NATIONAL CENTER *for* ANALYSIS *of* LONGITUDINAL DATA *in* EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

CALDER

AIR

Do More Effective Teachers Become More Effective Principals?

Dan Goldhaber

Kristian Holden

Bingjie Chen

WORKING PAPER No. 215-0119-1 • January 2019

Do More Effective Teachers Become More Effective Principals?

Dan Goldhaber
American Institutes for Research/CALDER
University of Washington

Kristian Holden
American Institutes for Research/CALDER

Bingjie Chen
American Institutes for Research/CALDER

Contents

Acknowledgments

This research was supported by the National Center for the Analysis of Longitudinal Data in Education Research (CALDER), which is funded by a consortium of foundations. For more information about CALDER funders, see www.caldercenter.org/about-calder. We thank James Cowan and Umut Özek for comments. All opinions expressed in this paper are those of the authors and do not necessarily reflect the views of our funders or the institutions to which the authors are affiliated.

CALDER working papers have not undergone final formal review and should be cited as working papers. They are intended to encourage discussion and suggestions for revision before final publication. Any opinions, findings, and conclusions expressed in these papers are those of the authors and do not necessarily reflect the views of our funders.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street NW, Washington, DC 20007
202-403-5796 • www.caldercenter.org

Do More Effective Teachers Become More Effective Principals?

Dan Goldhaber, Kristian Holden, Bingjie Chen
CALDER Working Paper No. 215-0119-1
January 2019

Abstract

Principals are widely seen as a key influence on the educational environment of schools, and nearly all principals have experience as teachers. Yet there is no evidence on whether we can predict the effectiveness of principals (as measured by their value added) based on their value added as teachers, an issue we explore using administrative data from Washington. Several descriptive features of the principal labor market stand out. First, teachers who become principals tend to have higher levels of educational attainment while teaching and are less likely to be female, but we find no significant differences in licensure test scores between those teachers who become principals and those we do not observe in the principalship. Second, principal labor markets appear to be quite localized: about 50 percent of principals previously taught in the same district in which they assumed a principalship. We find positive correlations between teacher and principal value added in reading (ELA) and similarly sized but less precise estimates in math. Teachers who become principals have slightly higher teacher value added, but the difference between the two groups is not statistically significant, suggesting that principals are not systematically selected based on their prior effectiveness when serving as a classroom teacher.

1. Introduction

Principals are widely seen as a key influence on the educational environment in the schools they lead, and a relatively new body of empirical evidence suggests they play an important role in affecting student outcomes.[1] Principals may affect their schools in a variety of ways. For example, they can serve as "instructional leaders" who promote high-quality instruction and create an environment conducive to student success. They may also influence the composition of the teacher workforce through hiring, counseling out, and retention. The idea that school leaders are important is buttressed by a broader economic literature on leadership in the private sector. This research finds that supervisors vary significantly in their effectiveness and replacing an ineffective supervisor with an effective one can significantly enhance the output of team production (e.g., Lazear et al., 2015).This can occur in a variety of ways, from influencing decision making to enhancing the productivity of those who are supervised (Lazear, 2012).

Despite a belief in the importance of leadership, the quality of school principals has received relatively little focus as a way to improve outcomes for students, and principals have often been an afterthought in school improvement efforts (Rowland, 2017; Rotherham, 2010). This is changing as policymakers are increasingly turning their attention to the ways that principals are developed, recruited, selected, and evaluated, and how various policy levers may influence the quality of the principal workforce. This is evidenced by the focus on the quality of school principals in the federal *Every Student Succeeds Act*, which requires states to submit plans for improving school leadership. These plans are quite varied—some states focus on the principal pipeline and requirements to become a principal, while others are focused on training and on-the-job supports (Newleaders.org, 2018).

Unfortunately, policymakers are operating in a bit of an empirical vacuum, as we know relatively little about the specific prior experience, training, or personal traits that predict which individuals will make effective principals (we discuss this in more detail in the next section). Importantly, however, nearly all principals were previously classroom teachers (Austin et al., 2019), offering the possibility that we might learn about the potential for school leadership based on an individual's performance as a teacher.

In this paper, we focus on the connections between teacher and principal effectiveness using administrative data from Washington state that allow us to estimate direct measures of effectiveness: teacher and principal contributions to student test achievement ("value added"). We find that teachers who become principals tend to have higher levels of educational attainment and are less likely to be female, yet the results suggest no significant differences in licensure test scores between those teachers who become principals and those we do not observe in the

[1] See, for instance, research on school leadership (Hallinger & Murphy, 1985; Blasé & Blasé, 2004; Bottoms & O'Neill, 2001; Glickman, Gordon, & Ross-Gordon, 2014; Hoy & Hoy, 2003), the organizational management of schools (March, 1978; Heck, 1992; van de Grift & Houtveen, 1999; Balu, Horng, & Loeb, 2010; Grissom & Loeb, 2011), or work linking principals to student outcomes (Branch et al., 2009; Clark, Martorell, & Rockoff, 2009; Miller, 2013; Dhuey and Smith, 2014; Grissom et al., 2015; Chiang et al., 2016); Austin et al., 2019.).

principalship.[2] As is the case for teacher labor markets (Boyd, et al., 2005; Goldhaber et al., 2013; Krieg et al., 2016; Ronfeldt et al., 2018), principal labor markets are quite localized: about half of principals have prior experience as a teacher in the same district, and 20% to 25% have experience teaching in the same school.

We find that teacher value added in reading is strongly predictive of principal value added in reading, and similarly sized effects that are less precisely estimated emerge in math.[3] We also find some evidence that teaching in tested grades for math (and thus having math value-added estimates) is positively predictive of principal effectiveness.[4] Our estimates are not sensitive to selection into the principalship; however, we note that there are conceptual reasons to be cautious about causal interpretation of principal value-added estimates. Even models that use within-school differences in principal effectiveness may reflect the characteristics of the previous principal, because the influence of one principal may transcend his or her spell at a school.

Research on private sector labor markets suggests that firms strongly prefer internal hires. Consistent with this literature, we find large differences in the characteristics of principals depending on whether they have prior teaching experience within the district, and within the school—with internal hires having less educational attainment. We add to prior research by considering whether these individuals are differentially effective. Contrary to the notion of positive specific human capital effects, we find evidence that internal hires within a school (teachers who are promoted to the principalship in the same school in which they once taught) are less effective relative to external hires, whereas the difference between hires internal to the school district (but not school) and external to the school district is not statistically significant.

Overall, this research lends support to a growing body of research that relates traits of principals to student achievement, but it also shows the sensitivity of the findings to the specification of principal value added. Additionally, the fact that we find little evidence that teacher effectiveness plays a role in determining who ends up in a principalship suggests there is significant scope for improvement in who is selected into a principalship, as teacher value added appears to be an indicator of principal performance.

[2] Right censoring is an important consideration because some of the newly hired teachers in our sample have not yet been observed as a principal but will eventually take on this role. We attempt to address this by estimating models that limit our sample to more experienced teachers.
[3] Washington tested students in "reading" from 1997-98 to 2013-14 and "English Language Arts (ELA)" from 2014-15 on. For simplicity, we refer both reading and ELA tests as "reading."
[4] Teachers from 2007-08 to 2016-17 can have teacher value-added estimates if they are in tested grades and subjects; however, individuals observed *before* this period can have experience in tested grades and subjects without value-added scores. We discuss this in more detail in Section 3 below.

2. Background on Path to the Principalship and Links to Effectiveness

2.1 What Do We Know About the Principal Pipeline?

There is only a sparse literature on who seeks to become a principal, but the great majority of principals have some prior experience as teachers. Austin et al. (2019), for instance, examine the path to the principalship across a number of states and find that, in all of them, 80 percent or more principals have prior experience as a teacher.

A growing body of research focuses on the process of selecting principals. This is important because research documents difficulties in hiring principals (see Cooley & Shen, 2000; Fenwick & Pierce, 2001; Hammond et al., 2001; Malone & Caddell, 2000; Whitaker, 2003; Winter & Morgenthal, 2002), which appears to be more problematic for disadvantaged schools (Loeb, Kalogrides, & Horng, 2010). Several studies suggest that states certify more administrators than required to fill vacancies (Pounder, Galvin, & Shepherd, 2003; Lankford, O'Connell, & Wyckoff, 2003), and open positions will typically receive multiple applications (Roza, 2003), suggesting a mismatch between supply and demand of principal candidates. In this section, we discuss how the selection process functions, and how the supply of candidates and the demand for candidates are likely to impact the hiring and quality of selected principal candidates.

The hiring process involves two parties. First, from the demand perspective, school representatives seek to fill principal positions.[5] School representatives define the needs of the role and recruit potential candidates, and are likely to influence who applies to positions through informal mechanisms. For example, research by Myung, Loeb, and Horng (2011) considers the informal process of selecting principal candidates via "tapping," where principals reach out to recruit teachers in their schools or districts who appear to have promising skills to be effective principals. Next, school representatives must select individuals from the pool of applicants to offer the position. Selection will depend on criteria chosen by the representatives, such as perceived candidate skills, experience, and attitude. Research by Rammer (2007) suggests that superintendents in Wisconsin tend to look for candidates with skills in the categories of communication, culture, outreach, and visibility. One area of concern motivating our own research is that school representatives report having difficulty identifying suitable candidates for the principal position—especially in identifying candidates with the skills they most value. Similarly, Whitaker (2003) find that 30.2% of surveyed superintendents rate the quality of principal candidates as either 1 or 2 on a 5-point scale. Lastly, Roza (2003) finds that 80% of surveyed superintendents indicate moderate or major problems identifying qualified school principals.

Second, from the supply perspective, applicants decide whether to apply for the position, and whether to accept the job when offered. A relatively large body of literature on the supply of

[5] See Hay Group (2006) for more discussion of the principal hiring process from the perspective of the hiring agency.

principal candidates suggests that applicants are deterred from the position by increasing demands of the job and accountability requirements, and the additional salary and status is not sufficient motivation (Harris et al., 2003; Whitaker, 2003). As summarized by Myung, Loeb, and Horng (2011), this appears to be less about a shortage of available candidates and more about the change in the requirements and skills for the job.[6]

In addition to school and district-level hiring practices, there are many state-level policies that could be used to influence principal effectiveness. All 50 states have adopted standards that are required in order to serve as a school principal, and many of these standards differ across states.[7] For example, 37 states require that principals have a master's degree and 3 years of teaching experience, and 38 require field experience. In addition to traditional preparation programs, 39 states allow for alternative requirements depending on applicant qualifications, which vary widely in their requirements.[8] Clearly, there are very different approaches used to determine who should be eligible for a principalship, but the lack of evidence on the relationship between principals' training and prior experience and their impacts on schools and students means that policy decisions, such as the implementation of standards, are largely being made in an empirical vacuum.

