EdWorkingPaper No. 20-223

Pre-Service Teacher Quality and Workforce Entry

Brendan Bartanen
Texas A&M University

Andrew Kwok
Texas A&M University

Using rich longitudinal data from one of the largest teacher education programs in Texas, we examine the measurement of pre-service teacher (PST) quality and its relationship with entry into the K–12 public school teacher workforce. Drawing on rubric-based observations of PSTs during clinical teaching, we find that little of the variation in observation scores is attributable to actual differences between PSTs. Instead, differences in scores largely reflect differences in the rating standards of field supervisors. We also find that men and PSTs of color receive systematically lower scores. Finally, higher-scoring PSTs are slightly more likely to enter the teacher workforce and substantially more likely to be hired at the same school as their clinical teaching placement.

VERSION: April 2020

Suggested citation: Bartanen, Brendan, and Andrew Kwok. (2020). Pre-Service Teacher Quality and Workforce Entry. (EdWorkingPaper: 20-223). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/jvss-mr72

Pre-Service Teacher Quality and Workforce Entry

Brendan Bartanen

Texas A&M University

Andrew Kwok

Texas A&M University

April 2020

Author Note

We are grateful to Jason Grissom, Aliza Husain, and R. Joseph Waddington for their feedback on the manuscript. We thank Kimberly Parish, Misti Corn, Mary Ronsonet, Tara Slaydon, Rahul Sharma, and Jonathan Hall for their assistance in data collection, and Jori Beck for insight on literature.

Abstract

Using rich longitudinal data from one of the largest teacher education programs in Texas, we examine the measurement of pre-service teacher (PST) quality and its relationship with entry into the K–12 public school teacher workforce. Drawing on rubric-based observations of PSTs during clinical teaching, we find that little of the variation in observation scores is attributable to actual differences between PSTs. Instead, differences in scores largely reflect differences in the rating standards of field supervisors. We also find that men and PSTs of color receive systematically lower scores. Finally, higher-scoring PSTs are slightly more likely to enter the teacher workforce and substantially more likely to be hired at the same school as their clinical teaching placement.

Pre-Service Teacher Quality and Workforce Entry

Introduction

Policymakers and researchers have started to examine more critically the experiences and training of pre-service teachers (PSTs). This attention is warranted, as roughly 80% of new teachers enter the profession through university-based teacher education programs, or TEPs (National Center for Education Statistics, 2018). TEPs thus present an attractive policy lever for improving the quality of the K–12 teacher workforce and, by consequence, the outcomes of students. Nonetheless, evidence on the extent to which TEPs contribute to variation in in-service teacher quality is mixed (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Gansle, Noell, & Burns, 2012; Henry et al., 2014; Koedel, Parsons, Podgursky, & Ehlert, 2015; Mihaly, McCaffrey, Sass, & Lockwood, 2013), despite knowing the vast majority of the variation exists within rather than between TEPs (Goldhaber, Liddle, & Theobald, 2013).

Despite increased focus on TEPs, there has been little attention to understanding the variation in PST quality, particularly within programs.[1] This gap is notable from both a practice and policy perspective. Tasked with developing PSTs' skills, behaviors, and mindsets, TEPs need to accurately measure performance to provide PSTs with specific feedback on their practice and to identify those who might need additional support prior to entering the classroom. Yet we have little systematic evidence about how TEPs measure PST quality and whether they can accurately distinguish among high and low performers. It is also critical to understand whether high-quality PSTs are more or less likely to enter the teacher workforce, as well as the extent to which TEPs influence this screening process.

Our limited understanding of PST quality and its relationship with workforce entry stems from two main gaps in the literature. First, there are few recent studies examining

[1] Prior research uses teacher quality as a broad term to mean teacher/teaching effectiveness or readiness. We use PST quality in the same manner, not differentiating the terms of teacher/teaching quality/effectiveness/readiness, though we recognize there are subtle differences in these terms.

measures of PST quality within TEPs. In particular, existing studies have largely relied on certification (or licensure) exam scores, as opposed to measures directly collected by TEPs, such as classroom observations during clinical teaching. Exam scores may be poor or incomplete proxies for teacher quality or readiness (e.g., D'Agostino & Powers, 2009; Goldhaber, Cowan, & Theobald, 2017; Goldhaber & Hansen, 2010; Goldhaber, Krieg, & Theobald, 2014), whereas observation scores are more plausibly related to the practice of teaching. The second gap is that prior work on TEPs—particularly studies that connect TEPs to their graduates' outcomes—has almost exclusively examined individuals who ultimately selected into the teacher labor market and received a teaching job (Goldhaber, 2019; Goldhaber et al., 2014), rather than the full set of individuals that completed a teacher education program. Thus, we know little about the factors within teacher education that influence PST entry into the teacher workforce.

This article helps to fill these gaps by examining the measurement of PST quality and, subsequently, the relationship between PST quality and entry into the public teacher workforce. We access a unique dataset from one of the largest TEPs in Texas, which contains detailed information on all undergraduates who enroll in the teacher education program. Importantly, we can access multiple measures or proxies of PST quality, including SAT scores, high school GPA, teacher certification exam scores, and observational evaluations during clinical teaching. Texas requires that all PSTs in a clinical teaching placement receive observational evaluations conducted by a trained field supervisor. While ostensibly for the purpose of developing PSTs' pedagogical skills, the scores must also be provided to the principal of the placement school. Thus, schools or districts may draw on these scores as a measure of PST quality when making hiring decisions. We link these data to state administrative records, which allows us to observe whether a PST entered the Texas K–12 public school workforce and, conditional on entry, where they worked.

We make a number of important contributions. First, we examine the observational ratings of PSTs during their clinical (or student) teaching, defined as full-time teaching in a field placement classroom near the completion of preparation. Despite their widespread use in teacher education programs (National Council for Accreditation of Teacher Education, 2010), we know little about what these observations are actually measuring. A growing body of work examines teaching observations in the in-service context, with mixed evidence as to whether these scores accurately measure teacher performance. While some studies find a positive relationship between observation scores and student achievement growth (Garrett & Steinberg, 2015; Grissom & Loeb, 2017; Kane, Taylor, Tyler, & Wooten, 2011), others have demonstrated these these scores in part reflect factors that teachers cannot control, such as their race, gender, classroom composition, and school context (Campbell & Ronfeldt, 2018; Cohen & Goldhaber, 2016; Steinberg & Garrett, 2016). Motivated by these findings, we examine the extent to which PSTs' scores appear to reflect their own performance, as opposed to factors outside of their control.

Second, we provide new evidence about the validity of certification exams, which are are designed to measure the minimum amount of professional knowledge required of all individuals wanting to teach and are a licensure requirement in nearly every U.S. state (Goldhaber et al., 2017; National Center for Education Statistics, 2017; Youngs, Odden, & Porter, 2003). While early evidence indicated that these exams may not provide reliable information about teaching effectiveness (D'Agostino & Powers, 2009), more recent studies find a positive link between teacher performance on these exams and student achievement (e.g., Clotfelter, Ladd, & Vigdor, 2007; Goldhaber, 2007; Goldhaber et al., 2017; Goldhaber & Hansen, 2010). We provide a different type of evidence by comparing certification scores to other measures or proxies of PST quality, including prior academic achievement (SAT score, high school GPA) and observation scores during clinical teaching.

Third, we add to the thin literature examining PST entry into the public school teaching profession. To our knowledge, only two studies have directly modeled this entry process, both of which use data from Washington state. Goldhaber et al. (2014) find that PSTs endorsed to teach in "difficult-to-staff" areas are more likely to find employment as public school teachers, as are younger PSTs, white PSTs, and PSTs who completed student teaching in suburban schools. Notably, they also find that PSTs with higher certification exam scores are no more likely to enter the public school teacher workforce, though they are somewhat more likely to be hired into the same school as their student teaching placement. In the other study, Goldhaber et al. (2017) find a positive association between licensure exam performance (edTPA) and the probability of employment as a public school teacher. Our analysis is novel in that we model entry as a function of multiple measures of PST quality, including observation scores and high school achievement.

To be specific, we seek to answer the following research questions:

1. To what extent are observation scores of pre-service teachers during clinical teaching explained by individual-level, school-level, and supervisor-level variation?
2. How are observation scores related to the characteristics of PSTs and their placement schools?
3. How are certification scores related to the characteristics of PSTs and their placement schools?
4. What is the relationship between PST quality and labor market outcomes, including entry into the public school teacher workforce and the characteristics of the full-time teaching school?

Using a hierarchical linear modeling (HLM) approach, we find that little of the variation in PST observation ratings is attributable to differences between PSTs (12% of total variance). Instead, the majority of the variation is explained by differences in the (arbitrary) rating standards of supervisors (65%), and a small portion is explained by differences between clinical teaching school placements (9%). In practical terms, this means that most of the information contained in these observation scores is unrelated to the performance, readiness, or quality of PSTs. We further illustrate this point by documenting that only weak relationships exist between observation scores and other plausible measures of PST quality, such as high school GPA and scores on pedagogy and content certification exams. However, we show that isolating the component of observation scores that reflects variance among PSTs substantially strengthens these correlations. In particular, we show that higher-quality PSTs, as measured by observation scores, earn higher scores on the pedagogy certification exam, even after accounting for their prior academic achievement.

Additionally, we extend findings from prior work examining observation score biases among in-service teachers by documenting that race and gender gaps also exist among pre-service teachers. Men score 0.29 SD below women, on average, while individuals of color score 0.11 SD below their white peers. These gaps persist even after accounting for a rich set of individual characteristics, placement school characteristics, and supervisor assignments. We also find that PSTs in high-poverty schools score lower than their peers in low-poverty schools, as do PSTs in rural schools.

