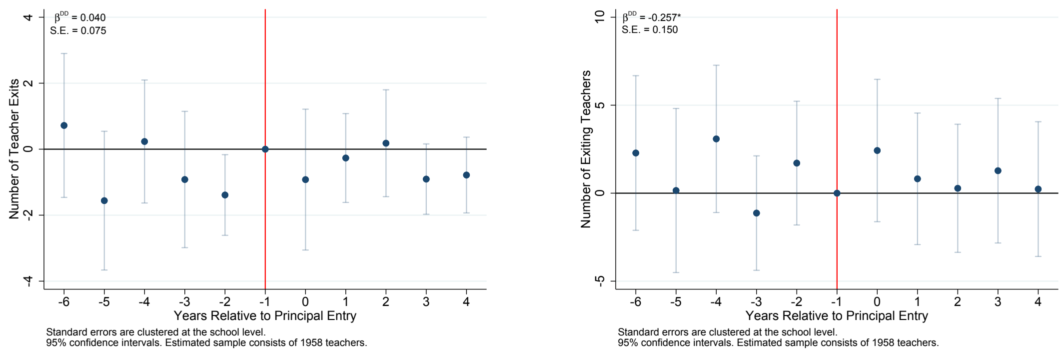
Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

A6 Appendix Tables and Figures

Figure A1: Principal Quality and Number of Entering and Exiting Teachers

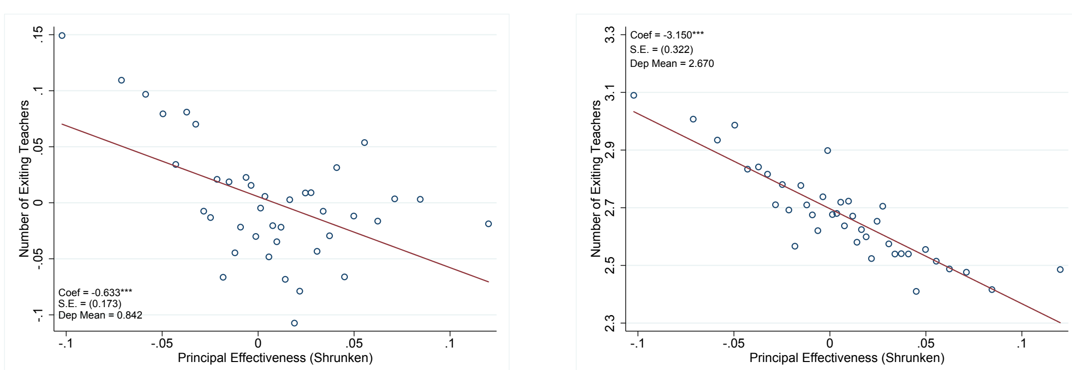
(a) Number of Entering Teachers (b) Number of Exiting Teachers



Notes: This figure plots the β coefficients from equation 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 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 A2: Principal Quality and Number of Entering and Exiting Teachers

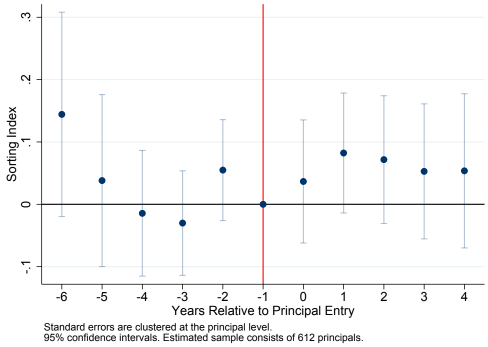
(a) Number of Entering Teachers (b) Number of Exiting Teachers



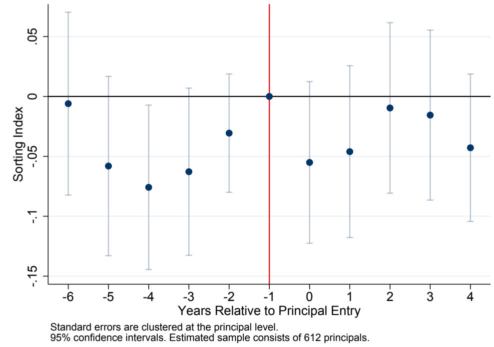
Notes: This figure plots the estimates from a binned scatter plot regression using the full sample of principals: $y_{s,t} = \beta_0 + \beta_1 \hat{\theta}_p^{eb} + X'_{s,t}\phi_2 + \alpha_t + \alpha_s + \epsilon_{s,t}$ where $y_{s,t}$ is either the number of entering existing teachers or the total number 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 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 principal level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A3: Principal Quality and Student Composition