2.2 Possible Links Between Teacher and Principal Effectiveness

There are several reasons to think that teachers might have important insights into the principal role and that more effective teachers would be expected to make for more effective principals. To begin, we would expect that teachers could learn a good deal about the role of principals through interacting with their principals directly, with observations of the teacher-principal relationship providing important insights on the job's requirements (Rammer, 2007).

There is also evidence that effective teachers can support the development of their peers. Papay et al. (2016) find that pairing high and low-performing teachers, and working on improving teaching skills can raise the performance of the teacher pairing, and the low-performing teacher in particular.[9] Since one of the roles of principal is to serve as an instructional leader who mentors struggling teachers (Hallinger & Murphy, 1985; Blasé & Blasé, 2004; Bottoms & O'Neill, 2001; Glickman, Gordon, & Ross-Gordon, 2014; Hoy & Hoy, 2003), it is

[6] Several studies suggest that states certify more administrators than required to fill vacancies (Pounder, Galvin, & Shepherd, 2003; Lankford, O'Connell, & Wyckoff, 2003), and open positions will typically receive multiple applications (Roza, 2003). Work by Roza (2003) suggests that districts struggle to hire school leadership positions because individuals do not possess the necessary skills to be successful, and research by Pounder & Merrill (2001) and Winter & Morgenthal, (2002) suggests that changing demands of the principalship deter potential applicants.
[7] As reported by the Education Commission of the States, https://www.ecs.org/50-state-comparison-school-leader-certification-and-preparation-programs, retrieved 9/26/18.
[8] For example, Utah makes exceptions for individuals with "exceptional professional experience," while Virginia allows for exceptions based on concentrations of graduate coursework in "school law, evaluation of instruction, and other areas of study required by the employing Virginia school superintendent."
[9] This is related to work by Goldhaber et al. (2018) and Ronfeldt et al. (2018) that considers the influence of mentor teachers on teacher candidates, and finds that assignment to more effective mentors is associated with the effectiveness of their mentees who enter the teacher labor market.

natural to think that effective teachers might also serve as role models and thus be more effective instructional leaders.

Finally, some of the determinants of teacher effectiveness could be associated with innate characteristics of the individual, such as ability and motivation. For example, a seminal paper by Weiss (1995) finds evidence that signaling models are better able to explain the hiring of workers to firms relative to human capital models, which suggests that the unobserved fixed traits of workers are more important than returns to schooling. To the degree that teaching and principal success depends on similarly fixed and unobserved traits of the individual (as opposed to human capital accumulation as a teacher), we might also expect these traits to influence the relationship between teacher and principal effectiveness.

Another important consideration is whether to hire a principal internally—from teachers within a school or district—or externally. This decision between internal and external hiring has been a focus of significant literature based on the private sector.[10] For example, Jovanovic (1982) presents a theoretical model that highlights the role of uncertainty about employee ability when hiring. Consistent with this idea, DeVaro, Kauhanen, and Valmari (2015) and Marita and Tang (2019) find that firms appear to have preferences for internal hiring; to overcome these preferences, external hires tend to have prior experience in the role, more educational attainment, and more experience.

Several recent studies have estimated principal value-added models to investigate the impact of principals on student achievement.[11] These value-added models attribute the improvements in student achievement between a student's current test scores and previous test scores to principals while taking into account the observable characteristics of students, classes, and schools. Given the challenges of estimating principal value-added (which we describe in greater detail below in Section 3.2), it may not be surprising that there are some significant differences in the estimated variation in principal effectiveness across studies. For example, Branch et al. (2009) use data from Texas and find that a 1–standard deviation increase in principal value added is associated with a 0.11–standard deviation increase in math scores while Dhuey and Smith (2014) consider a unique setting in British Colombia that creates principal mobility because principals regularly rotate between schools; they find that a 1–standard deviation improvement in principal value-added associated with an increase in student achievement of 0.289 to 0.408 standard deviations in reading and math between Grades 4 and 7, suggesting an average improvement of 0.01 to 0.14 per grade. Lastly, cross-state research by Austin et al. (2019) estimates principal value-added models across 6 states and finds standard deviations range between 0.06 and 0.10. While there are differences in methods and estimated magnitudes of principal effects across studies, these papers tend to suggest that there are important differences in the effectiveness of principals.

[10] Seminal work by Baker, Gibbs, and Holmstrom (1994) use twenty years of personnel data from one private firm to describe the "hierarchical structure" of the firm, which suggests that employees tend to be promoted internally in predictable and stable pathways within the firm.

[11] For evidence on the import of managers in the private sector, see Lazear et al. (2012).

There are myriad mechanisms through which principals could affect students but influencing teacher quality is surely a key one given what we know about the importance of teachers for student outcomes (Anderson et al., 2007; Chetty et al., 2014; Jackson, 2018; Rivkin et al., 2005). For example, they could affect the quality of teachers in their schools by changing the performance of incumbent teachers—for example, by creating a particular learning environment or by changing the mix of teachers in their schools. Recent evidence by Cohen et al. (2018) provides encouraging evidence that principals' beliefs about their ability to affect teacher effectiveness do matter. Specifically, principals who perceive that they have greater agency are more likely to utilize evaluation and tenure review policies and practices aligned with the strategic goal of improving the quality of the teachers they supervise. And, Grissom and Bartanen (2018) find that more effective principals are associated with differential teacher turnover where retention is concentrated among high-performing teachers.

Yet despite the evidence that principals affect student outcomes, there is little evidence that preservice principal characteristics predict their effectiveness. Clark et al. (2009), for instance, find little evidence that the level of degree attainment or prestige of the degree-granting institution are associated with student outcomes.[12] To our knowledge, there is no existing evidence on whether teacher effectiveness predicts principal effectiveness; and though there is evidence from the private sector that more effective employees are more likely to be promoted (Baker, Gibbs, & Holmstrom, 1994), it is not yet established that these individuals are more likely to turn out to be more effective managers.

3. Methodology and Data

3.1 Estimating Teacher Value-Added Models

There is a significant body of research that estimates and validates value-added models of teacher effectiveness, through simulations (Goldhaber & Chaplin, 2015; Guarino, Reckase, & Wooldridge, 2015), quasi-experimental designs (Bacher-Hicks et al., 2014; Chetty et al., 2014), and experimental designs (Kane et al., 2008, 2013). On the whole, this literature supports the notion that, if properly specified, value-added models provide estimates of teacher contributions to student test score gains that are likely to have limited to no bias (Koedel et al., 2015).

Based on this, we estimate teacher value-added models having the following general form:

$$Y_{isjt} = f(Y_{it-1}) + \alpha_1 X_{it} + \tau_j + \varepsilon_{isjt} \quad (1)$$

where Y_{isjt} represents student i's test score in subject s, taught by teacher j in year t. The first term on the right, $f(Y_{it-1})$, is a cubic polynomial of prior standardized test scores in math and reading for student i, specified as a cubic polynomial. X_{it} represents a vector of student-level

[12] This mirrors many comparable findings from the teacher value-added literature on degree attainment and credentials (e.g., Chingos & Peterson, 2011; Goldhaber & Brewer, 2000).

controls such as gender, ethnicity, and participation in free or reduced lunch (FRL), special education services, and limited English proficiency (LEP) programs, and indicators for grade and school year. Lastly, ε_{isjt} is a mean-zero error term.[13] The coefficient of interest is τ_j, the estimated teacher value added score for teacher j. Teacher value added is then normalized across the sample within grade and year. We drop cases where fewer than 10 students are associated with a given teacher.

One area of contention in the specification of value-added models is whether to include classroom level student covariates to capture peer effects between students. Controlling for peer effects (which would imply, for instance, that having a classroom with highly concentrated poverty is more challenging) is appealing (Isenberg et al., 2016). However, models with classroom controls may *over* control for peer effects by removing true variation in teacher quality (Goldhaber et al., 2016). Since specifications with and without classroom covariates have been validated in validity tests of value added,[14] we estimate teacher value added ("TVA") using both models to check the robustness of our findings. As we describe in Section 3.4, however, the two estimates are highly correlated, and our findings are little influenced by the choice of TVA specification.

Our primary specification pools teacher value-added estimates over time, which is appealing from a reliability standpoint (Koedel & Betts, 2011). But, we also follow the general practice of using a Bayesian shrinkage procedure where we weight the mean of teacher value added more heavily as the standard error for a teacher's individual value added estimate increases; in simple terms, this adjustment shrinks imprecise estimates of teacher value added towards the mean (e.g., Herrmann, Walsh, Isenberg, & Resch, 2013).[15] This process reduces the impact of measurement error and attenuation bias in our analysis.

A tradeoff of using teacher value added pooled across years is that it will obscure important variation over a teacher's career. In particular, returns to experience in value added suggest that teachers will tend to be more effective at the end of a teacher's career, so that end-of-career TVA may more accurately reflect their ability at the time that they are being considered

---

[13] We estimate TVA models separately by grade span, K-8 and high school, as K-8 models are estimated using lagged student test scores from their previous school year (t-1). High school TVA models require the use of lagged student test scores from earlier grades for some students, depending on which end-of-course (EOC) exams are available (e.g. 8th grade general math tests can be linked to either 9th grade Algebra or 10th grade Algebra EOC exams depending on course taking). For more discussion of high school TVA models, see Theobald, Goldhaber, Gratz, and Holden (2018).

[14] Chetty et al. (2014) examine teachers who switch grades or schools and employ a quasi-experimental approach to validate value-added specifications that include peer effects. But, an investigation of students admitted or excluded from elite public schools based on a lottery (Abdulkadiroğlu, Angrist, & Pathak, 2014) suggests that peer effects have little influence on student test achievement.

[15] Our empirical Bayes estimates are calculated as follows:

$$\hat{\tau}_i^{EB} = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \widehat{se}_i^{\,2}} * \hat{\tau}_i + \left(1 - \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \widehat{se}_i^{\,2}}\right) * \bar{\tau} = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \widehat{se}_i^{\,2}} * \hat{\tau}_i$$

Where $\hat{\tau}_i$ is the estimated value-added score for teacher i, $\bar{\tau}$ is the average value-added score normalized to zero, $\hat{\sigma}^2$ is the estimated variance of value-added, and $\widehat{se}_i$ is the estimated standard error of value-added for teacher i.

for a principal position. There is a clear trade-off with precision, however, as single year estimates will tend to contain more sampling error (for instance, see Koedel et al., 2015).

Another issue is that teachers will tend to have different periods of time between teaching and the principal role. On one hand, this may not matter much as a significant component of teacher effectiveness appears to be persistent over time (Goldhaber & Hansen, 2013), suggesting that pooling across years will tend to capture the *persistent* component. Alternatively, as discussed in Section 2, effective teachers may become more effective principals because they serve as "instructional leaders" and use their prior teaching experience. If this experience is many years in the past, TVA may be less representative of the abilities of the individual. We address these possibilities by estimating alternative models (see Appendix A) that use single-year estimates of teacher value added that are more proximal to the principalship, and, as we discuss below, by directly controlling for the number of years between the value-added estimates and the time as a principal.