Finally, we link PST quality to labor market outcomes—both the decision to become a full-time teacher in public K–12 schools and, conditional on entry, the characteristics of their school. We show that higher-rated teachers are more likely to enter. However, when we isolate the individual-, school-, and supervisor-level variance components of PSTs' observation scores, we find that the supervisor-level component is the main driver of this relationship. In other words, positive selection of higher-scoring PSTs into the K–12 workforce seems to be related to arbitrary supervisor assignments, rather than the actual quality of PSTs. Conditional on entry, higher-rated PSTs are substantially more likely to be hired into the school where they completed their clinical teaching. This pattern holds even when comparing PSTs within the same clinical teaching placement, suggesting that hiring principals seek to strategically hire higher-quality PSTs as full-time teachers.

Background

Teacher education is vital in preparing new teachers for the profession (National Council for Accreditation of Teacher Education, 2010). Important to this function is accurate measurement of teacher and teaching effectiveness within preparation programs. As stated by Guarino, Santibañez, and Daley (2006, p.176), "The issue of teacher quality is integrally related to the interplay of supply and demand. Because not all teachers are alike, quality is an important variable that can be adjusted by policymakers in their efforts to bring supply in line with demand." As such, we begin by reviewing studies that measure and identify PST quality, specifically through certification exams and observation evaluations. We then outline factors that influence PST entry into the profession. Finally, we bridge these two areas to frame the contribution of our study.

Measuring Pre-Service Teacher Quality

Despite increased interest in evaluating the contributions of TEPs to in-service teacher quality, little is known about PST quality, though evidence suggests it tends to be greater within programs rather than across programs (Goldhaber et al., 2013). We explore PST quality through two commonly used measures in teacher education detailed below: certification exams and observation scores.

Certification Exams. Certification exams (or professional licensure exams) are required in nearly every U.S. state and can include separate tests for measuring basic skills, content knowledge, and pedagogical knowledge (Goldhaber et al., 2017; National Center for Education Statistics, 2017; Youngs et al., 2003). These exams provide a baseline standard of knowledge required to enter the profession and have been positively linked to in-service teacher evaluation scores (Jacob & Walsh, 2011). However, studies have differed as to whether success on these exams are positively associated with student achievement (Goldhaber, 2007; Goldhaber et al., 2013; Harris & Sass, 2011; Wayne & Youngs, 2003).

On the one hand, certification exams can be seen as a barrier into the profession for individuals who do not exhibit a baseline standard of professional knowledge to become a teacher. For example, D'Agostino and Powers (2009) found in their meta-analysis that these exams could be used to identify the lowest tier of teachers and prevent them from entering the classroom. As a result, some studies have argued that because individual states determine the passing cut score, they could raise that standard to increase the quality of teachers entering the profession (Memory, Coleman, & Watkins, 2003; Shuls, 2018). Shifting the cut score, however, could disproportionately restrict teachers based on ethnicity (Angrist & Guryan, 2008), gender (Goldhaber & Hansen, 2010), and content area (Gitomer, 2007). Understanding this trade-off between quality and teacher characteristics is vital for the supply of new teachers (Goldhaber, 2007). Yet, few studies have investigated how success on certification exams may influence entry into the profession, which we discuss more fully below.

On the other hand, certification could be seen as a signal for teacher effectiveness. Examination of the relationship between certification scores and teacher characteristics has largely been confined to traditional demographics (i.e., race, gender, content area). Because certification is a form of standardized testing, however, measures of high school achievement could be comparatively more important predictors of success. On this point, Goldhaber et al. (2017) calls for investigation of how certification scores "are related to other, broader measures of teacher performance, such as observational ratings" (p. 390). While this specific reference is to in-service evaluations, we hypothesize that various signals of PST quality should be correlated with one another. Alternatively, we may find that certification scores provide no signal about PST quality above and beyond other measures of academic achievement. Our study advances this understanding of PST quality, then, by examining how PST characteristics—including demographics and high school achievement—are related to certification scores. Further, we provide new evidence examining the relationship between certification scores and observation scores from clinical teaching, which we describe next.

Teacher Observation Evaluations. Teacher education programs consistently use observation evaluations —during clinical teaching, specifically—to develop instructional practice (American Association of Colleges for Teacher Education, 2018). Throughout this preparation structure, PSTs participate in professional activities under the guidance of a mentor (or cooperating) teacher for at least one full semester (Anderson & Stillman, 2013; Feiman-Nemser & Buchmann, 1987; National Council for Accreditation of Teacher Education, 2010). During this experience, PSTs are assessed on their pedagogical skills and their professional learning, often by university supervisors, who tend to be part-time faculty providing an external perspective on teaching quality (Clift & Brady, 2005; Ronfeldt, Reininger, & Kwok, 2013; Slick, 1998). Despite evidence suggesting supervisors' importance towards PST development, teacher education policies have increasingly reduced the requirements to become a supervisor (Darling-Hammond, 2014; Grossman, Hammerness, McDonald, & Ronfeldt, 2008; Wasburn-Moses & Noltemeyer, 2018).

According to the National Council for Accreditation of Teacher Education (2010), increased use of clinical teaching observational data is one recommendation for effective teacher preparation to raise PST quality. Unfortunately, the research exploring observation ratings and teacher education is extremely limited. In perhaps the most relevant study, Henry et al. (2013) used five years of data on elementary teacher education graduates from one large preparation program to examine how various measures collected by the program were correlated with in-service teacher value-added. The authors examined five different measures of PST quality, including clinical teaching ratings and basic skills certification exams, and found that these measures 1) signaled only one underlying construct of PST quality, and 2) were not associated with math and reading value-added scores.

Two additional studies have used *in-service* observation scores to draw conclusions about the quality of TEPs. Using statewide data from Tennessee, Ronfeldt and Campbell (2016) found variation in the average observation scores of graduates from different teacher preparation programs. Bastian, Patterson, and Pan (2018) reached a similar conclusion using data from North Carolina, but also found a relationship between in-service observation scores and both high school achievement—including SAT score and high school GPA—and school characteristics.

Otherwise, research on teacher observation ratings is almost exclusively confined to the in-service context. Prior work suggests that teaching observations can provide evidence for teacher-student interactions, the quality of teaching practices across multiple dimensions, and a general marker for teacher effectiveness (Garrett & Steinberg, 2015; Grossman, Loeb, Cohen, & Wyckoff, 2013). Despite vast variation in protocols and instruments to conduct classroom observations, they have been shown to, at the very least, distinguish between weak and sufficient teaching (Grossman, Cohen, Ronfeldt, & Brown, 2014). Scholars have also argued that observations tend to be more predictive of quality instructional practices that are otherwise not encapsulated in—or correlated with—student achievement and value-added model scores (Bell et al., 2012; Grossman et al., 2013; Kane et al., 2011; Strunk, Weinstein, & Makkonen, 2014). As a result, observations are one piece of multiple-measure teacher evaluation systems that tend to be well accepted by all stakeholders (Garrett & Steinberg, 2015). Nonetheless, recent work has demonstrated that classroom observations may suffer from race and gender biases (Campbell & Ronfeldt, 2018) as well as biases stemming from the student composition of a teacher's classroom (Garrett & Steinberg, 2015).

Our study extends knowledge about teacher evaluation in the context of teacher education, where evidence is noticeably scant. Data may have been difficult to systematically collect in this context due to the multiple stakeholders involved (e.g., PSTs, teacher education programs, school districts), as well as the archaic nature of programs still relying on hand-written evaluations.

Connecting Pre-Service Teacher Quality and Entry into the Profession

One particularly important of aspect of PST quality is its relationship with entry into the K–12 teaching profession (Goldhaber, Gross, & Player, 2011; Goldhaber & Ronfeldt, 2020; Guarino et al., 2006). This mechanism not only affects the average quality of the workforce, but also the distribution of teacher quality across schools, which in turn reduces or exacerbates student achievement gaps. Work in this area, however, remains limited. Guarino et al. (2006) argue that, "very few research studies exist, however, that combine issues of recruitment and retention with the issue of teacher quality" because "few sources of data exist that permit researchers to identify effective teachers and examine the factors that promote their recruitment and retention" (p. 176). While increasingly accessible administrative datasets (particularly at the state level) combined with measures of teacher quality (e.g., value-added, observation scores) has spurred a large number of studies examining the relationship between teacher quality and retention, our understanding of recruitment or entry remains limited, absent some notable exceptions that we build upon.

Lankford, Loeb, and Wyckoff (2002) pioneered this work by constructing a quality measure based on certification exam failure (alongside teaching experience and college competitiveness). They found that high-quality teachers in New York tend to work in more advantaged schools, though they also leave these schools sooner than their peers. Similarly, Krieg, Theobald, and Goldhaber (2016) found that more qualified PSTs (based on their certification score and undergraduate GPA) tend to work in more advantaged clinical teaching schools, and that clinical teaching school was a strong predictor of first teaching site. Relatedly, Goldhaber et al. (2014) identified that approximately one-sixths of PST hires were in the same school as their clinical teaching (internship) placement. While certification exam score was not a significant predictor for entry, it was associated with being hired into the same school. The authors posit that clinical teaching was possibly used by districts or schools "screening time" to consider PSTs for future hire. Finally, (Goldhaber et al., 2017) investigated the predictive validity of edTPA (a high-stakes observational score of pedagogy) and found that scores were highly predictive of entry into Washington public schools. However, validity issues cloud the use of edTPA, particularly as it relates to consequential decisions for both PSTs and teacher preparation programs (Gitomer, Martínez, Battey, & Hyland, 2020).