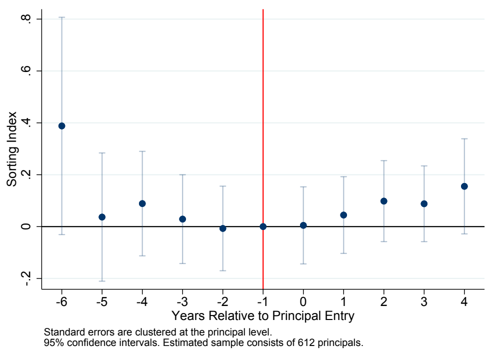
(a) Black or Hispanic



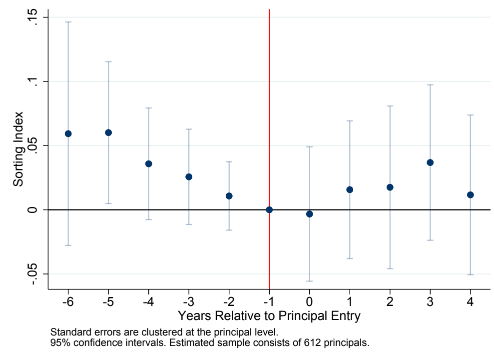
(b) Male



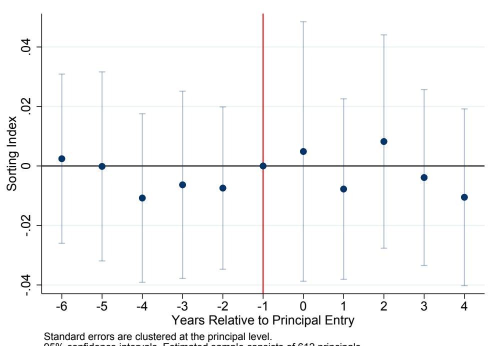
(c) FRLP



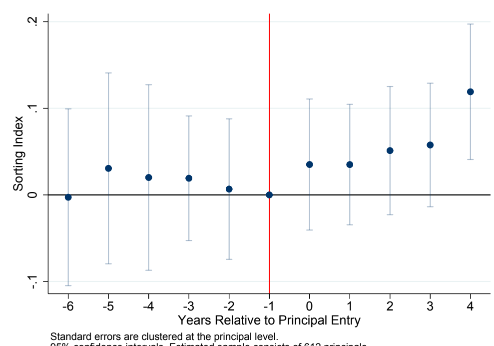
(d) LEP



(e) Repeat Subject



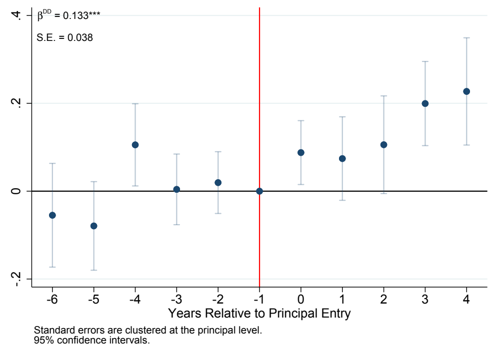
(f) Academically Gifted Students



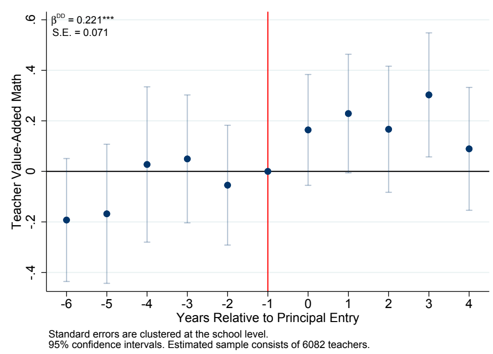
Notes: This figure plots the β coefficients from equation 12 where the outcome variables are the share of students who are Black or Hispanic, are male, are eligible for free or reduced-price lunch (FRLP), have limited English proficiency (LEP), have repeated a subject from a previous year, or are classified as academically gifted. Standard errors are clustered at the principal level, and 95% confidence intervals are reported.