3.2 Estimating Principal Value-Added Models

In the case of teacher VA models, questions about which specifications limit bias appear to come down to whether to include classroom covariates. In contrast, the concerns for principal value-added models are much more fundamental. The specifications of principal value-added models have not been thoroughly explored and their validity have not been tested.[16] Moreover there are conceptual reasons to be skeptical that any principal value-added model should be interpreted as the causal contribution of principals to student test achievement (Austin et al., 2019).

One challenge in estimating principal value added ("PVA") is the difficulty of separating a principal effect from other school-level factors, such as the collegiality of teachers, that may influence student achievement (Grissom et al., 2015; Chiang et al., 2016). Another problem, noted by Clark, Martorell, and Rockoff (2009), is that students are repeatedly served by the same principal, which is likely to cause autocorrelated errors and inconsistent estimates. For example, middle school principals can affect a cohort of students in grades 6, 7, and 8, so that in later grades, lagged student achievement is endogenous to the principal's effectiveness.[17] But there is a more fundamental problem: the influence of one principal may transcend that person's spell at a particular school, meaning the influence of one principal may be misattributed to the principal that next assumes the principalship (Austin et al., 2019). In what follows, we address how the models we estimate do or do not address the above concerns.

[16] There is evidence that principal value added is correlated with principal observation ratings (Grissom et al., 2015), but there are no validation studies along the lines of Chetty et al. (2014) or Kane et al. (2013) for teachers. Moreover, unlike the case for teachers where the distribution of value added is quite consistent across different settings (Hanushek and Rivkin, 2010), the principal value-added studies we reviewed in Section 2.2 finds quite varied differences in the distribution of principal effectiveness, suggesting sensitivity to specification and context.

[17] One could explore principal value-added models limited to students who are initially served or students who exit the school, but if the results differ, it would not be clear whether the models are biased or if student contributions differ for these subpopulations.

We begin with a specification which we refer to as the "school value-added model" that *does not* control for the characteristics of the school or use within-school variation to identify principal value added. As such, there are many reasons to suspect that this model will result in biased estimates of principal effectiveness (Chiang et al., 2016). So, while we *do not* favor this specification, we do estimate it as it has been previously used by both researchers and policy as a measure of principal effectiveness,[18] and we provide further evidence on how this specification diverges from more preferred specification described below.

$$Y_{ispt} = \alpha_0 Y_{it-1} + \alpha_1 X_{it} + \alpha_2 \bar{S}_{it} + \delta_{ps} + e_{ispt} \tag{2}$$

Where, similar to teacher value-added models discussed above, Y_{isjt} represents student i's test score in a given subject, for school s, under principal p in year t. The first term on the right, Y_{it-1}, is a vector of prior test scores in math and reading for student i, specified as a cubic polynomial. X_{it} represents a vector of student-level controls such as gender, ethnicity, and participation in FRL, special education services, and LEP programs, indicators for grade and school year, $\bar{S}_{it}$, which represents school-average characteristics, and e_{ispt} is a mean-zero error term.[19] [20] The coefficient of interest is δ_p, the estimated principal value-added score for principal p. This model will attribute all achievement gains (that are not explained by student covariates X_{it}) which are common across students in a school to principal p.[21]

A shortcoming of the above approaches is that they attribute adjusted student achievement gains within schools to principals (Grissom et al., 2014). Clearly many school characteristics are not under the control of principals, particularly, newly hired principals. For example, previous research suggests that time spent on teacher selection is associated with improved student outcomes, but most principals are not responsible for hiring most of their teaching staff; instead, they inherit teachers selected by previous principals. Thus, as Chiang et al. (2016) find, school value-added measures provide poor estimates of a principals' persistent effectiveness.

Next, we describe our preferred approach for estimating principal value added that uses within-school variation in achievement, as introduced in recent work by Austin et al. (2019). We begin by estimating models as described in Equation (2) to store δ_{ps}, which is our estimate of principal-by-school fixed effects. This contains information about principal value added, but is likely confounded by the issues discussed above. Consistent with Austin et al. (2019), we attempt to remove the influences of fixed school factors by demeaning δ_{ps} within schools using the school-average value of PVA, $\overline{\delta_s} = \sum_{p=1}^{P_s} \pi_p \delta_{ps}$, where π_p is the ratio of years principal p

[18] For example, see a discussion of this issue in (see Chiang, Lipscomb, & Gill, 2016).
[19] Research by Altonji and Mansfield (2018) suggests that school-averages may control for sorting on unobservable characteristics of students and schools in some settings.
[20] Like TVA models, we estimate PVA models separately by grade span, K-8 and high school. See footnote 12 for more discussion.
[21] We attempt to address these concerns by also estimating "initial" principal effectiveness, limited to each principal's first year of employment, and compare results across principals with similar experience.

leads school s to the total number of years school s appears in the data panel, and is the number of principals who served at school s over the course of the data panel. As such, our estimate of within-school principal value added is $\delta_{ps} - \overline{\delta_s}$. Note that our focus on principal by school fixed effects differs from studies such as Grissom et al. (2015) and Chiang et al. (2016), where principal value added is calculated using models that include school fixed effects. We estimate similar models and report the results in Appendix A. These estimates are qualitatively similar to the approach we consider above, though school fixed effects models produce less precise estimates.

These models account for the time-invariant unobservable characteristics of schools that could bias principal value-added scores. An obvious limitation of the within school estimate of principal effectiveness is that it, by definition, ignores differences in principal effectiveness across schools. This may mask an important component of the variance of principal effects, if, for instance, schools typically hire principals from the same strata of the principal performance distribution, but that strata differs between schools. Moreover, within school models can only estimate effects when principals can be compared to other individuals within the same school, so that these estimates depend on the amount of overlap in our sample. This excludes 8% to 21% of principal observations depending on the specification. Also, importantly, these models will not address time-*varying* unobservable characteristics of schools; as cautioned by Austin et al. (2019), complex dynamics and contributions of other factors challenges the interpretation of within-school differences as measures of differences in principal effectiveness.

But it is also the case that within school estimates of principal effectiveness do not necessarily guarantee the recovery of the within school variation in principal performance. One issue is that the estimates may capture transitory fluctuations in student achievement that should not be attributed to the principal. Miller (2013), for instance, finds that principal turnover is preceded by declines in student achievement, so newly hired principals may appear to rapidly improve in their effectiveness due to mean reversion. Moreover, many features of the school are fixed if a principal is hired in the fall (e.g., teaching staff, curriculum, assignments), so there may be relatively little malleable factors for the principal to control. We consider this possibility by estimating models that exclude the first year of a principal’s tenure. But note that exclusion of the first year does not necessarily fully address the more fundamental issue of the potential misattribution of one principal’s influence on a school to the principal who follows him or her. Thus, we return to this problem in the robustness section below.

An issue that arises specifically in our context, where we are seeking to link TVA and PVA, is that students could contribute both to the estimate of teacher and principal value added, creating a mechanical correlation between the two. In practice, this is not likely to be a concern because most staff take several years to transition between teaching positions and principal positions. About 50 percent take 3 years or longer. Nevertheless, we consider specifications that employ a jackknife procedure where value-added models are estimated using non-overlapping student test scores from different time periods.[22]

3.3 Estimating Associations Between Teacher and Principal Value Added

The general model we utilize to relate PVA to TVA, or to teachers not having estimates of value added given their prior teaching assignments, is:[23]

$$PVA_i = \gamma_1 NoTVA_i + \gamma_2 TVA_i + \gamma_3 TchChar_i + u_i \quad (4)$$

where PVA_i and TVA_i are value-added estimates for individual i, $NoTVA_i$ indicates that the individual did not teach in a tested grade or subject, and u_i is a mean-zero error term. The key variable of interest is γ_2, which estimates the associated change in PVA for a one unit increase in TVA, and positive estimates indicate that high TVA individuals tend to have high PVA.[24] We are also interested in γ_1, which estimates the average effectiveness of teachers who do not teach tested grades and subjects.[25] We interpret NoTVA as indicating that teachers are likely to have less direct exposure to accountability because they do not teach a core tested academic subject. It is important to note, however, that not observing TVA is possible for several reasons. First, only teachers who serve between 2007-08 and 2016-17 in tested grades and subjects can have TVA scores, so some teachers may teach tested grades and subjects prior to 2007-08 and not have TVA. Second, we censor TVA for teachers who are linked to fewer than 10 students, and these teachers clearly have experience in tested grades and subjects.

In some specifications, we include $TchChar_i$, which is a vector of individual characteristics, to explore whether more effective teachers become more effective principals *over and above the influence of their personal characteristics*. We include indicators for gender, race/ethnicity, WEST-B licensure test scores, and education and experience prior to becoming a principal. As discussed above in Section 3, we also include a variable for the number of years between the last teaching position and the first principalship to account for potential changes in the relationship between TVA and PVA due to TVA drift or skill loss.

In this model, identification comes from cross-sectional variation in TVA across principals, and as such, there are several challenges when estimating Equation (4). One concern is the potential that unobserved school or district factors influence both the effectiveness of

[22] As noted by Chetty et al., (2014a) jackknife procedures are important in their analysis of TVA so that estimation errors do not appear on both the left and right side of the regression equation.
[23] In Washington state, given the state's testing regime, the TVA model specifications we describe in Section 3.1 can be used to estimate TVA in grades 4-8, which covers about 16 percent of the state's teacher workforce.
[24] While TVA estimates are clustered at the classroom level to reflect correlated errors among students, we do not cluster standard errors in or PVA models because we do not use adjustments to PVA, such as the EB adjustment for TVA, or use estimated standard errors for PVA.
[25] Individuals with missing values for TVA have scores that are imputed to 0, commonly referred to as the "dummy variable method." Research by Abrevaya and Donald (2013) indicates that such models require assumptions in addition to "missing at random"; as such, we have also estimated models where we restrict our sample to individuals with TVA and find very similar results, which are available on request.

teachers and the effectiveness of principals, leading to bias in estimates of the relationship of the two.[26] While one may attempt to address this by estimating models that control for school or district fixed effects, this is not necessary as some specifications of the PVA model (Equation (3) discussed above) are based on within-school estimates of principal effectiveness that should purge any school level, time-invariant factors from the PVA estimate.[27]

Another concern is non-random selection of principals (from a policy perspective, we would hope that schools and districts are able to non-randomly select more effective principals). This is a concern because we only observe PVA for individuals who are hired as a principal. If the propensity to be hired as a principal is correlated with principal value added as well as unobserved individual characteristics that influence teacher value added, then our findings may suffer from selection bias. For example, a principal may have low teacher value added and be hired regardless because, conditional on their low TVA, they have good organizational management skills, which may attenuate our estimates. We assess the degree to which this is an issue by estimating the propensity of observing teachers as principals, which allows us to sign the likely direction of bias in the relationship between teacher and principal value added.