We study the relationship between teacher workforce entry and PST quality as measured through observational evaluations from supervisors—which are ubiquitous throughout teacher education but have yet to be systematically explored—and certification scores from pedagogy and content exams. In this way, our study addresses Goldhaber's (2019) call for additional research about how experiences in teacher preparation, namely clinical teaching, connect to professional outcomes such as PSTs' choice to enter the profession.

Data

The study sample includes roughly 2,100 undergraduate students majoring in Interdisciplinary Studies in a college of education at a Texas public university—one of the largest producers of teachers for the state every year. These pre-service teachers were concurrently enrolled in one of the following teacher certification programs between the fall semester of 2012 through the winter semester of 2018: Early Childhood through 6th Grade Core Subjects (50% of PSTs), Middle Grades 4-8th grade Math/Science (26%), or Middle Grades 4-8th grade English Language Arts/Social Studies (24%). Below, we describe in more detail the various data used in our analyses.

Demographic and admissions data

For each PST, we can access basic demographic information, including race/ethnicity, gender, and a categorical measure of family income. As shown in Table 1, our PST sample is 84% White and 95% female, both of which are slightly above the national averages (80%

and 89%, respectively) for elementary school teachers.[2] We also access admissions data, including a categorical measure of family income, high school GPA, SAT score, and whether the PST is a transfer student or first-generation college student. Overall, our sample is fairly affluent and high-achieving, with a mean high school GPA of 3.5, a mean SAT score of 1100 (with 14% missing), and 44% of participants from families earning more than $100,000 per year. A relatively high percentage of PSTs (36%) are transfer students, largely driven by individuals completing program prerequisites at a nearby community college.

Certification data

Teachers applying for initial licensing in Texas during the study period were required to pass two certification exams as part of the Texas Examinations of Educator Standards (TExES). The first exam (henceforth referred to as “content”) varies by program. PSTs seeking elementary certification (EC-6) were required to pass a core subjects exam that covered all content areas, whereas middle grades PSTs took an exam that corresponded to their specific content area (math/science or ELA/social studies). The second exam is the Pedagogy and Professional Responsibilities (henceforth referred to as “pedagogy”) exam. PSTs were required to complete the content exam prior to clinical teaching—as per Texas law—and the pedagogy exam prior to graduation and in order to receive their teaching certificate. In addition to whether the PST passed the exam, we also have the scaled scores, which were used to construct a standardized score for pedagogy and content. Standardization is straightforward for pedagogy, since there is a single exam. For content, we standardized within each exam and pooled scores into a single measure.

Clinical teaching evaluation data

The program consists of three semesters of field placements. However, only the final semester includes a clinical teaching requirement. During this time, PSTs are expected to

[2] https://nces.ed.gov/programs/coe/indicator_clr.asp

take over much of the classroom responsibilities. Over the course of the semester, PSTs receive four 45-minute formal observations. These observations are conducted by field supervisors, who are part-time contractors of the university. The observation instrument used mimics the state evaluation rubric for (in-service) beginning teacher competency; PSTs are evaluated in the pedagogical areas of planning, instruction, and creating a learning environment. While primarily intended for developmental purposes to provide PSTs with feedback about their teaching practice, observations can also serve as a red flag system to identify and provide support to struggling individuals. Finally, per state law, observation scores are automatically distributed to both the mentor teacher and principal at the clinical teaching placement site.

There were 148 field supervisors from across the state in our sample, 77% of whom were female. Under state law, supervisors must hold current teacher certification in the same area as the PST's classroom, have three years of teaching experience, and complete a state-approved training in teacher observation. Additionally, supervisors may not be employed at the clinical teaching placement school. The university also requires the completion of an online new supervisor training and twice a semester, supervisors facilitate meetings with the PST and mentor teacher to collectively discuss PST development.

In the semester prior to their clinical teaching, PSTs are required to submit their school district preference to the program. Conditional on a district's willingness to accept student teachers, PSTs are often placed in their preferred district. However, school and mentor teacher placements are determined by the individual districts according to their own placement criteria. Among the roughly 2,100 PSTs in our sample, we observe 129 unique districts and 609 unique schools, which are scattered throughout the state. Because placements are so widespread, the university hires and assigns supervisors based on their proximity to placement schools.

For each PST, we have scores from each of their four observations, as well as identifiers for the supervisor and school. For this analysis, we use the average score from the set of rubric items on a 1 to 4 scale. Using the average score is advantageous because the rubric items vary somewhat across years. Additionally, we conducted factor analyses in each semester/year and found consistent evidence that scores measured a single underlying construct, similar to (Henry et al., 2013). Finally, we standardized these average scores within each semester and year.

Job placement data

Each year, the TEP provides a list of graduates to the Texas Education Agency, which then provides a matched list with information that includes school placement and job title. We use this matched list to construct a binary indicator of whether a PST enters the K–12 public school system. We cannot observe whether graduates are employed in private schools or jobs outside of education. Additionally, the full list of graduates (beginning in 2012) is matched each year, which allows us to observe delayed entries. This placement information also allows us to construct indicators for whether PSTs enter the same school or district as their clinical teaching placement.

Methods

Variance Decomposition of Observation Scores

Our first research question examines the extent to which variation in PST observation scores during clinical teaching is attributable to differences between PSTs, differences between placement schools, and differences between supervisors. As discussed above, PSTs receive four rubric-based observations during their semester of clinical teaching, each of which is performed by the same supervisor. Further, supervisor assignments are geographic, such that schools are almost perfectly nested within supervisor.[3] We exploit this nested structure using a hierarchical linear modeling (HLM)

[3] The nesting is not perfect because there was some turnover in supervisors across years. Among the 609 unique schools, 78% had a single supervisor, 15% had two supervisors, and 7% had three or more supervisors. One way to address this is to model supervisors as a crossed random effect. However, this

approach with observations at level 1, teacher candidates at level 2, placement schools at level 3, and supervisors at level 4.[4] This approach allows us to partition the variability in observation scores into each of the four levels. While we show a range of specifications, our preferred variance decomposition model allows for random intercepts and slopes at levels 2, 3, and 4. Specifically, we estimate the following growth model at level 1:

$$\text{Level 1}: \quad Y_{itjs} = \pi_{0tjs} + \pi_{1tjs} Order_{itjs} + \epsilon_{itjs} \tag{1}$$

where Y is a standardized observation score and i, t, j, and s index observations, teacher candidates, schools, and supervisors, respectively. In the unconditional model, a PST's score for observation i (1st through 4th) is a function of their initial level, π_{0tjs}, their improvement rate, π_{1tjs}, and a random error term, $\epsilon_{itjs} \sim N(0, \sigma^2)$. We treat π_{0tjs} and π_{1tjs} as having both fixed and random components that vary by PST, school, and supervisor. This leads to sets of models for the slope and intercept at levels 2, 3, and 4:

$$\begin{aligned} \text{Level 2}: \quad & \pi_{0tjs} = \beta_{00js} + r_{0tjs} \quad \text{and} \quad \pi_{1tjs} = \beta_{10js} + r_{1tjs}, \\ & \text{where} \begin{bmatrix} r_{0tjs} \\ r_{1tjs} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{r_{0tjs}} & \sigma_{r_{0tjs}, r_{1tjs}} \\ \sigma_{r_{1tjs}, r_{0tjs}} & \sigma^2_{r_{1tjs}} \end{bmatrix} \right) \end{aligned} \tag{2}$$

$$\begin{aligned} \text{Level 3}: \quad & \beta_{00js} = \gamma_{000s} + \mu_{00js} \quad \text{and} \quad \beta_{10js} = \gamma_{100s} + \mu_{10js}, \\ & \text{where} \begin{bmatrix} \mu_{00js} \\ \mu_{10js} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\mu_{00js}} & \sigma_{\mu_{00js}, \mu_{10js}} \\ \sigma_{\mu_{10js}, \mu_{00js}} & \sigma^2_{\mu_{10js}} \end{bmatrix} \right) \end{aligned} \tag{3}$$

approach yields very similar results to our preferred approach, which is to simply model schools as supervisor-by-schools to obtain perfect nesting. The estimated school random effects (BLUPs) from these models are highly correlated ($r = 0.81$). Given the small variance component for schools, we opted for the simpler approach, which also allows us to more easily model random slopes for supervisors.

[4] We also considered an additional level for school district, but found little evidence of substantial between-district heterogeneity. Specifically, we estimated an intercepts-only model with supervisors specified as a crossed random effect and school districts at level 4. The variance proportion for district was 0.7%.

$$\text{Level 4}: \quad \gamma_{000s} = \phi_{0000} + \zeta_{000s} \quad \text{and} \quad \gamma_{100s} = \phi_{1000} + \zeta_{100s},$$
$$\text{where} \begin{bmatrix} \zeta_{000s} \\ \zeta_{100s} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\zeta_{000s}} & \sigma_{\zeta_{000s},\zeta_{100s}} \\ \sigma_{\zeta_{100s},\zeta_{000s}} & \sigma^2_{\zeta_{100s}} \end{bmatrix} \right) \quad (4)$$

Here, the parameters of interest are the variances of the random intercepts and slopes at each level along with the covariance between the intercept and slope within each level. For instance, $\sigma^2_{r_{0tjs}}$ is the model-based estimate of the variance of r_{0tjs}, which is the deviation of pre-service teacher t's intercept from the school mean intercept. $\sigma^2_{r_{1tjs}}$ follows equivalently for the slope. The covariance term, $\sigma_{r_{1tjs},r_{0tjs}}$, allows for the possibility that, for instance, PSTs who initially receive higher scores experience lower growth over the course of the semester. From here, we can estimate the proportion of variance that is (1) within PSTs, (2) among PSTs within schools, (3) among schools within supervisors, and (4) among supervisors.[5]

Beyond describing the variability at each level, we can also draw on the predicted random effects (best linear unbiased predictions or "BLUPs") from these HLM models to be used as right-hand-side variables for our analyses of certification scores and labor market outcomes. These BLUPs represent the "contribution" of an individual supervisor, school, or PST to a PST's observation score—analogous to the interpretation of teacher random effects in a value-added model as a teacher's contribution to student test scores.[6]

Although the exact mechanisms that produce the variation in observation scores at each level are unobserved, we propose the following conceptualizations of these variance components. We argue that supervisor effects likely reflect arbitrary differences in rating standards among supervisors (i.e., some supervisors grade more harshly than others), given that supervisor assignment to PST is effectively random conditional on geographic location

[5] The variance components by level are (starting from level 1): σ^2; $\sigma^2_{r_{0tjs}} + \sigma^2_{r_{1tjs}} + 2\sigma_{r_{1tjs},r_{0tjs}}$; $\sigma^2_{\mu_{00js}} + \sigma^2_{\mu_{10js}} + 2\sigma_{\mu_{00js},\mu_{10js}}$; $\sigma^2_{\zeta_{000s}} + \sigma^2_{\zeta_{100s}} + 2\sigma_{\zeta_{100s},\zeta_{000s}}$. To calculate the proportion of variance at each level, we divide an individual variance component by the sum of the four variance components.