Figure A4: Changes in Principal Quality on Teacher Composition

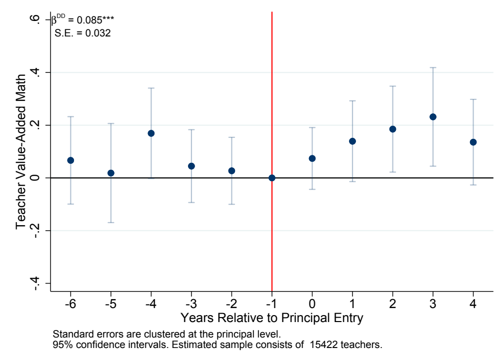
(a) Average Teacher Effectiveness



(b) Entering Teachers



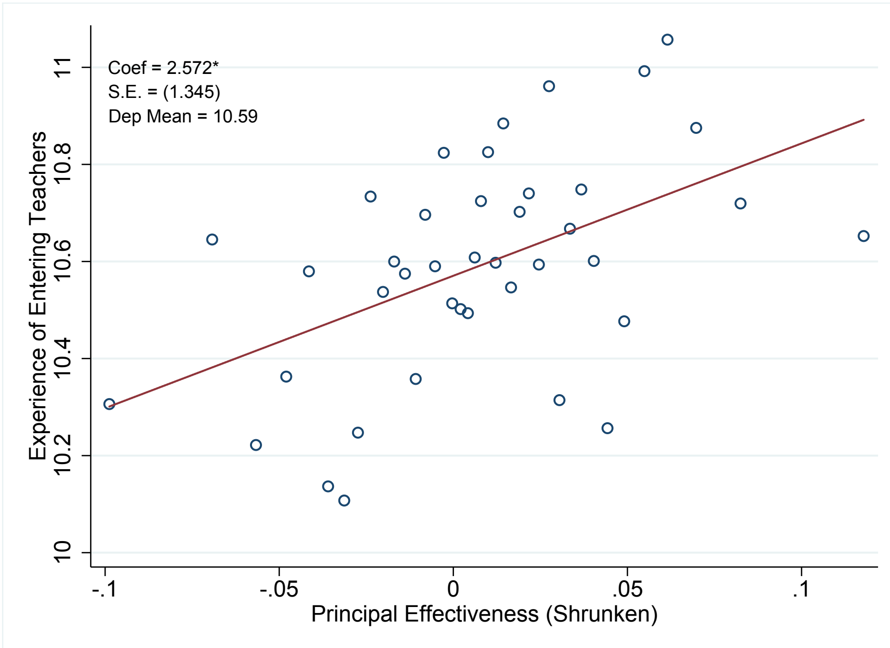
(c) Exiting Teachers



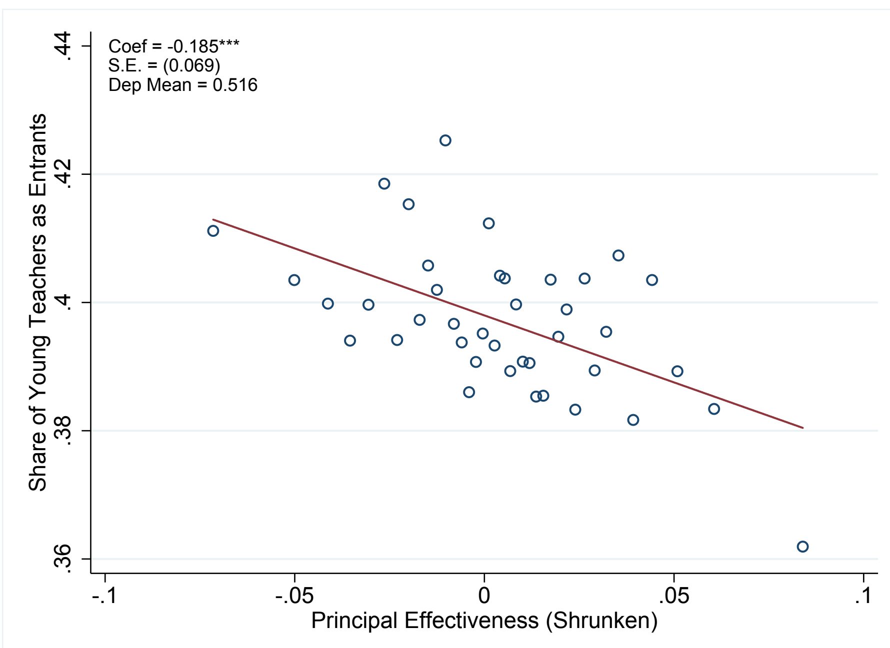
Notes: This figure plots the β coefficients from equation 12, which examines the effect of within-school changes to principal quality on average (school-level) and entering and exiting teacher value-added. No restrictions are placed on the number of observed years for the entering and exiting principal. Reported coefficients correspond to the difference-in-differences analogue of equation 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 A5: Characteristics of Entering Teachers