Finally, research on managerial promotions in the private sector suggests that relatively greater uncertainty about the likely future productivity of managerial candidates external to the organization influences the types of internal and external hires. Specifically, hiring officials are likely to have less first-hand information about an external candidate's productivity, hence risk-averse organizations should seek a "compensating premium" of qualifications thought to be predictive of future performance. Consistent with this idea, research (Morita & Tang, 2019; DeVaro, Kauhanen, & Valmari, 2019) finds that individuals who are hired internally tend to have lower qualifications, such as prior experience or education, relative to external hires. The fact that internal and external candidates tend to have different observable characteristics might suggest that they also vary along *unobservable* dimensions (e.g., hiring of external candidates with more motivation).

In addition to the managerial ability level of internal and external principal candidates, it is possible that familiarity with context could play a role in principal success through job matching effects (there is an extensive literature on firm specific human capital; for examples, see Parsons, 1972; Hashimoto, 1981; Neal, 1995; Lazear, 2009). Principals who have previously worked as a teacher within a school or district may be more familiar with the needs of students and staff, and thus, they may be more effective than someone hired outside the school. Alternatively, internally hired principals may face challenges when managing former peers if

[26] For example, suppose that high-performing districts have better hiring practices (e.g. HR districts, hiring committees), which lead to the hiring of more effective principals as well as more effective teachers. As such, a naive comparison would suggest that teacher value-added is positively correlated to principal value added while the relationship is driven by district factors.

[27] Even with the inclusion of school fixed effects in Equation (3), school and district factors could still bias estimates. For example, schools may have strong trends in effectiveness over time, perhaps due to consecutive year shocks, and principals who are promoted within these schools will have higher teacher value added as well as higher principal value added. To investigate this, we consider specifications limited to principals who are hired from different schools or districts.

these individuals do not view the principal as a school leader, or the principal is less able to make changes that affect his or her prior teacher peers. For all these reasons we explore models that allow for differential relationships between PVA and TVA of internal and externally hired principals.

3.4. Washington State Data on Teachers and Principals

Our analysis of Washington State teachers and principals leverages three administrative data sets. The first is the S-275 personnel reporting system, maintained by the Washington State office of the Superintendent for Public Instruction (OSPI). This data includes detained records on employee demographics (such as overall experience, gender, race/ethnicity, and education level). And for a subset of individuals (those who became teachers after 2002-03), we observe their scores on basic skills licensure tests in math and reading, known as the "WEST-B." Key to our study, the S-275 also includes information on whether individuals are working as teachers, principals or in other administrative positions in public schools, location, and full-time equivalency, as well as a unique certification ID number which can be used to track individuals over time and link to other administrative data. This information allows us to identify teachers, principals, and individuals who transition from teaching to the principalship. The S-275 is uniquely suited for this study because the data cover a long period of time, from 1983-84 to 2016-17.

Test scores used in the estimation of both TVA and PVA come from two administrative data sets. The Core Student Records System (CSRS) reports student test scores on state tests from 2006-07 to 2009-10, and we use data on exam proctors to match students to teachers for this period, and unique school and district IDs to match students to principals.[28] Due to the nature of this match, we only estimate TVA for elementary and middle school teachers. From 2009-10 to 2016-17, the Comprehensive Education Data and Research System (CEDARS) is used to follow students over time. This includes information on course assignments, and teacher files that allow us to create student-teacher links.[29] Similar to CSRS data, we use unique school and district IDs to match students to principals.

We impose several restrictions on our sample of teachers and principals. We define teacher and principal positions as having at least 0.5 FTE, and do not consider principals who are employed in multiple schools.[30] Given our focus on teacher characteristics, we also restrict our focus to individuals who are observed working as a teacher at some point in the S-275 data. A small number of individuals are observed working as teachers before and after their first

[28] The proctor of the state assessment was used as the teacher-student link for at least some of the data used for analysis. The 'proctor' variable was not intended to be a link between students and their classroom teachers, so this link may not accurately identify those classroom teachers.
[29] CEDARS data includes fields designed to link students to their individual teachers, based on reported schedules. However, limitations of reporting standards and practices across the state may result in ambiguities or inaccuracies around these links.
[30] We include settings where two principals work in the same school and we apply the same PVA estimate to both individuals.

principalship position (about 0.6%); we avoid conflating changes in TVA due to prior experience as a principal by only considering an individual's "first spell" of teaching prior to their first principal position.

We cannot determine whether teachers are applying for positions as principals, and, if so, receiving job offers; we only observe employment status, which in the case of principal positions would mean that a match occurred indicating an application, a job offer, and an acceptance of that offer. In total, we have a 11-year panel (SY 2006-7 to 2016-17) for which we observe this match, and based on the above restrictions, the analytic dataset we utilize includes 1,708,548 teacher-year observations (154,464 unique teachers), 55,531 principal-year observations (7,429 unique principals), and 80,522 individual-year observations (3,102 unique individuals) of employees who work in both roles at some point in their career.

Table 1 presents sample statistics for four distinct subsamples of teachers: according to whether teachers are at some point in the data observed as principals, and whether they have TVA estimates.[31] Observations are unique at the individual teacher level, and all characteristics reflect the last year of teaching. The tests of significance are for teachers in each subsample (value added or not) who we do or do not observe as principals (i.e., the means of column 1 vs. column 2, and then the means of column 3 vs. column 4).

As mentioned above, the majority of teachers are not observed as principals and do not have value-added estimates (Column (1), about 83% of all teachers). Like most teaching populations, individuals in the sample tend to be female, at about 70% of the sample. A very high percentage of teachers are white, about 90%.[32] About half have either a bachelor's or master's degree, and less than 1% have a Ph.D. WEST-B licensure tests are slightly below average (normalized to zero across the sample of WEST-B takers), though the differences are not statistically significantly different from any other group represented in the table.

Next, we compare teachers *without* value added who are and are not observed as principals (Column 1 vs. 2). Future principals are far less likely to be female, 52% relative to 70% of those who are not observed as principals. There are also large differences in degree attainment for future principals, with 82% attaining a master's or higher degree prior to exiting teaching (which is required for the principalship), compared to 55% of other teachers. Principals tend to have higher licensure tests, by 11-13% of a standard deviation, then those not observed as principals in the non-TVA sample, though these differences are not statistically significant. Lastly, the localness of principal labor markets is demonstrated by the fact that about half of principals in this subsample have prior experience teaching in the same district they become principal, and about 18% have prior experience teaching in the same school.

[31] The TVA estimates in the table are based on the TVA specification that average math and reading and do not include classroom fixed effects. However, as we report below, correlations within subject are very high across model specification.

[32] Washington state has only a small percentage of non-white teachers, roughly 13%, about half of whom are underrepresented minority teachers (for more discussion, see Goldhaber, Theobald, & Tien, 2018).

Comparisons between teachers with value added who are not observed as principals (Column 3) to those who are observed as principals (Column 4) show many similar patterns to the non-value added sample comparisons. In particular, teachers observed as principals are far less likely to be female, far more likely to have an advanced degree, and we again see that many principals are hired from the same district or school in which they taught. By contrast, in the value-added sample, teachers who become principals have *lower* licensure test scores in both math and reading, however these results are not statistically significant. Teachers in the value-added sample who we observe as principals have slightly higher value added, but the difference between the two groups is not statistically significant, suggesting that principals are not systematically selected according to their prior effectiveness when serving as a classroom teacher.[33]

3.5. Correlations between different TVA and PVA specifications

In Table 2 we report correlations across different value-added specifications for teachers. The correlations between the different TVA specifications: math, reading, with and without classroom covariates. This sample is limited to 17,506 teachers who have estimates for all eight TVA specifications. Consistent with the existing literature (Aaronson, Barrow, & Sander, 2007; Goldhaber et al., 2012; Ehlert et al., 2014; McCaffrey et al., 2004), we find that the inclusion of different types of covariates has little impact on the within subject correlations; for example, the correlation coefficient for both math and reading TVA with and without classroom covariates is about 0.98. Given the high correlation between TVA models with and without classroom covariates, we only present findings for models that exclude classroom covariates, but consistent with the high correlation between the two specifications, the findings are quite consistent regardless of which TVA specification is utilized.[34] The correlations across subject are notably smaller, about 0.60, but these too are in the same neighborhood of what has been previously found for (unadjusted for sampling error) TVA (e.g., Koedel & Betts, 2007; Loeb, Kalgorides, & Beteille, 2012; Teh, Resch, Walsh, Isenberg, & Hock, 2013; Value-Added Research Center, 2010).[35]

Next, in Table 3, we present correlations between different specifications of PVA. The sample is limited to 2,464 principals who have estimates for all eight specifications: math and reading; specifications with and without school demeaning; and specifications that do or do not drop the first year of the principalship.[36] Starting with models that include all principal observations, the *cross-subject correlations within specification* correlations are about 0.40 to 0.50. Dropping the first principal year leads to lower cross-subject correlations, as low as 0.14

---

[33] In Table 1 we report TVA that is based on a teacher's full career (before a first principalship). But we also find little difference in TVA across categories when we use a teacher's first value-added estimate in the comparison.
[34] Results for models with classroom covariates are available on request.
[35] Goldhaber, Cowan, and Walsh (2013) note that TVA is measured with sampling error, which will cause raw correlations to understate the true correlation of TVA estimates across models. They find substantially higher correlations of 0.7 to 0.8.
[36] The distribution of PVA estimates range from 0.13 to 0.16 SDs for both school value-added models and within-school models. This varies substantially from models with school fixed effects reported in Appendix A, at 0.24 to 0.34 SDs for. While school value-added models and within-school models are comparable to estimates from prior studies, school-fixed effect estimates are considerably larger.

when comparing the school value-added model for math and reading. The *within subject across specification* correlations are somewhat less highly correlated than the TVA specifications reported in Table 2; specification choice is thus more likely to have important implications for estimating the relationship with TVA. For instance, math models with and without demeaning have correlations of 0.80 to 0.94 depending on whether the correlation is for math or reading and whether a principal's first year is included in the PVA estimate. Given these differences, we report results for all PVA specifications and discuss how results vary across models.