[6] To account for the fact the BLUPs themselves are estimates, when using them as predictors we compute bootstrapped standard errors.

of the placement school. School effects, on the other hand, are more ambiguous. Average differences in PST observation scores among placement schools (which are nested within supervisor) could reflect variation in professional environment (e.g., Kraft & Papay, 2014), which leads to actual differences in PST development and effectiveness. On the other hand, supervisors may simply be reacting (i.e., giving higher or lower scores to PSTs) to school or classroom conditions that are unrelated to PST performance (e.g., Campbell & Ronfeldt, 2018; Garrett & Steinberg, 2015). Finally, PSTs may sort non-randomly to schools, such that certain schools tend to have higher- or lower-quality PSTs, on average. Finally, PST effects reflect observation score differences between PSTs after accounting for supervisor and school-level heterogeneity. We conceptualize this as the component of observation scores that is attributable to PSTs. We note, however, that this component includes actual quality differences and any race or gender biases that may exist in the scores (Campbell & Ronfeldt, 2018).[7]

To illustrate the utility of using the BLUPs as predictors, we present a simplified conceptual framework in Figure 1. In Panel A, we consider a simple test of concurrent validity that estimates the correlation between PST observation scores and certification exam scores. Because observation scores in part reflect supervisor and school effects—which we hypothesize are unrelated to PST quality—the observed correlation is attenuated relative to the true correlation between PST quality (more specifically, the component of PST quality captured by observational evaluations) and certification exam scores. By partitioning the variance into supervisor, school, and PST effects, we can

[7] One way to check the plausibility of these conceptualizations (e.g., that the supervisor random effect is not picking up variation that should be attributed to PSTs) is to regress the supervisor and school BLUPs on PST characteristics, which we show in Appendix Table A1. Specifically, we model the supervisor (school) BLUPs as a function of the mean PST characteristics within that supervisor (school). If the BLUPs are truly independent of PST quality, we should expect to find no evidence of relationships. Consistent with our expectations, we find no strong evidence of a relationship between supervisor BLUPs and the average characteristics of PSTs to which they are assigned. We do find some evidence of sorting to schools, however, which is unsurprising given that PSTs can choose which district they go to. For comparison, we also show that there are clear relationships between PST BLUPs and their own characteristics.

effectively obtain a disattenuated correlation, which provides a better indication of the extent to which observation scores and certification exam scores capture a common underlying construct of PST quality.

Panel B considers the relationship between PST observation scores and entry into a K–12 teaching position. Here, we hypothesize that the observed correlation may conflate two distinct sets of processes. First, PSTs may be more likely to enter the profession if they receive a positive signal about their effectiveness (in the form of higher observation scores), even if that signal is unrelated to their actual performance. Hiring schools or districts may also respond positively to higher scores, since state law mandates that principals at the clinical teaching placement school receive them. Second, higher-quality PSTs may actually be more or less likely to enter the profession. By modeling entry as a function of supervisor, school, and PST effects, rather than observation scores, we can potentially shed light on the relative importance of each of these processes. Specifically, we propose that any relationship between supervisor effects and entry would more plausibly reflect a signaling mechanism, whereas PST effects more plausibly capture the individual teacher or teaching quality of PSTs.

Predicting Observation Scores and Certification Scores

Our second and third research questions examine the relationships among observation scores, certification scores, and the characteristics of PSTs, placement schools, and supervisors. For observation scores, we estimate our preferred HLM specification described in the previous section, but now add covariates. For certification scores, we estimate models via OLS at the PST level, as we found no evidence of supervisor or school effects.[8]

[8] Specifically, we estimated three-level HLM models with PSTs nested within schools nested within supervisors. Likelihood ratio tests comparing the HLM and OLS specifications were not statistically significant for either exam ($p = 0.86$ and $p = 0.81$ for pedagogy and content, respectively). While not definitive evidence, the lack of supervisor or school effects on certification scores suggests that these variance components in the observation score models are, as we argue in this analysis, unrelated to the performance/quality of PSTs. Specifically, if the presence of supervisor or school effects in the observation scores models reflects that supervisors or schools are actually contributing differentially to PST development, we would expect to observe non-zero effects in the certification score models.

PST characteristics include gender, race/ethnicity, high school grade point average, SAT composite score, family income, transfer student, and first-generation college student. Placement school characteristics include school level, locale, and the proportion of students qualifying for free/reduced-price lunch. We also considered models that included school-level averages of student racial composition, but found a high level of multicollinearity with free/reduced-price lunch status and no evidence of improved model fit. Finally, supervisor characteristics include gender and a flag for whether the supervisor and PST are the same gender.

A potential concern with the HLM models is that they introduce bias through a violation of the assumption that the random effects are uncorrelated with the independent variables. As a check, we run Hausman tests to compare estimates from fixed effects and random effects models at various levels: supervisor, school, and supervisor-by-school. The results, shown in Appendix Table A2, show no evidence that the random effects assumption is violated. Additionally, Appendix Table A3 shows that our findings are robust to a wide range of fixed effects specifications, including those modeling unit-specific trends.

Modeling Entry into K–12 Public Schools

The final piece of our analysis examines the relationship between PST quality and entry into the K–12 public school system in Texas as a full-time teacher. Specifically, we estimate linear probability models of the general form:

$$Y_{ijs} = \beta_1 Quality_i + \beta_2 PST_i + \beta_3 School_j + \tau(Cohort \times Program)_i + \gamma_s + \epsilon_{ijs} \tag{5}$$

where *Quality* is a vector of measures of PST quality, including certification exam scores and PST observation scores or, alternatively, the BLUPs from variance decomposition models. *PST* and *School* are vectors of PST and school characteristics, respectively, which mirror those used in the observation and certification score models. In all models, we include cohort-by-program fixed effects to account for any program-specific differences in hiring dynamics as well as year-to-year shocks to the labor market. We also estimate specifications that include fixed effects for clinical teaching school (γ_s). While our models are inherently descriptive, comparing the results with and without school fixed effects helps to elucidate the extent to which the observed patterns are driven by nonrandom sorting of PSTs to clinical teaching placements. For example, one explanation for a positive relationship between PST quality and entry would be that high-quality PSTs seek out clinical teaching placements in schools or districts that have a strong history of hiring student teachers after they graduate. If the pattern persists conditional on school fixed effects, however, we can feel more confident that this sorting behavior is not the major driver of this relationship. Finally, we cluster standard errors at the school district level.

Results

Our analysis proceeds in two main parts. First, we consider the measurement of PST quality, both in terms of observation scores and certification exam scores. Second, we examine the relationship between PST quality and entry into Texas K–12 public schools as a full-time teacher.

What Do PST Observation Scores Measure?

Table 2 shows unconditional (i.e., without covariates) variance decomposition models for PST observation scores. We start by specifying a two-level model that nests individual observations within PST. Column 2 shifts to a two-level growth model by including observation order (1st through 4th observation) and a squared term to test for non-linearity. We find that scores increase by 0.50 SD with each additional observation, on average. There is no evidence of a non-linear relationship. Columns 3 and 4 expand to three- and four-level models, respectively. Each additional level substantially improves model fit, as both the supervisor and school random effects are large in magnitude and statistically significant. The four-level model in column 4 shows that a large proportion of the total variation in observation scores is between supervisors. The model-based estimate of the standard deviation of supervisor effects is 0.61 SD, compared to 0.22 SD for schools and 0.37 SD for PSTs.

Columns 5 through 7 test for heterogeneity in the *improvement* trajectory of PSTs by adding random effects for the slope (observation order). We also allow for an unstructured covariance matrix to examine whether a higher initial starting level (i.e., the random intercepts) is correlated with higher or lower growth. We focus on our preferred specification in column 7, which includes random intercepts and slopes for each level. We find strong evidence of heterogeneity in both intercepts and slopes at each level, but particularly for supervisors. The standard deviation of supervisor intercepts is 0.98 SD and the standard deviation of supervisor slopes is 0.21 SD. However, these parameters have a large negative correlation ($r = -0.85$), meaning that supervisors essentially trade off between giving higher initial scores versus increased scores with each additional observation. In sum, differences between supervisors in terms of their rating standards accounts for nearly two-thirds of the total variation in observation scores, whereas only 9% is attributable to schools and 12% to PSTs themselves.