(a) Experience of Entering Teachers

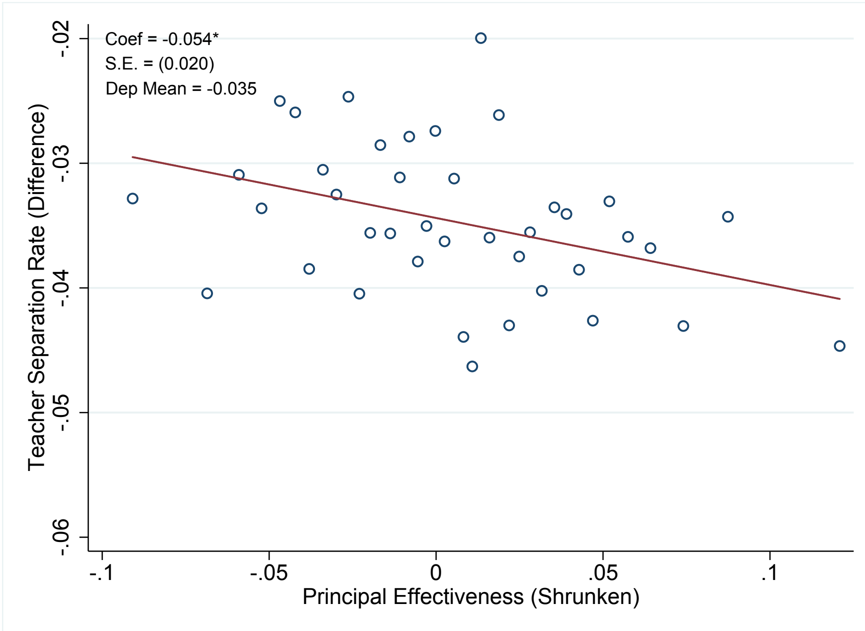


(b) Teachers with Less than 4 Years of Experience as Share of New Hires



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 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 school-year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

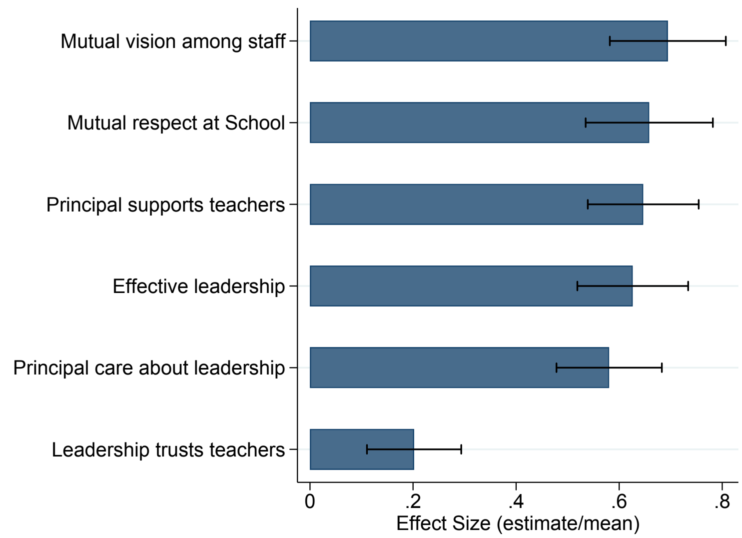
Figure A6: Principal Quality and Teacher Attrition



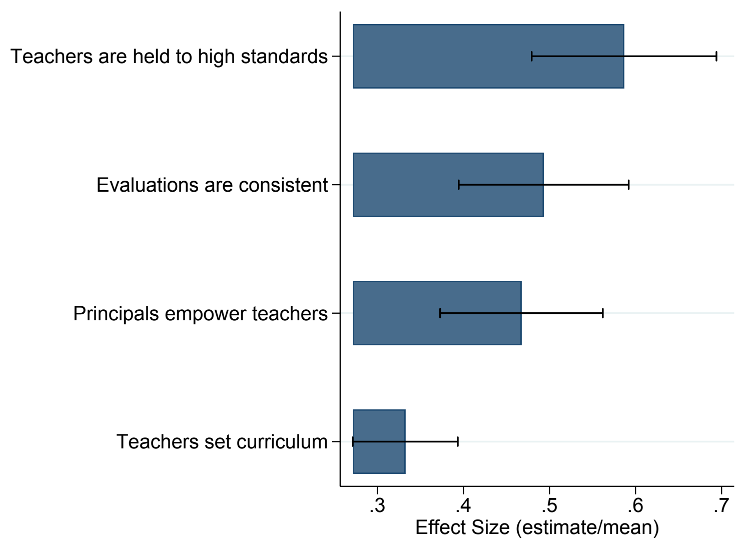
Notes: This figure plots the binned scatter plot of equation 13 where the outcome variable is the difference in attrition rate between teachers above and below median value-added in mathematics and the main dependent variable is mean PVA. The binned scatter plots are divided into 40 equally sized bins 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 12), school fixed effects and year fixed effects.The coefficient and standard error correspond to the identical regression at the school-year level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A7: 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 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 13, the outcome variable corresponds to the degree to which respondents agree with the given outcome variable.