4. Results

4.1. Estimating the relationship between different TVA and PVA specifications

In this subsection, we present our findings on the relationship between TVA and PVA. Table 4 reports the OLS regressions of Equation (4) for TVA and PVA for principal math value added (Panel A) and reading value added (Panel B). Each column represents a model with a particular specification of PVA: school value-added models over all principal observations (Column 1); models that exclude a principal's 1st year at a school (Column 2); models for within school PVA (Column 3); models that include school fixed effects but exclude a principal's 1st year at a school (Column 4); and parallel specifications that include teacher covariates (Columns 5-8). PVA is measured in student level standard deviations.[37]

We begin by focusing on PVA math in Panel A. The main coefficient of interest is the point estimate on TVA, which represents the change in PVA for a one-standard deviation change in TVA.[38] The coefficient on TVA in math is consistently positive, though it is not statistically significant for some of the PVA specifications. It is worth noting that all models suggest fairly similar point estimates between 1 to 2 percent of a standard deviation. The coefficient on No TVA Observed is negative and significant in the PVA specifications that do not include school fixed effects; the point estimates suggest these principals have lower math PVA by 1 to 5 percent of a standard deviation of student achievement. The coefficients in models with school fixed effects are also negative, but imprecisely estimated. And, there is almost no difference in the estimated coefficients in the specifications that exclude teacher covariates (columns 1-4) and those that include teacher covariates (columns 5-8), which is not terribly surprising given the literature showing a relatively weak relationship between teacher covariates and value added (Aaronson, Barrow, & Sander, 2007; Goldhaber et al., 2012; Ehlert et al., 2014; McCaffrey et al., 2004).

[37] And while not reported, it is worth noting that, in both math and reading, the number of years separating TVA and PVA appears to have very little impact on our estimates; the coefficient on this variable is not statistically significant, consistent in sign, and it tends to be close to zero.
[38] We also estimate models with only "own subject" TVA in each model. For math PVA, we find that the coefficients on TVA math are significant for specifications (2) and (6), which are also significant for reading in Table 4. Results for reading PVA are very similar to those reported in Table 4, Panel B. These are available on request.

Panel B reports parallel specifications for reading PVA. In contrast to PVA math, there is much stronger evidence that teacher value added predicts principal value added and little evidence that whether a teacher has a value-added estimate is predictive of PVA reading. We find that a one standard deviation increase in TVA is associated with an increase in PVA of 1 to 2 percent of a student-level standard deviation. Across all specifications, a teacher's TVA in reading is positively related to their value added in reading, though is not statistically significant (at the 95% confidence level) in specifications that use within-school PVA and exclude the principal's first year in the school.

In Appendix Table B.2, we report results that include both TVA math and TVA reading in the same estimation equation. These results suggest that, for some specifications, TVA in reading appears to predict PVA in math. This could imply two possibilities. First, this is consistent with the idea that TVA is picking up some other school contextual factors that are biasing PVA estimates (e.g., some districts may have better HR departments that hire better teachers and principals). Second, TVA reading may simply be more important for determining the success of the principal. One example is that principals with higher TVA in reading may have better communication skills in general, and they are better able to serve both math and reading teachers relative to principals with higher TVA in math.[39]

4.2. Internal and External Hires

Next, we estimate models, similar to those presented in Table 4, that include indicators for whether the principal is hired externally (with no teaching experience in the school or district in which a principal is employed), hired internally to the district *but not the principal's school* (any teaching experience within the district), or hired internally to the school (any teaching experience within the same school).[40] Table 5 presents results for different specifications of PVA in math and reading, as well as interactions between hiring type and TVA in math and reading.[41]

We start by discussing the indicators for the type of hire where the omitted group is external hires. In math (Panel A), the coefficient on internal to the district is positive and significant for the school value-added models, but it is not significant in our preferred models that utilize within school PVA estimates. But, the coefficient on internal to the school is significant and negative across all math specifications. This is contrary to expectations about specific human capital effects associated with having direct knowledge about a school having served as a teacher in that school prior to assuming the principalship. Note also that the strength

[39] As noted above, this difference in predictive power is interesting because prior research on TVA suggests that there are relatively high correlations between TVA subjects (e.g. Koedel and Betts, 2007; Loeb, Kalgorides, & Beteille, 2012; Teh, Resch, Walsh, Isenberg, & Hock, 2013; Value-Added Research Center, 2010). Other work by Goldhaber and Hansen (2010) suggests that TVA in math predicts future student performance in reading, but not vice versa.

[40] Internal to district is interpreted as not internal to school because we include indicators for both types of internal hires, and all hires that are internal to the school are also internal to the district.

[41] We do not report specifications that include teacher characteristics as these have very little effect on the coefficient of TVA, but these findings are available from the authors upon request.

of the relationship between TVA in math and PVA in math is stronger for principals who are hired externally.

The PVA in reading models (Panel B) we find that principals hired from outside a district are not found to have different PVA than those hired from inside the district by from a different school. And, consistent with the PVA and math results, there is some evidence that teachers who become principals in a school where they once taught tend to have lower PVA (though these findings are only marginally significant in one of the within school PVA models). And, again similar to the math findings, the strength of the relationship between TVA and PVA in reading is stronger for external hires.

4.3 Robustness Checks

In this subsection we address two issues: 1) the extent to which the relationship between PVA and TVA is likely to be affected by sample selection; and 2) whether the predictive power of TVA is influenced by the proximity of the TVA estimates to an individual's tenure as a principal; 3) the degree to which censoring of teacher observations could cause bias in the relationship between PVA and TVA.

There is scant research examining the probability that teachers move into the principalship. Thus, this line of inquiry is interesting in general, however, we are particularly concerned about the degree to which TVA appears to predict the likelihood of observing teachers as principals. We follow Brewer (1996) and estimate probit regression models of the following form:

$$probit(1(hired_t)) = \alpha_0 + \alpha_1 TVA + \alpha_2 T_t + u \tag{5}$$

where $hired_t$ indicates that a former teacher is observed as a principal in year t. T represents a vector of teacher characteristics in year t. Like other studies of promotions (e.g., Brewer, 1996; for research from the private sector, see DeVaro, Kauhanen, & Valmari, 2019), we include variables for gender, ethnicity, educational attainment, experience in teaching, and school year indicators, and in some specifications, TVA.[42]

In Appendix Table B.1 we present selected coefficients (the marginal effects) of teacher characteristics on the likelihood of observing a teacher as a principal. The first column presents results for all teachers in our sample to give a broad sense of selection into the principalship, and in the second column we focus on the subsample of teachers for whom TVA is available.

We first focus on the full sample of teachers (column 1). Perhaps the most striking finding is the significant and large discrepancy in the likelihood that female teachers are observed as principals: they are about 1 percentage point less likely to be observed in the principalship than male teachers, conditional on other attributes. Given that only about 3 percent of teachers are observed as principals, this represents a gender difference in the likelihood of becoming a principal of about 33 percent, which is roughly comparable to findings reported by

[42] Standard errors are clustered at the individual teacher level.

Brewer (1996) that only 26 percent of principals and assistant principals are female relative to 56 percent for teachers.[43] Given that we only observe an individual in the principalship if he or she applies, is selected, and accepts the job, this finding does not necessarily indicate discrimination, but it does suggest that more work on this topic is needed. While not reported, we also find an "inverted u" pattern in the relationship between teacher experience and the likelihood of observing teachers as principals, this is consistent with findings from Brewer (1996), where teachers are initially less likely to be employed as principals, more likely with around 6 to 8 years of experience teaching, and less likely afterward.

Next, we consider the second column which presents results for individuals with TVA. The findings are generally consistent with the full sample (in column 1). And, importantly, there is little evidence that a teacher's specific TVA is associated with the likelihood of being observed as a principal. The estimated coefficients on math and reading TVA are all quite small, statistically insignificant, and precisely estimated. Moreover, both coefficients are positive, which if anything, suggests that the relationship between PVA and TVA (presented in Table 4) is a lower bound.

Second, we address the concern that the career estimates of TVA that we utilize in the models in Tables 4 and 5 could mask the ability of TVA to predict PVA. Recall that the estimates of TVA used in Table 4 were based on as many years of matched student and teacher data that were available in our data. There is evidence that much of a teacher's value added is fixed over the course of that teacher's career (Atteberry et al., 2015; Goldhaber and Hansen, 2013), nevertheless, we assess the possibility that TVA estimates more proximal to a principal's tenure are more predictive by estimating teacher-year specifications of equation (1) and using the year of TVA most proximal to the time that teachers assumed a principalship in estimating PVA.

As predicted above, these results are much less precise relative to TVA calculated over a teacher's career.[44] That said, they are positive for all specifications. Results for PVA math are very consistent with point estimates in Table 4, and like our previous results, they are statistically significant for specifications that exclude the principal's first year. Results for reading are somewhat consistent, but considerably less precise, with only marginally significant coefficients that are closer to zero.

5. Policy Implications and Conclusions

In this study, we provide a first look at whether value-added effectiveness of teachers is predictive of principal value added. Prior to focusing on the implications of our primary focus, several ancillary findings are worth emphasizing. First, our findings highlight the sensitivity of principal value-added estimates to model specification choices. While this is merely a replication of prior findings in a new context, it is an important policy consideration.

[43] Gates et al. (2006) also find significant gender disparities in who becomes principals.
[44] We do not report these results because they are qualitatively similar to those in Table 4, but they are available on request.

Second, we find evidence that principal labor markets also appear to be segmented and localized. This is similar to findings in teacher labor markets, but had not previously been documented for principals.

Third, while the teaching profession is predominately female, all else equal, principals are far less likely to be female. There is an extensive literature that explores gender-based labor market discrimination in the promotion to management positions (e.g., Bertrand et al., 2001), but little evidence on the degree to which gender may influence promotion into managerial positions in the public sector. Thus, the finding for the gender disparity in the likelihood of becoming a principal merits further exploration.

In terms of the primary focus of the paper, we find evidence that value-added measures of teacher effectiveness are predictive of value-added measures of principals. Yet there is little evidence that a teacher's value added is considered when it comes to making decisions about who should serve in the principalship. This suggests policymakers have considerable opportunities to improve the effectiveness of the principal workforce through more purposeful selection of teachers according to their value added. That said, one should be careful interpreting these results because our estimates are not significant in all specifications. This is likely due to small sample sizes, as relatively few principals can be linked to both measures of value added. Moreover, more work is needed to validate principal value added as a measure that can remove school contextual factors from the influence of the principal.

Lastly, a literature on promotions in the private sector shows that external candidates promoted to management positions have higher qualifications than internal candidates. We contribute to this broader literature in our investigation of promotions in the public sector. More specifically, we provide the first evidence on the performance of internal and external candidates, finding that principals that were promoted from their school's teacher workforce tend to be less effective. This too may have important policy implications, as hiring officials are likely to consider the value of the knowledge that internal candidates have about their schools, but this knowledge may come at the cost of having less flexibility to make changes once the candidate moves into a position of school leadership.

References

Aaronson, D., Barrow, L. and Sander, W., 2007. Teachers and student achievement in the Chicago public high schools. *Journal of labor Economics*, *25*(1), pp.95-135.

Abdulkadiroğlu, A., Angrist, J. and Pathak, P., 2014. The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica*, *82*(1), pp.137-196.