Figure 2 shows graphically the results of column 7. Specifically, we obtain best linear unbiased predictions (BLUPs) of the random effects for each level, then construct the estimated slopes and intercepts for each PST by adding their individual BLUP to the BLUPs for their supervisor and school. While the plot includes all of the roughly 2000 PSTs, we also color-code PSTs for three supervisors (A, B, and C) and schools within supervisor A to illustrate the logic of the variance decomposition. First, we can see the strong negative correlation between the intercept and slope. PSTs who initially receive lower scores “improve” at higher rates. However, PSTs are tightly clustered by supervisor, which corresponds to the large variance component for supervisors. In other words, arbitrary differences between supervisors’ rating standards explains much of the variation in PST observation scores. The school-level variance is illustrated by clustering among schools that are nested within supervisors. As we show for supervisor A, school placement helps explain additional variation within supervisor, as PSTs within the same school tend to be clustered together. Finally, the PST-level variation is represented by the differences among PSTs within the same school. However, the amount of school-level and PST-level variation is dwarfed by the differences among supervisors.

Measures of PST Quality and the Characteristics of PSTs and School Placements

Table 3 shows results for predicting observation scores as a function of PST characteristics, clinical teaching school characteristics, and supervisor characteristics. Specifically, we estimate HLM models that follow the specification shown in column 7 of Table 2. We also check the robustness of these results to various fixed effects specifications, which are shown in Appendix Table A3.

Column 1 shows baseline differences by PST race and gender. Similar to prior work examining in-service teachers (e.g., Campbell & Ronfeldt, 2018), we find that men score lower than women (-0.26 SD) and PSTs of color score lower than white PSTs (-0.09 SD). Column 2 incorporates additional PST characteristics, including high school GPA, indicators for transfer student and first-generation student, SAT composite score, and family income. Including these characteristics slightly reduces both the race and gender gap. We find a positive relationship between high school GPA and observation scores; a 1-unit increase in GPA (i.e., moving from a 2.5 to a 3.5) is associated with a 0.2 SD increase in student-teaching observation scores, on average. However, we find no relationship between observation scores and SAT scores.[9] There is a small gap between transfer and non-transfer students (-0.06 SD), but no relationship between observation scores and first-generation status. Finally, we find a U-shaped relationship for family income. PSTs from poorer families (< \$60,000) score 0.11 SD lower than their peers from middle-income families (\$60,000–\$100,000), as do PSTs from families with annual incomes

[9] This null finding holds even if we remove the other covariates from the model.

above $100,000 (-0.07 SD).

Columns 3 and 4 add school placement characteristics and supervisor gender, respectively. Again, we extend findings from prior work on in-service teachers (Campbell & Ronfeldt, 2018; Steinberg & Garrett, 2016) to the pre-service context by showing that PSTs' observation scores appear to measure, in part, the characteristics of schools and the students they serve. Specifically, we find that PSTs in lower-income schools tend to receive lower observation scores, as do PSTs in rural placements. Turning to supervisor gender, we find no evidence of a main effect or the interaction between supervisor and PST gender (i.e., gender matching). However, we note the relatively large standard errors for these coefficients, given the low number of male PSTs and supervisors.

Table 4 shows results for predicting PST certification scores on pedagogy and content exams. Columns 1 and 6 only include PST characteristics and placement school characteristics. We found no consistent evidence of a relationship with school characteristics and omit them from the table for brevity (see Appendix Table A4 for full results). Comparing across both exams, we find large, positive relationships with high school GPA and SAT scores. Whereas there was no relationship between SAT scores and observation scores, the exam score difference between PSTs with SAT composite scores below 1000 and above 1200 is 1 SD for pedagogy and 1.3 SD for content. We also find that transfer students perform substantially worse on both exams, even conditional on their observable academic characteristics. PSTs of color score slightly lower on both exams, on average, though the difference is only statistically significant for content scores. Finally, we observe gender gaps for both exams, but they run in opposite directions. Men score 0.17 SD lower than women on the pedagogy exam but score 0.31 SD higher on the content exam, on average.

Columns 2 and 7 examine the relationship between PSTs' average observation score and certification scores, without controlling for other PST characteristics. We find a positive and statistically significant correlation for both pedagogy and content, but the magnitudes are small. Columns 3 and 8 show that these correlations are somewhat attenuated after including PST characteristics.

One explanation for the weak relationship between teacher observations and certification exam scores is that the observation scores contain substantial noise from the supervisor- and school-level variance components. Specifically, Table 2 showed that the bulk of the variation in observation scores reflects factors that are unrelated to the "performance" of PSTs themselves. Conceptually, if we decompose the observation scores into the PST, school, and supervisor effects, we should find the relationship between observation scores and certification scores is explained by PST effects, rather than school or supervisor effects. Further, to the extent that the school and supervisor effects are unrelated to PST performance/knowledge, isolating the PST effect should help to disattenuate the estimated relationship with certification scores.

To test this explanation, we first compute the best linear unbiased predictions (BLUPs) for the PST, school, and supervisor random effects from the variance decomposition model. For the sake of parsimony, we use the intercepts-only model (Table 2, Column 4). Models that include BLUPs for intercepts and slopes are extremely similar but less precise, given the very high correlations between the slope and intercept parameters for supervisor and school. The BLUPs are, effectively, predictions (shrunken towards the mean) for the contribution of PST, school, and supervisor to a PST's observation score. We then re-scale the BLUPs into PST-, school-, and supervisor-level standard deviations using the model-based estimates of the standard deviation of each random effect.[10] Finally, we re-estimate the model using the BLUPs in place of the average observation scores, which isolates the sources of variation from supervisors, schools, and PSTs.[11]

[10] We use the model-based estimate of the standard deviation of the PST/school/supervisor effects instead of the standard deviation of the distribution of the BLUPs. The variance components model provides unbiased estimates of the magnitude of the effects, whereas the standard deviation of the BLUP distribution will be a downwardly biased estimate of the true standard deviation due to shrinkage.

[11] As an additional check, we run models that treat the PST, school, and supervisor effects as fixed (i.e., individual parameters to be estimated), then use these fixed effects estimates instead of the BLUPs. The results, as shown in Appendix Table **??**, are similar.

The results of incorporating these BLUPs are shown in columns 4–5 and 9–10. First, we repeat the specification shown in column 2 (without PST characteristics). For pedagogy scores (column 4) we find a positive correlation with PST BLUP (i.e., the source of variation in observation scores that is attributable to PSTs). Further, the magnitude of the relationship is much larger than would be implied by column 2; a 1 SD increase in PST quality (as measured by observation scores) is associated with a 0.18 SD increase in pedagogy exam score. By contrast, neither the supervisor nor school BLUPs are correlated with pedagogy exam scores, which is consistent with our hypothesis that these effects do not contain information about PST quality. Column 5 shows that this positive correlation remains even after controlling for other PST characteristics, such as GPA and SAT scores. Substantively, this suggests that the pedagogy exam and observational evaluations do contain some common signal about PST quality that is not captured by observable academic or demographic characteristics. However, the vast majority of the explained variation in the pedagogy exam is attributable to high school GPA and SAT scores, rather than a PST's performance as a student teacher.

Turning to content exam scores in columns 9 and 10, we find a slightly different story. Column 9 shows that both the PST and school BLUPs are positively associated with context exam scores, though the magnitudes are somewhat small. However, once we condition on PST characteristics in column 10, neither of these relationships are statistically significant. In other words, we find little evidence that clinical teaching observation scores contain a signal about content knowledge (as measured by certification exam scores) that is independent from the signal provided by general measures of academic achievement.

Pre-Service Teacher Entry into Texas K–12 Public Schools

While the first part of our analysis focused on the measurement of PST quality, we now turn to entry into the profession and its relationship with PST quality. We start with

Table 5, which estimates linear probability models for entry into the K–12 public school teacher workforce. In terms of PST characteristics, the most important predictor of entry is SAT score. There is a fairly steep downward gradient, such that PSTs with higher SAT scores are less likely to enter K–12 public schools in Texas. Compared to PSTs with fairly average scores (1000–1200 out of 1600), those scoring below 1000 are 5.5 percentage points more likely to enter upon graduation, while PSTs scoring above 1200 are 5.4 percentage points *less* likely to enter (column 2). We also find that first-generation students are more likely to enter than non-first-generation students, while individuals from low-income families are less likely to enter than others from middle- or high-income families. When we add school fixed effects in column 3, we find that these relationships are attenuated somewhat, though there is also a sizeable increase in the standard errors.

We find no consistent relationships between clinical placement school characteristics and the probability of entering. Notably, we do not find evidence of a gender difference in entry, though the standard errors are fairly large due to the low number of male PSTs. Similarly, while PSTs of color enter at slightly lower rates, the difference is not statistically significant at conventional levels. Finally, while transfer students score lower on both observational evaluations and certification exams, we find no evidence that they are more or less likely to enter K–12 public schools in Texas.

Turning to observation and certification scores, column 2 shows a positive relationship between a PST's average observation score as a clinical teacher and their probability of entry, even conditional on other PST characteristics. Specifically, a 1 SD increase in observation scores is associated with a 3.4 percentage point increase in the probability of entry. This pattern does not appear to be explained by nonrandom sorting to clinical teaching placements, as the coefficient from the school fixed effects model in column 3 is identical. We also find a small positive association between pedagogy certification exam score and entry ($\beta = 0.018$), though not with the content exam.