Abrevaya, J. and Donald, S.G., 2017. A GMM Approach for Dealing with Missing Data on Regressors. *Review of Economics and Statistics*, *99*(4), pp.657-662.

Altonji, J.G. and Mansfield, R.K., 2018. Estimating group effects using averages of observables to control for sorting on unobservables: School and neighborhood effects. *American Economic Review*, *108*(10), pp.2902-46.

Atteberry, A., Loeb, S. and Wyckoff, J., 2015. Do first impressions matter? Predicting early career teacher effectiveness. *AERA Open*, *1*(4), p.2332858415607834.

Austin et al. (2019). Path to the Principalship and Value Added: A Cross-state Comparison of Elementary and Middle School Principals. Unpublished working paper.

Baker, G., Gibbs, M. and Holmstrom, B., 1994. The internal economics of the firm: evidence from personnel data. *The Quarterly Journal of Economics*, *109*(4), pp.881-919.

Balu, R., Horng, E.L. and Loeb, S., 2010. Strategic personnel management: How school principals recruit, retain, develop and remove teachers. *School Leadership Research, Working Paper*, pp.10-6.

Bottoms, G. and O'Neill, K., 2001. *Preparing a new breed of school principals: It's time for action*. Atlanta, GA: Southern Regional Education Board.

Boyd, D., Grossman, P., Ing, M., Lankford, H., Loeb, S. and Wyckoff, J., 2011. The influence of school administrators on teacher retention decisions. *American Educational Research Journal*, *48*(2), pp.303-333.

Branch, G., Hanushek, E., & Rivkin, S. (2009). *Estimating principal effectiveness* (Working Paper 32). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Brewer, D. J. (1993). Principals and student outcomes: Evidence from U.S. high schools. *Economics of Education Review*, *12*(4), 281–292.

Brewer, D.J., 1996. Career paths and quit decisions: Evidence from teaching. *Journal of Labor Economics*, *14*(2), pp.313-339.

Chetty, R., Friedman, J.N. and Rockoff, J.E., 2014. Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), pp.2593-2632.

Chiang, H., Lipscomb, S., & Gill, B. (2016). Is school value added indicative of principal quality? *Education Finance and Policy*, *11*(3), 283–309.

Chingos, M.M. and Peterson, P.E., 2011. It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review*, *30*(3), pp.449-465.

Clark, D., Martorell, P., & Rockoff, J. (2009). *School principals and school performance* (Working Paper 38). Washington, DC: National Center for Analysis of longitudinal data in Education research.

Cooley, V.E. and Shen, J., 2000. Factors influencing applying for urban principalship. *Education and Urban Society*, *32*(4), pp.443-454.

Dhuey, E. and Smith, J., 2014. How important are school principals in the production of student achievement? *Canadian Journal of Economics/Revue canadienne d'économique*, *47*(2), pp.634-663.

Dhuey, E., & Smith, J. (2018). How school principals influence student learning. *Empirical Economics*, *54*(2), 851–882.

Ehlert, M., Koedel, C., Parsons, E. and Podgursky, M.J., 2014. The sensitivity of value-added estimates to specification adjustments: Evidence from school-and teacher-level models in Missouri. *Statistics and Public Policy*, *1*(1), pp.19-27.

Fenwick, L.T. and Pierce, M.C., 2001. The principal shortage: crisis or opportunity? *Principal*, *80*(4).

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

Glickman, C.D., Gordon, S.P. and Ross-Gordon, J.M., 2005. *The basic guide to supervision and instructional leadership*. Pearson/Allyn & Bacon.

Goldhaber, D. and Chaplin, D.D., 2015. Assessing the "Rothstein Falsification Test": Does it really show teacher value-added models are biased? *Journal of Research on Educational Effectiveness*, *8*(1), pp.8-34.

Goldhaber, D. and Hansen, M., 2013. Is it just a bad class? Assessing the long-term stability of estimated teacher performance. *Economica*, *80*(319), pp.589-612.

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, *22*(2), 129–145.

Goldhaber, D., Cowan, J. and Walch, J., 2013. Is a good elementary teacher always good? Assessing teacher performance estimates across subjects. *Economics of Education Review*, *36*, pp.216-228.

Goldhaber, D., Grout, C. and Huntington-Klein, N., 2017. Screen twice, cut once: Assessing the predictive validity of applicant selection tools. *Education Finance and Policy*, *12*(2), pp.197-223.

Goldhaber, D., Krieg, J. and Theobald, R., 2018. *Exploring the impact of student teaching apprenticeships on student achievement and mentor teachers*. CALDER Working Paper 207-1118-1.

Goldhaber, D., Quince, V. and Theobald, R., 2016. Reconciling different estimates of teacher quality gaps based on value-added. *Washington, DC: National Center for the Analysis of Longitudinal Data in Education Research (CALDER), American Institutes for Research*.

Goldhaber, D.D. and Brewer, D.J., 2000. Does teacher certification matter? High school teacher certification status and student achievement. *Educational evaluation and policy analysis*, *22*(2), pp.129-145.

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

Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal*, *48*(5), 1091–1123.

Grissom, J. A., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, *37*(1), 3–28.

Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher*, *42*(8), 433–444.

Grissom, J.A. and Bartanen, B., 2018. Strategic Retention: Principal Effectiveness and Teacher Turnover in Multiple-Measure Teacher Evaluation Systems. *American Educational Research Journal*, p.0002831218797931.

Grissom, J.A., 2011. Can good principals keep teachers in disadvantaged schools? Linking principal effectiveness to teacher satisfaction and turnover in hard-to-staff environments. *Teachers College Record*, *113*(11), pp.2552-2585.

Grissom, J.A., Mitani, H. and Woo, D.S., 2018. Principal Preparation Programs and Principal Outcomes. *Educational Administration Quarterly*, p.0013161X18785865.

Guarino, C.M., Reckase, M.D. and Wooldridge, J.M., 2015. Can value-added measures of teacher performance be trusted? *Education Finance and Policy*, *10*(1), pp.117-156.

Hallinger, P. and Heck, R.H., 2011. Conceptual and methodological issues in studying school leadership effects as a reciprocal process. *School Effectiveness and School Improvement*, *22*(2), pp.149-173.

Hallinger, P. and Murphy, J., 1987. Assessing and developing principal instructional leadership. *Educational leadership, 45*(1), pp.54-61.

Hallinger, P., & Heck, R. H. 1996. Reassessing the principal's role in school effectiveness: A review of empirical research, 1980-1995. *Educational Administration Quarterly*, *32*(1), 5–44.

Harris, A., Muijs, D. and Crawford, M., 2003. Deputy and assistant heads: Building leadership potential. *National College for School Leadership, University of Warwick, Coventry*.

Hashimoto, M., 1981. Firm-specific human capital as a shared investment. *The American Economic Review*, *71*(3), pp.475-482.

Heck, R.H., 1992. Principals' instructional leadership and school performance: Implications for policy development. *Educational evaluation and policy analysis, 14*(1), pp.21-34.

Heckman, J.J., Stixrud, J. and Urzua, S., 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, *24*(3), pp.411-482.

Herrmann, M., Walsh, E., Isenberg, E. and Resch, A., 2013. WORKING.

Hitt, C., McShane, M.Q. and Wolf, P.J., 2018. Do impacts on test scores even matter? Lessons from long-run outcomes in school choice research. *American Enterprise Institute. Retrieved from http://www.aei.org/publication/do-impacts-on-test-scores-even-matterlessons-from-long-run-outcomes-in-school-choice-research*.

Horng, E. L., Klasik, D., & Loeb, S. (2010). Principal's time use and school effectiveness. *American Journal of Education*, *116*(4), 491–523.

Hoy, A.W., Hoy, W.K. and Hoy, A.W., 2003. Instructional leadership: A learning-centered guide.

Isenberg, E., & Walsh, E. (2014). Measuring school and teacher value added in DC, 2012-2013 school year: Final report. Policy Report, Mathematica Policy Research (01.17.2014).

Isenberg, E., Max, J., Gleason, P., Johnson, M., Deutsch, J., & Hansen, M. (2016). Do low-income students have equal access to effective teachers? Evidence from 26 districts. NCEE 2017-4007. Institute of Education Sciences.

Jackson, C.K., 2018. What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, *126*(5), pp.2072-2107.

Jacob, R., Goddard, R., Kim, M., Miller, R. and Goddard, Y., 2015. Exploring the causal impact of the McREL Balanced Leadership Program on leadership, principal efficacy, instructional climate, educator turnover, and student achievement. *Educational Evaluation and Policy Analysis, 37*(3), pp.314-332.

Kane, T.J., McCaffrey, D.F., Miller, T. and Staiger, D.O., 2013. Have we identified effective teachers? Validating measures of effective teaching using random assignment. In *Research Paper. MET Project. Bill & Melinda Gates Foundation*.

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

Koedel, C. and Betts, J.R., 2011. Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education Finance and policy*, *6*(1), pp.18-42.

Koedel, C., Mihaly, K. and Rockoff, J.E., 2015. Value-added modeling: A review. *Economics of Education Review*, *47*, pp.180-195.

Krieg, J.M., Theobald, R. and Goldhaber, D., 2016. A foot in the door: Exploring the role of student teaching assignments in teachers' initial job placements. *Educational Evaluation and Policy Analysis*, *38*(2), pp.364-388.

Ladd, H.F., 2011. Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement? *Educational Evaluation and Policy Analysis*, *33*(2), pp.235-261.

Lankford, H., O'Connell, R. and Wyckoff, J., 2003. Identifying the next generation of school leaders. *Center for Policy Research, University at Albany, SUNY*.

Lazear, E.P., 2012. Leadership: A personnel economics approach. *Labour Economics*, *19*(1), pp.92-101.

Lazear, E.P., Shaw, K.L. and Stanton, C.T., 2015. The value of bosses. *Journal of Labor Economics*, *33*(4), pp.823-861.

Loeb, S., Kalogrides, D. and Horng, E.L., 2010. Principal preferences and the uneven distribution of principals across schools. *Educational Evaluation and Policy Analysis*, *32*(2), pp.205-229.

Malone, B.G. and Caddell, T.A., 2000. A crisis in leadership: Where are tomorrow's principals? *The Clearing House*, *73*(3), pp.162-164.

March, J.G., 1978. American public school administration: A short analysis. *The School Review*, *86*(2), pp.217-250.

Marzano, R.J., Waters, T. and McNulty, B.A., 2005. *School leadership that works: From research to results*. ASCD.

McCaffrey, D.F., Lockwood, J.R., Koretz, D., Louis, T.A. and Hamilton, L., 2004. Models for value-added modeling of teacher effects. *Journal of educational and behavioral statistics*, *29*(1), pp.67-101.

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

Myung, J., Loeb, S. and Horng, E., 2011. Tapping the principal pipeline: Identifying talent for future school leadership in the absence of formal succession management programs. *Educational Administration Quarterly*, *47*(5), pp.695-727.

Neal, D., 1995. Industry-specific human capital: Evidence from displaced workers. *Journal of labor Economics*, *13*(4), pp.653-677.

Newleaders.org, 2018. Prioritizing Leadership: An Analysis of State ESSA Plans. Accessed Dec. 13, 2018.

Papay, J.P., Taylor, E.S., Tyler, J.H. and Laski, M., 2016. *Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data* (No. w21986). National Bureau of Economic Research.

Parsons, D.O., 1972. Specific human capital: An application to quit rates and layoff rates. *Journal of political economy*, *80*(6), pp.1120-1143.

Pounder, D.G., Galvin, P. and Shepherd, P., 2003. An analysis of the United States educational administrator shortage. *Australian Journal of Education*, *47*(2), pp.133-145.

Rammer, R.A., 2007. Call to action for superintendents: Change the way you hire principals. *The Journal of Educational Research*, *101*(2), pp.67-76.

report to the Legislative Education Study Committee. Senate Joint Memorial 3: Reports

Rivkin, S.G., Hanushek, E.A. and Kain, J.F., 2005. Teachers, schools, and academic achievement. *Econometrica*, *73*(2), pp.417-458.

Robinson, V.M., Lloyd, C.A. and Rowe, K.J., 2008. The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational administration quarterly*, *44*(5), pp.635-674.

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

Ronfeldt, M., Brockman, S.L. and Campbell, S.L., 2018. Does Cooperating Teachers' Instructional Effectiveness Improve Preservice Teachers' Future Performance? *Educational Researcher*, *47*(7), pp.405-418.

Rotherham, A. 2010 "Paging Principal Skinner: Evaluating School Leaders" *Time*. Oct. 21. Accessed Dec. 13, 2018.

Rowland, C., 2017. Principal Professional Development: New Opportunities for a Renewed State Focus. *Education Policy Center at American Institutes for Research*.

Roza, M., 2003. A Matter of Definition: Is There Truly a Shortage of School Principals? Seattle, WA: Center on Reinventing Public Education, University of Washington.

Rusch, E. (2005). Institutional

Teh, B.R., Resch, A., Walsh, E., Isenberg, E. and Hock, H., 2013. Is the stability of value added underestimated. In *AEFP 38th Annual Conference, New Orleans, LA*.

Theobald, R.J., Goldhaber, D.D., Gratz, T.M. and Holden, K.L., 2018. Career and technical education, inclusion, and postsecondary outcomes for students with learning disabilities. *Journal of learning disabilities*, p.0022219418775121.

van de Grift, W., & Houtveen, A. A. M. (1999). Educational leadership and pupil achievement in primary education. *School Effectiveness and School Improvement*, *10*(4), 373–389.

Weiss, A., 1995. Human capital vs. signalling explanations of wages. *Journal of Economic perspectives*, *9*(4), pp.133-154.
DeVaro, J., Kauhanen, A. and Valmari, N., 2015. Internal and external hiring: the role of prior job assignments. In *Fourth SOLE-EALE World Meeting, Montreal*.

Whitaker, K.S., 2003. Principal role changes and influence on principal recruitment and selection: An international perspective. *Journal of educational administration*, *41*(1), pp.37-54.

Winograd, P., Garcia, V., & Dasenbrock, R. (2008). Strong Leaders for New Mexico Schools: A A Report to the Legislative Education Study Committee. *Senate Joint Memorial 3: Report & Recommendations*

Winter, P.A. and Morgenthal, J.R., 2002. Principal recruitment in a reform environment: Effects of school achievement and school level on applicant attraction to the job. *Educational Administration Quarterly*, *38*(3), pp.319-340.

Tables and Figures

Table 1. Teacher Subsamples by Principal and Value-Added Status

	Teachers *without* value-added estimates		Teachers *with* value-added estimates	
	(1) *Not* observed as principals	(2) Observed as principals	(3) *Not* observed as principals	(4) Observed as principals
Teacher value-added, math	N/A	N/A	-0.0002	0.0175
Teacher value-added, reading	N/A	N/A	0.0010	-0.0693
Female	0.698	0.515***	0.761	0.600***
White	0.912	0.892***	0.902	0.862**
Highest degree during teaching:				
Bachelors	0.436	0.182***	0.309	0.127***
Masters or higher	0.545	0.815***	0.690	0.873***
WEST-B math	-0.023	0.094	0.056	-0.104
WEST-B reading	-0.014	0.111	0.032	-0.054
Internal hire: district	N/A	0.507	N/A	0.561
Internal hire: school	N/A	0.175	N/A	0.206
Unique observations	123,705	2,747	26,064	355

Notes: Each column represents four subsamples: Teachers with and without TVA, and teachers who are and are not observed as principals with PVA estimates. Each cell reports unweighted means for the relevant statistic. Internal hires are defined as whether an individual has any previous experience teaching within the school or district where they first serve as a principal. WEST-B scores are standardized within subject. Values of significance are calculated from two-tailed tests between columns (1) and (2), & (3) and (4): *p<0.10, **p<0.05, ***p<0.01.

Table 2. Correlations Across Teacher Value-Added Models

		Math		Reading	
		No classroom covariates	Classroom covariates	No classroom covariates	Classroom covariates
Math	No classroom covariates	1.000			
	Classroom covariates	0.974	1.000		
Reading	No classroom covariates	0.611	0.576	1.000	
	Classroom covariates	0.595	0.587	0.984	1.000

Notes: Each element reports the raw correlation coefficient between relevant specifications. Within subject correlations are pairwise, and cross-subject comparisons are limited to individuals with estimates in both specifications and subjects. All TVA models include a cubic in prior reading and math scores and controls for student demographics. The sample is limited to 17,506 teachers who have both reading and math TVA scores.

Table 3. Correlations Between Principal Value-Added Models

			Math				Reading			
			School value added		School fixed effect		School value added		School fixed effect	
			All obs	No Yr 1	All obs	No Yr 1	All obs	No Yr 1	All obs	No Yr 1
Math	School value added	All obs	1.000							
		No Yr 1	0.893	1.000						
	School fixed effect	All obs	0.937	0.836	1.000					
		No Yr 1	0.802	0.931	0.863	1.000				
Reading	School value added	All obs	0.505	0.500	0.443	0.428	1.000			
		No Yr 1	0.466	0.522	0.408	0.461	0.939	1.000		
	School fixed effect	All obs	0.424	0.428	0.433	0.429	0.923	0.866	1.000	
		No Yr 1	0.141	0.193	0.214	0.289	0.794	0.865	0.823	1.000

Notes: Each element reports the raw correlation coefficient between relevant specifications. All PVA models include a cubic in prior reading and math scores, controls for student demographics, and school-average covariates. The sample is limited to 2,464 principals who have both reading and math PVA scores.

Table 4. Relationship Between Teacher Value Added and Principal Value Added

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Dependent variable is PVA math								
TVA math	0.011	0.023***	0.008	0.016**	0.012	0.024***	0.009	0.016**
	(0.009)	(0.009)	(0.008)	(0.007)	(0.009)	(0.009)	(0.008)	(0.007)
No TVA observed	-0.045***	-0.022**	-0.033***	-0.012	-0.039***	-0.016*	-0.029***	-0.009
	(0.012)	(0.009)	(0.011)	(0.007)	(0.012)	(0.009)	(0.011)	(0.008)
Within-school PVA	No	No	Yes	Yes	No	No	Yes	Yes
Exclude Yr 1	No	Yes	No	Yes	No	Yes	No	Yes
Teacher covariates	No	No	No	No	Yes	Yes	Yes	Yes
Observations	3102	2706	2846	2464	3102	2706	2846	2464
Panel B: Dependent variable is PVA reading								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TVA reading	0.015***	0.022***	0.010**	0.010	0.017***	0.021***	0.012***	0.010
	(0.006)	(0.008)	(0.004)	(0.007)	(0.005)	(0.007)	(0.004)	(0.007)
No TVA observed	-0.017**	-0.017*	-0.007	-0.005	-0.013	-0.012	-0.004	-0.005
	(0.008)	(0.009)	(0.007)	(0.009)	(0.008)	(0.009)	(0.007)	(0.009)
Within-school PVA	No	No	Yes	Yes	No	No	Yes	Yes
Exclude Yr 1	No	Yes	No	Yes	No	Yes	No	Yes
Teacher covariates	No	No	No	No	Yes	Yes	Yes	Yes
Observations	3102	2706	2846	2464	3102	2706	2846	2464

Notes: Each column and panel represent a separate regression where the dependent variable is a PVA score estimated as indicated in the relevant column. Teachers missing TVA scores have imputed values to zero, and the variable "No TVA observed" indicates where values are imputed. Robust standard errors are reported. Values of significance are calculated from two-tailed tests: *p<0.10, **p<0.05, ***p<0.01.

Table 5. Characteristics of Principals by Hiring Type, Internal Relative to External

	Panel A: Dependent variable is PVA math				Panel B: Dependent variable is PVA reading			
External	Omitted Categories				Omitted Categories			
Internal to district	0.011*	0.019***	0.001	0.006	0.008	0.009	0.006	-0.003
	(0.006)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)
Internal to school	-0.033***	-0.038***	-0.019***	-0.018***	-0.016**	-0.016**	-0.011*	0.003
	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)
TVA	0.030**	0.035***	0.026**	0.030***	0.013**	0.020*	0.011***	0.006
	(0.013)	(0.012)	(0.012)	(0.011)	(0.006)	(0.011)	(0.004)	(0.009)
TVA *	-0.018	-0.007	-0.019	-0.009	0.007	0.004	0.000	0.003
Internal to district	(0.020)	(0.019)	(0.017)	(0.016)	(0.015)	(0.020)	(0.011)	(0.019)
TVA *	-0.017	-0.012	-0.011	-0.016	-0.003	0.002	-0.001	0.009
Internal to school	(0.022)	(0.022)	(0.019)	(0.017)	(0.021)	(0.019)	(0.016)	(0.019)
No TVA observed	-0.045***	-0.021**	-0.032***	-0.011	-0.017**	-0.016*	-0.007	-0.005
	(0.012)	(0.009)	(0.011)	(0.008)	(0.008)	(0.009)	(0.007)	(0.009)
Within-school PVA	No	No	Yes	Yes	No	No	Yes	Yes
Exclude yr 1	No	Yes	No	Yes	No	Yes	No	Yes
Observations	3102	2706	2846	2464	3102	2706	2846	2464

Notes: Each column represents a separate regression on the sample of principals where the dependent variables are listed in the relevant column. Robust standard errors are reported. Values of significance are calculated from two-tailed tests: *p<0.10, **p<0.05, ***p<0.01.