In column 4, we replace average observation score with the variance components

(BLUPs) for supervisor, school, and PST. We find that while the coefficients are positive for each of the BLUPs, the magnitude is largest for supervisor. This suggests that the positive selection of higher-scoring PSTs into the teacher workforce is explained less by actual PST quality and more by arbitrary differences in supervisor rating standards. When we isolate the PST-level source of variation, we find only a slight positive relationship. One explanation is that the pattern is driven by hiring schools/districts or PSTs themselves reacting to the observation score as signal of PST quality. In this scenario, the supervisor BLUP will be more important than the school or PST BLUP because of the large contribution of supervisors to the observation score. Contrast this with an alternative scenario where observation scores are not used in the hiring process and are simply a proxy measure for PST quality. In this case, we would expect to find no relationship between supervisor BLUP and entry, since supervisor-level variation in observation scores is unrelated to PST quality. An additional way to illustrate this point is to simply unscale the BLUPs (i.e., do not divide by the supervisor-, school-, and PST-level standard deviations) and compare the coefficients from the same model. If stakeholders (i.e., schools, districts, PSTs) are responding to the score itself, these coefficients should be similar in magnitude. We show this in column 5. While the unscaled supervisor BLUP remains slightly larger than the school or PST BLUP, we cannot reject the null hypothesis that these coefficients are jointly equal ($p = 0.87$).

Examining Alternative Entry Outcomes

One question raised by Table 5 is the extent to which the relationship between observation scores and PST entry into the teaching profession reflects supply-side versus demand-side dynamics in the labor market. In other words, are higher-scoring PSTs more likely to become teachers or are schools/districts more likely to hire high-scoring PSTs? While we lack district application data that could investigate this question (Engel, Jacob, & Curran, 2014), we can provide suggestive evidence by examining a set of alternative entry outcomes that we hypothesize are more likely to reflect demand-side behavior. Specifically, we consider *immediate* entry (i.e., becoming a full-time teacher in the school-year immediately following graduation) and entry into the same school as where a PST completed their clinical teaching. Whereas we observe that 83% of PSTs enter K–12 public schools in Texas, 74% enter immediately and 17% of those that enter are at the same school as their clinical teaching placement.

We show the results for these outcomes in Table 6. For immediate entry, we find patterns similar to Table 5, but nearly all of the relationships are larger in magnitude. One notable difference is high school GPA, where the coefficient in the baseline model (column 1) has more than doubled in magnitude. We also observe a substantially steeper SAT gradient. Both of these relationships reinforce that PSTs with stronger high school academic backgrounds are less likely to enter the teacher workforce. However, these patterns also run counter to our expectation that immediate entry would more strongly reflect demand-side behavior. Instead, these results may suggest a supply-side dynamic whereby PSTs with better academic credentials are more likely to pursue opportunities outside of K–12 public schools.

For observation and certification scores, the results are consistent with our expectation that higher-quality PSTs are more likely to immediately enter the teacher workforce. Both of these relationships are larger in magnitude than for the baseline entry models. Interestingly, when we decompose the observation score, we find that the PST BLUP is comparatively more important in explaining immediate entry than simple entry.

The last three columns in Table 6 show results for predicting, conditional on entry, hiring at the same school as where the PSTs completed their clinical teaching. Here, we find no evidence of a relationship with high school GPA or SAT score, nor with certification scores. However, there is a large, positive association between entry into the same school and PST observation scores. In column 4, a 1 SD increase in observation score is associated with a 5.9 percentage point (35%) increase in the probability of being hired at the same school. The pattern is nearly identical when we restrict to comparisons of PSTs within the same school (column 5). Turning to the variance components in column 6, we again find that the PST BLUP is a relatively more important predictor of entering the same school than simple entry, though here the pattern is even starker than for immediate entry.

That observation scores are a salient predictor of entering the same school as clinical teaching is perhaps not a surprise, given state law requires that principals at the clinical teaching placement school have access to these scores. The results in Table 6 suggest that principals may draw on these scores to identify effective PSTs to hire as full-time teachers.[12]

As a final analysis, we examine the characteristics of the full-time teaching school among PSTs who enter Texas K–12 public schools. Specifically, we construct an "advantage index" using principal components analysis to reduce dimensionality.[13] The index is scaled to have a mean of zero and a standard deviation of one, where higher scores indicate a more "advantaged" school in terms of demographic characteristics (e.g., fewer low-income students, fewer students from historically marginalized backgrounds). Column 1 shows that among PSTs who enter Texas K–12 public schools, there are fairly substantial relationships between school advantage and PST characteristics. Male and nonwhite PSTs enter substantially less advantaged schools, on average, as do PSTs from lower-income backgrounds. For instance, the difference in advantage index between male and female

[12] One question raised by the results in Table 6 is whether the association between observation scores and entry into K–12 public schools operates completely through the same-school hiring mechanism. We can provide suggestive evidence by re-estimating our entry models on the restricted sample of PSTs who do not enter the same school as their clinical teaching placement. The results in Appendix Table A6 show that the pattern holds. However, we observe some attenuation in the relationship when dropping all PSTs who stayed in the same district as their clinical teaching (37% of all hires). The estimated coefficient for average observation remains positive, but it is roughly half the size of the main estimates in Table 5. This suggests that observation scores may also be associated with same-district hiring, which we confirm in Appendix Table A7.

[13] We construct the index using the following school characteristics: title I schoolwide program eligibility, proportion of FRPL students, proportion of students by race/ethnicity, and locale type. We use the population of Texas K–12 public schools in 2012–2018 to perform the PCA and produce a predicted score for the first component. Then, we standardize this predicted score within year and school level (elementary, middle, high, other). Finally, we impute for 2019 using 2018 data. Appendix Table A8 shows results from models that predict individual characteristics of entry schools instead of the index. Appendix Table A9 shows a correlation matrix for the advantage index and school characteristics.

graduates is 0.38 SD. We also find positive relationships for high school GPA and transfer student status. Notably, while SAT scores are an important predictor of entry, they are not correlated with the characteristics of the full-time school placement, conditional on entry.

Column 2 adds controls for the characteristics of the PST's placement school for student teaching. Given prior findings that new teachers tend to find jobs in schools that are demographically similar to where they completed their clinical teaching (Krieg et al., 2016), adding these characteristics helps to reveal the extent to which sorting patterns to full-time teaching school shown in column 1 can be explained by the initial sorting of PSTs to clinical teaching placements. Indeed, column 2 shows that the characteristics of the clinical teaching school are correlated with the advantage level of the teacher's initial job placement. Nonetheless, the estimated relationships between school advantage and PST characteristics are largely unchanged, except for a complete attenuation of the transfer student coefficient and partial attenuation for family income. Column 3 adds average observation score and certification exam scores. The estimated coefficients are small in magnitude.[14] When we isolate to variation within the same clinical teaching placement school in column 4, we do find some changes in the coefficients for the PST characteristics. Both the gender and race patterns are somewhat attenuated, while the positive correlation between high school GPA and advantage index is larger in magnitude. We again find no evidence of relationship with average observation score. However, when we replace observation score with the variance components, we find a positive relationship between the school-level BLUP and the advantage index of a PST's full-time school. Given the small proportion of variance in observation scores attributable to schools, this relationship is essentially washed out when using the average observation score.

[14] While the certification exam scores are individually significant at the 90% level, they are not jointly significant nor are they significant when included in the model individually.

Discussion, Implications, and Limitations

Mounting evidence points to teacher education as a key policy lever for improving the quality of the teacher workforce. Understandably, researchers have increased attention towards identifying high-quality TEPs and how the training or experiences PSTs receive may influence their effectiveness as full-time teachers. Yet our understanding of the variation within TEP—how individuals range in PST quality—and its association with labor outcomes remains underdeveloped. We provide new evidence on the measurement of PST quality and its connection to entry into the public teaching workforce.

Our results show that little of the variation in PST observation evaluations during clinical teaching is attributable to PST quality. Instead, differences in PSTs' scores largely reflect differences in the rating standards of their field supervisors. Consistent with prior findings from the in-service context (Campbell & Ronfeldt, 2018), we also find race and gender gaps in observation scores. On average, men score 0.26 SD below women, and PSTs of color score 0.09 SD below their white classmates. While the magnitude of these gaps may seem modest, they are large relative to the amount of between-PST variation. The estimated gender and race gaps translate to 74% and 27% of a standard deviation in the distribution of PST quality, respectively.[15] These gaps largely persist across a wide range of model specifications, including those that adjust for high school achievement, clinical teaching placement school context, and certification exam scores.

We find only a weak correlation between observation scores and certification exam scores, in part because so much of the variation in observation scores is unrelated to PST quality. Isolating the component of observation scores attributable to PSTs strengthens these correlations somewhat. However, certification scores appear mainly to proxy for academic achievement, rather than instructional effectiveness or readiness (at least as measured by observational evaluations).

[15] We reach these numbers by comparing the estimated gaps in Table 3 Column 1 to the model-based estimate of the standard deviation of PST random intercepts in Table 2 Column 4.

While more than 80% of teacher education students in our sample eventually become teachers in Texas K–12 public schools, we find some evidence that lower-quality PSTs—both in terms of observation scores and certification exam scores—are less likely to enter the teacher workforce. Consistent with Goldhaber et al. (2014), the relationship between observation scores and entry is particularly salient for PSTs to remain in the same school as their clinical teaching, which suggests that schools or districts may use these placements to screen for quality and fit of prospective teachers. Indeed, state law requires that PST observation scores are provided to principals at the clinical teaching site. Our results here provide additional evidence documenting that principals behave strategically to manage human capital in their schools (Goldring et al., 2015; Grissom & Bartanen, 2019; Rockoff, Staiger, Kane, & Taylor, 2012). A key implication, then, is that policies that provide administrators with information about PST quality can help to incentivize strategic decision-making. That said, the quality of the measure matters—arming principals or districts with a measure that lacks validity or reliability may undercut efforts to improve the quality of their workforce.