Appendix A. Principal value added estimated with school fixed effects

Several studies (e.g., Branch et al., 2009; Grissom et al., 2014) use school fixed effects to exploit within-school variation and separate principal performance from the school context. Consistent with this broader literature, we utilize a specification of the following form:

$$Y_{isptb} = \beta_0 Y_{it-1} + \alpha\beta_1 X_{it} + \delta_p + \delta_b + u_{isptb} \quad (3)$$

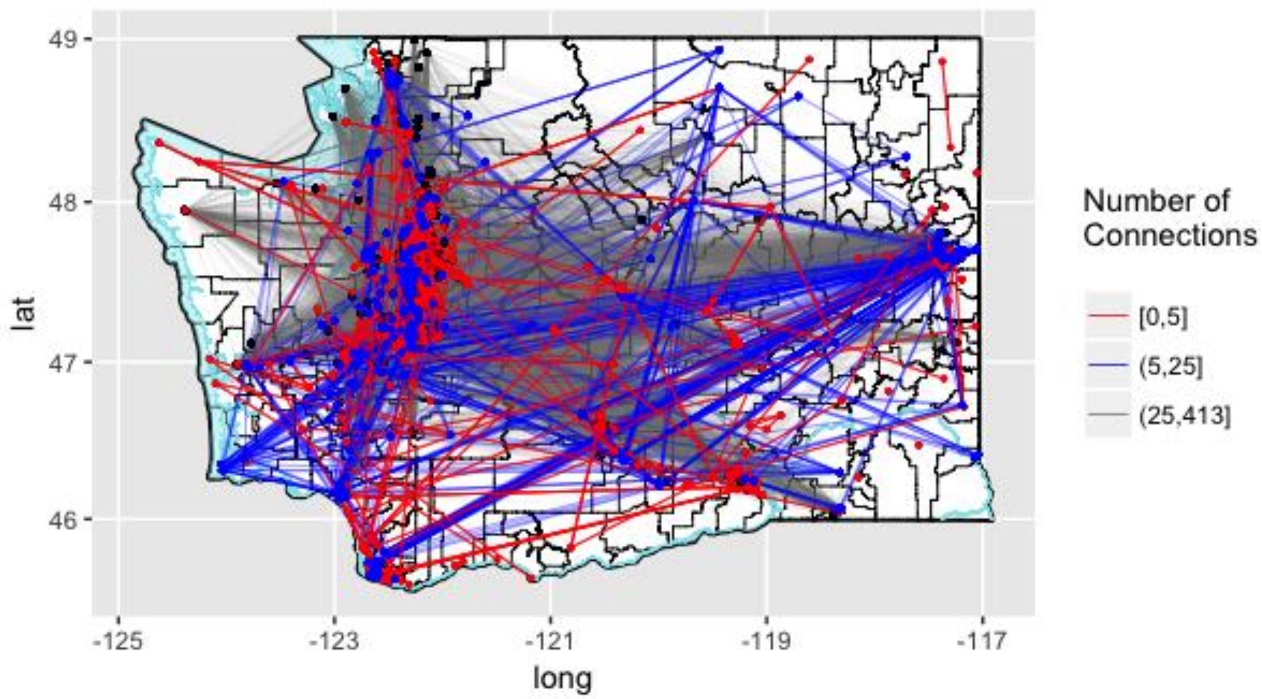
where, similar to principal value-added model discussed above, Y_{isptb} represents student i's test score in subject s, under principal p in year t for school building b. Here, δ_b represents a school fixed effect, and principal fixed effects are estimate relative to other individuals who serve in the same school setting.[45]

These models account for the time-invariant unobservable characteristics of schools that could bias principal value-added scores. That said, within-school models impose several restrictions on the data. First, they can only estimate effects when more than one principal is observed in a school, so that these estimates depend on the amount of mobility within our sample; this excludes 35% to 46% of principal observations depending on the specification. Second, we can only interpret the size of value-added estimates when there is overlap in principal mobility across school settings; in other words, if principals tend to participate in distinct labor markets, then we cannot compare estimates across these settings.

We can only interpret the size of value-added estimates when there is overlap in principal mobility across school settings; in other words, if principals tend to participate in distinct networks, then we cannot compare estimates across these settings.[46] We explore these principal networks in Figure 1. Each point represents a school site, and each line represents a connection between two schools due to principal mobility. Networks are grouped together according to the number of connections: red indicates less than five, blue is 5 to 25, and black is more than 25. The four largest networks contain 44% of principal observations, while about a third of principals are connected in smaller, disjoint networks. About 23% of principals are connected by very small networks of 2 to 5 connections. These networks are considerably more connected than those studied in previous research; for example, Chiang et al. (2016) find that 3,428 out of 5,238 principal-grade observations involve single-school networks. This could be due to their notably shorter panel period (2007-08 to 2012-13) or contextual differences between Pennsylvania and Washington.

[45] Unlike previous specifications, Equation (3) is estimated jointly across grade spans in order to capture mobility across different types of schools.
[46] This idea is closely related to research by Mihaly et al. (2013) who consider the modeling challenges of including school fixed effects when estimating the effectiveness of education preparation programs. They find that even though there is sufficient overlap due to teacher mobility, those individuals who connect schools tend to differ in their observable characteristics which may distort cross-market comparisons. For principal research, similar concerns are likely present as individuals who are more mobile could differ substantially from their peers.

Figure 1. Network connections between principals for within-school principal value-added models



Notes: Network connections are from PVA models that do not exclude the first year. Each point on the figure represents a school site in Washington state and connected lines between sites represent settings where principals can be compared to each other. The different colors represent different networks of connections: Red shows networks with between zero and five connections, blue indicates networks with 5 to 25 connections, and black represents networks with more than 25 connections. The largest network has 413 connections.

Table A.1 Relationship Between Teacher Value Added and Principal Value Added

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Dependent variable is PVA math								
TVA math	0.011	0.023***	0.005	0.012	0.012	0.024***	0.006	0.013
	(0.009)	(0.009)	(0.014)	(0.011)	(0.009)	(0.009)	(0.014)	(0.011)
No TVA observed	-0.045***	-0.022**	-0.055***	-0.040***	-0.039***	-0.016*	-0.054***	-0.034***
	(0.012)	(0.009)	(0.017)	(0.011)	(0.012)	(0.009)	(0.017)	(0.012)
School FE PVA	No	No	Yes	Yes	No	No	Yes	Yes
Exclude Yr 1	No	Yes	No	Yes	No	Yes	No	Yes
Teacher covariates	No	No	No	No	Yes	Yes	Yes	Yes
Observations	3102	2706	3102	2706	3102	2706	3102	2706
Panel B: Dependent variable is PVA reading								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TVA reading	0.015***	0.022***	0.003	0.019*	0.017***	0.021***	0.007	0.021**
	(0.006)	(0.008)	(0.011)	(0.010)	(0.005)	(0.007)	(0.011)	(0.010)
No TVA observed	-0.017**	-0.017*	-0.013	-0.005	-0.013	-0.012	-0.021	-0.010
	(0.008)	(0.009)	(0.016)	(0.012)	(0.008)	(0.009)	(0.016)	(0.012)
School FE PVA	No	No	Yes	Yes	No	No	Yes	Yes
Exclude Yr 1	No	Yes	No	Yes	No	Yes	No	Yes
Teacher covariates	No	No	No	No	Yes	Yes	Yes	Yes
Observations	3102	2706	3102	2706	3102	2706	3102	2706

Notes: Each column and panel represent a separate regression where the dependent variable is a PVA score estimated as indicated in the relevant column. Teachers missing TVA scores have imputed values to zero, and the variable "No TVA observed" indicates where values are imputed. Robust standard errors are reported. Values of significance are calculated from two-tailed tests: *p<0.10, **p<0.05, ***p<0.01.

Appendix B. Robustness Checks

Table B.1 Marginal Effects for Probit Regressions of Principal Selection

	Dependent variable is 1(principal)	
TVA math		0.0011
		(0.001)
TVA reading		0.0001
		(0.001)
Female	-0.0109***	-0.0076***
	(0.0008)	(0.003)
White	-0.0052***	0.0063
	(0.0012)	(0.008)
MA	0.0228***	0.0011***
	(0.0010)	(0.001)
Ph.D.	0.0187***	0.0001
	(0.0038)	(0.001)
Observations	1,574,407	52,644

Notes: Each column represents a separate probit regression where the dependent variable is whether a Teacher was hired as a principal. The first column includes all teachers, and the second column includes only individuals with non-missing TVA. Select coefficients reported. Standard errors are clustered on teacher IDs and reported. Values of significance are calculated from two-tailed tests: *p<0.10, **p<0.05, ***p<0.01.

Table B.2 Relationship Between Teacher Value Added for Math and Reading and Principal Value Added

	Panel A: Dependent Variable is PVA math							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TVA math	0.005	0.012	0.007	0.011	0.008	0.014	0.007	0.011
	(0.010)	(0.010)	(0.009)	(0.008)	(0.010)	(0.010)	(0.009)	(0.008)
TVA reading	0.010	0.022**	0.003	0.011*	0.009	0.020**	0.003	0.010
	(0.006)	(0.009)	(0.005)	(0.006)	(0.006)	(0.009)	(0.005)	(0.006)
No TVA observed	-0.046***	-0.023**	-0.033***	-0.013*	-0.039***	-0.017*	-0.030***	-0.010
	(0.011)	(0.009)	(0.010)	(0.007)	(0.011)	(0.009)	(0.010)	(0.008)
Exclude Yr 1	No	Yes	No	Yes	No	Yes	No	Yes
Teacher covariates	No	No	No	No	Yes	Yes	Yes	Yes
Observations	3102	2706	2846	2464	3102	2706	2846	2464

	Panel B: Dependent Variable is PVA reading							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TVA math	0.002	-0.005	0.002	-0.004	0.003	-0.004	0.003	-0.004
	(0.010)	(0.012)	(0.008)	(0.010)	(0.010)	(0.012)	(0.008)	(0.010)
TVA reading	0.015**	0.024***	0.010**	0.011	0.016***	0.022***	0.011**	0.012
	(0.006)	(0.009)	(0.005)	(0.008)	(0.006)	(0.009)	(0.004)	(0.008)
No TVA observed	-0.017**	-0.017*	-0.006	-0.006	-0.012	-0.013	-0.004	-0.005
	(0.008)	(0.009)	(0.007)	(0.009)	(0.008)	(0.009)	(0.007)	(0.009)
Exclude Yr 1	No	Yes	No	Yes	No	Yes	No	Yes
Teacher covariates	No	No	No	No	Yes	Yes	Yes	Yes
Observations	3102	2706	2846	2464	3102	2706	2846	2464

Notes: Each column and panel represent a separate regression where the dependent variable is a PVA score estimated as indicated in the relevant column. Teachers missing TVA scores have imputed values to zero, and the variable "No TVA observed" indicates where values are imputed. Robust standard errors are reported. Values of significance are calculated from two-tailed tests: *p<0.10, **p<0.05, ***p<0.01.