Echoing prior work from the in-service context, the existence of race and gender means there there may also be unintended consequences of observational evaluation systems (Campbell & Ronfeldt, 2018), particularly when those scores can be used for hiring and retention decisions. Beyond basic fairness, this concern is particularly relevant given calls to increase the diversity of the teacher workforce (e.g., U.S. Department of Education, 2016). While our results most strongly link observation scores to entry into the same school as clinical teaching, negative signals received by PSTs in the form of lower observation scores may in part discourage them from ever entering K–12 schools.

These results have distinct implications for teacher education programs, which occupy a critical role as the primary training ground for new teachers. Effectively fulfilling this role requires accurately measuring the performance of PSTs to provide targeted feedback and to identify those who need additional support. Our analysis suggests that attention to the validity and reliability of observational evaluations—specifically, reducing supervisor heterogeneity in rating standards and addressing race and gender biases—are important to improving the measurement of PST quality. This might include more intensive and standardized training for field supervisors or using multiple raters for evaluations (Henry et al., 2013). Online professional development or frequent communication with supervisors could be first steps towards changing this structure, echoing best practices of investing in evaluators (Weisberg, Sexton, Mulhern, & Keeling, 2009).

Given these implications, we also acknowledge that this study faces several limitations, each of which suggests important avenues for future work. First, our analysis of PST quality is based on a single teacher education program in Texas. While examining a specific context affords us access to unique data with a rich set of variables collected before, during, and after each PST's time in the program, our results may not generalize to other TEPs in Texas or nationally. Given the limited prior work in this area, it is important for future studies to establish whether these patterns hold in other contexts. That said, we replicate a number of patterns from other contexts, which provides some suggestion that our findings may generalize to other teacher education programs.

Second, while all of our analyses are fundamentally descriptive, we reiterate that the causal mechanisms driving the PST entry and allocation results are unclear. The higher rates of entry among PSTs with higher observation scores could reflect PST preferences, hiring school preferences, or both. As an example, while we posit that principals' preferences to hire effective teachers helps to explain the tendency for higher-scoring PSTs to become full-time teachers in their clinical teaching school, this pattern could also reflect greater willingness among higher-scoring PSTs to accept a job offer in that school. More generally, our entry results point to a number of patterns that are worthy of additional exploration. The lower entry rates of PSTs with high SAT scores and those from low-income families are consistent with selection away from public school teaching towards occupations with higher status or pay (Guarino et al., 2006). Future work should continue to investigate the pipeline from TEPs to full-time teaching placements, possibly by examining other types of field experiences within preparation and their relationship with labor market outcomes, which can help to inform policies that both increase the overall quality of the teacher workforce and remedy teacher quality gaps between advantaged and disadvantaged schools.

Finally, given their importance as a source of variability in observation scores, we are limited in that we have little information about field supervisors, including their demographic information and prior work experiences. These characteristics may be important in explaining the substantial heterogeneity in rating standards. Further, while supervisors were required to complete a state-approved observation training, we do not know when that training was obtained or the extent to which the training varied by geographic context. Relatedly, while not formally involved in the evaluation process for PSTs in our sample, we acknowledge the importance of mentor teachers in shaping PSTs' experiences in their clinical placement. Given evidence that PSTs who have an effective mentor teacher tend to be more effective themselves as full-time teachers (Goldhaber, Krieg, & Theobald, 2020; Ronfeldt, Brockman, & Campbell, 2018), it is important to understand the extent to which mentor teachers may also influence PSTs' entry decisions.

References

American Association of Colleges for Teacher Education. (2018). *A Pivot Toward Clinical Practice, Its Lexicon, and the Renewal of Educator Preparation* (Tech. Rep.).

Anderson, L. M., & Stillman, J. A. (2013). Student Teaching's Contribution to Preservice Teacher Development: A Review of Research Focused on the Preparation of Teachers for Urban and High-Needs Contexts. *Review of Educational Research*, *83*(1), 3–69.

Angrist, J. D., & Guryan, J. (2008). Does teacher testing raise teacher quality? Evidence from state certification requirements. *Economics of Education Review*, *27*(5), 483–503.

Bastian, K. C., Patterson, K. M., & Pan, Y. (2018). Evaluating Teacher Preparation Programs With Teacher Evaluation Ratings: Implications for Program Accountability and Improvement. *Journal of Teacher Education*, *69*(5), 429–447.

Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., Pianta, R. C., & Qi, Y. (2012). An Argument Approach to Observation Protocol Validity. *Educational Assessment*, *17*(2-3), 62–87.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher Preparation and Student Achievement. *Educational Evaluation & Policy Analysis*, *31*(4), 416–440.

Campbell, S. L., & Ronfeldt, M. (2018). Observational evaluation of teachers: Measuring more than we bargained for? *American Educational Research Journal*, *55*(6), 1233–1267.

Clift, R. T., & Brady, P. (2005). Research on Methods Courses and Field Experiences. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the aera panel on reseach and teacher education* (pp. 309–424).

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). *How and why do teacher credentials matter for student achievement?* Cambridge.

Cohen, J., & Goldhaber, D. (2016). Building a More Complete Understanding of Teacher Evaluation Using Classroom Observations. *Educational Researcher*, *45*(6), 378–387.

D'Agostino, J. V., & Powers, S. J. (2009). Predicting teacher performance with test scores and grade point average: A meta-analysis. *American Educational Research Journal*, *46*(1), 146–182.

Darling-Hammond, L. (2014). Strengthening Clinical Preparation: The Holy Grail of Teacher Education. *Peabody Journal of Education*, *89*(4), 547–561.

Engel, M., Jacob, B. A., & Curran, F. C. (2014). New evidence on teacher labor supply. *American Educational Research Journal*, *51*(1), 36–72.

Feiman-Nemser, S., & Buchmann, M. (1987). When is student teaching teacher education? *Teaching and Teacher Education*, *3*(4), 255–273.

Gansle, K. A., Noell, G. H., & Burns, J. M. (2012). Do Student Achievement Outcomes Differ Across Teacher Preparation Programs? An Analysis of Teacher Education in Louisiana. *Journal of Teacher Education*, *63*(5), 304–317.

Garrett, R., & Steinberg, M. P. (2015). Examining Teacher Effectiveness Using Classroom Observation Scores. *Educational Evaluation and Policy Analysis*, *37*(2), 224–242.

Gitomer, D. H. (2007). *Teacher Quality in a Changing Policy Landscape: Improvements in the Teacher Pool* (Tech. Rep.). Princeton, NJ: Educational Testing Service.

Gitomer, D. H., Martínez, J. F., Battey, D., & Hyland, N. E. (2020). Assessing the Assessment: Evidence of Reliability and Validity in the edTPA. *American Educational Research Journal*.

Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources*, *42*(4), 765–794.

Goldhaber, D. (2019). Evidence-Based Teacher Preparation: Policy Context and What We Know. *Journal of Teacher Education*, *70*(2), 90–101.

Goldhaber, D., Cowan, J., & Theobald, R. (2017). Evaluating Prospective Teachers: Testing the Predictive Validity of the edTPA. *Journal of Teacher Education*, *68*(4), 377–393.

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

Goldhaber, D., & Hansen, M. (2010). Race, Gender, and Teacher Testing: How Informative a Tool Is Teacher Licensure Testing? *American Educational Research Journal*, *47*(1), 218–251.

Goldhaber, D., Krieg, J., & Theobald, R. (2014). Knocking on the door to the teaching profession? Modeling the entry of prospective teachers into the workforce. *Economics of Education Review*, *43*, 106–124.

Goldhaber, D., Krieg, J., & Theobald, R. (2020). Effective like me? Does having a more productive mentor improve the productivity of mentees? *Labour Economics*, *63*(May 2019), 101792.

Goldhaber, D., Liddle, S., & Theobald, R. (2013). The gateway to the profession: Assessing teacher preparation programs based on student achievement. *Economics of Education Review*, *34*, 29–44.

Goldhaber, D., & Ronfeldt, M. (2020). Toward Causal Evidence on Effective Teacher Preparation. In J. E. Carinci, C. Jackson, & S. J. Meyer (Eds.), *Linking teacher preparation program design and implementation to outcomes for teachers and students* (pp. 211–236). Information Age Publishing.

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

Grissom, J. A., & Loeb, S. (2017). Assessing Principals' Assessments: Subjective Evaluations of Teacher Effectiveness in Low- and High-Stakes Environments. *Education Finance and Policy*, *12*(3), 369–395.

Grossman, P., Cohen, J., Ronfeldt, M., & Brown, L. (2014). The Test Matters: The Relationship Between Classroom Observation Scores and Teacher Value Added on Multiple Types of Assessments. *Educational Researcher*, *43*(6), 293–303.

Grossman, P., Hammerness, K. M., McDonald, M., & Ronfeldt, M. (2008). Constructing Coherence. *Journal of Teacher Education*, *59*(4), 273–287.

Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for measure: The relationship between measures of instructional practice in middle school english language arts and teachers' value-added scores. *American Journal of Education*, *119*(3), 445–470.

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

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

Henry, G. T., Campbell, S. L., Thompson, C. L., Patriarca, L. A., Luterbach, K. J., Lys, D. B., & Covington, V. M. (2013). The Predictive Validity of Measures of Teacher Candidate Programs and Performance: Toward an Evidence-Based Approach to Teacher Preparation. *Journal of Teacher Education*, *64*(5), 439–453.

Henry, G. T., Purtell, K. M., Bastian, K. C., Fortner, C. K., Thompson, C. L., Campbell, S. L., & Patterson, K. M. (2014). The Effects of Teacher Entry Portals on Student Achievement. *Journal of Teacher Education*, *65*(1), 7–23.

Jacob, B. A., & Walsh, E. (2011). What's in a rating? *Economics of Education Review*, *30*(3), 434–448.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources*, *46*(3), 587–613.

Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2015). Teacher preparation programs and teacher quality: Are there real differences across programs? *Education Finance and Policy*, *10*(4), 508–534.

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

Krieg, J. M., Theobald, R., & 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), 364–388.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis. *Educational Evaluation and Policy Analysis*, *24*(1), 37–62.

Memory, D. M., Coleman, C. L., & Watkins, S. D. (2003). Possible Tradeoffs in Raising Basic Skills Cutoff Scores for Teacher Licensure. *Journal of Teacher Education*, *54*(3), 217–227.

Mihaly, K., McCaffrey, D., Sass, T. R., & Lockwood, J. R. (2013). Where you come from or where you go? Distinguishing between school quality and the effectiveness of teacher preparation program graduates. *Education Finance and Policy*, *8*(4), 459–493.

National Center for Education Statistics. (2017). *Required testing for initial certification of elementary and secondary school teachers, by type of assessment and state: 2015 and 2016.*

National Center for Education Statistics. (2018). *Characteristics of Public School Teachers Who Completed Alternative Route to Certification Programs* (Tech. Rep.).

National Council for Accreditation of Teacher Education. (2010). *Transforming Teacher Education through Clinical Practice: A National Strategy to Prepare Effective*

Teachers. Report of the Blue Ribbon Panel on Clinical Preparation and Partnerships for Improved Student Learning. (Tech. Rep.).

Rockoff, J. E., Staiger, D. O., Kane, T. J., & Taylor, E. S. (2012). Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools. *The American Economic Review*, *102*(7), 3184–3213.

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

Ronfeldt, M., & Campbell, S. L. (2016). Evaluating Teacher Preparation Using Graduates' Observational Ratings. *Educational Evaluation and Policy Analysis*, *38*(4), 603–625.

Ronfeldt, M., Reininger, M., & Kwok, A. (2013). Recruitment or Preparation? Investigating the Effects of Teacher Characteristics and Student Teaching. *Journal of Teacher Education*, *64*(4), 319–337.

Shuls, J. V. (2018). Raising the Bar on Teacher Quality: Assessing the Impact of Increasing Licensure Exam Cut-Scores. *Educational Policy*, *32*(7), 969–992.

Slick, S. K. (1998). The university supervisor: A disenfranchised outsider. *Teaching and Teacher Education*, *14*(8), 821–834.

Steinberg, M. P., & Garrett, R. (2016). Classroom Composition and Measured Teacher Performance: What Do Teacher Observation Scores Really Measure? *Educational Evaluation and Policy Analysis*, *38*(2), 293–317.

Strunk, K. O., Weinstein, T. L., & Makkonen, R. (2014). Sorting out the signal: do multiple measures of teachers' effectiveness provide consistent information to teachers and principals? *Education Policy Analysis Archives*, *22*(0), 1–42.

U.S. Department of Education. (2016). *The State of Racial Diversity In the Educator Workforce* (Tech. Rep.). Washington, D.C.: U.S. Department of Education.

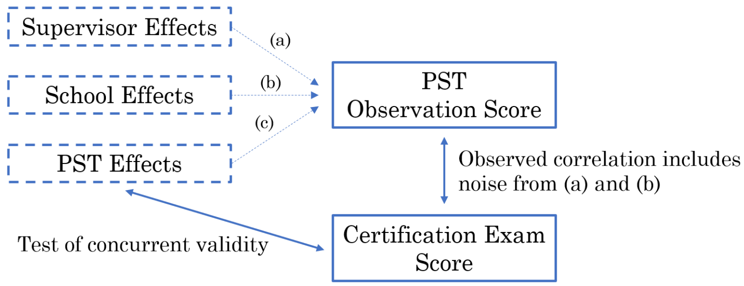
Wasburn-Moses, L., & Noltemeyer, A. (2018). Effectiveness of Campus Mentors, an alternative school–university partnership. *Preventing School Failure*, *62*(3), 190–197.

Wayne, A. J., & Youngs, P. (2003). Teacher Characteristics and Student Achievement Gains: A Review. *Review of Educational Research*, *73*(1), 89–122.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness* (Tech. Rep.). The New Teacher Project.

Youngs, P., Odden, A., & Porter, A. C. (2003). State Policy Related to Teacher Licensure. *Educational Policy*, *17*(2), 217–236.

(a) Examining the Relationship between PST Observation Scores and Certification Exam Scores



(b) Examining the Relationship between PST Observation Scores and Entry into a K–12 Teaching Position

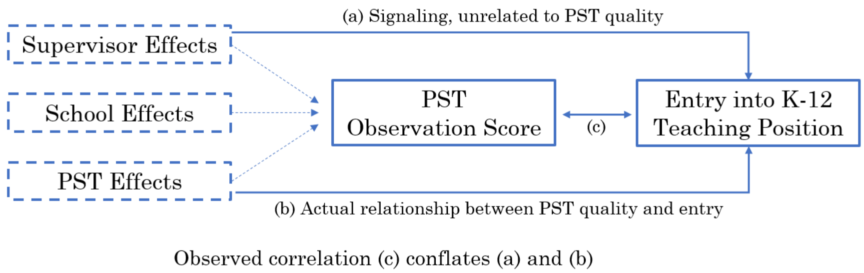


Figure 1. Conceptual Framework for Variance Decomposition of PST Observation Scores

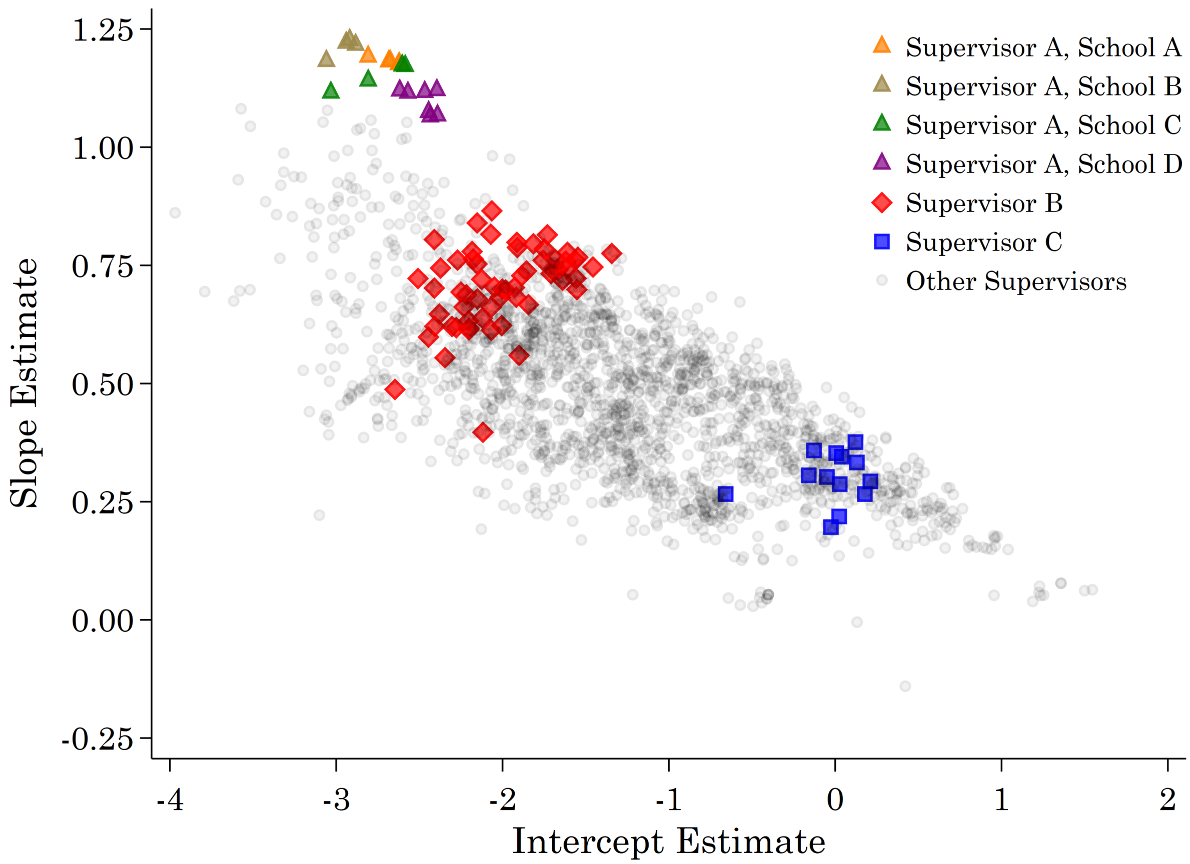


Figure 2. Pre-Service Teacher Random Intercepts and Slopes from Observation Score Models

Notes: Plot shows slope and intercept estimates for the full sample of PSTs from the HLM model estimated in Table 2 Column 7. The intercept is calculated as the sum of four components: constant term, PST random intercept, school random intercept, and supervisor random interept. The slope is calculated at the sum of five components: coefficient estimates for observation order and squared term, PST random slope, school random slope, and supervisor random slope. Each PST is represented by a single point. The intercept represents their estimated score (in standard deviation units) for their first observation, while the slope represents the average increase in each subsequent observation (up to four total). Color-coding is added to emphasize clustering by supervisor and school.

Table 1
Descriptive Statistics

	N	Mean	SD	Min	Max
Pre-Service Teacher Characteristics					
Male	2140	0.05			
Asian	2140	0.01			
Black	2140	0.02			
Hispanic/Latinx	2140	0.11			
White	2140	0.84			
Other Race/Ethnicity	2140	0.02			
High School GPA	2140	3.47	0.29	2.59	4.00
Transfer Student	2140	0.36			
First-Generation Student	2140	0.23			
SAT Composite Missing	2140	0.14			
SAT Composite < 1000	2140	0.18			
SAT Composite 1000–1200	2140	0.47			
SAT Composite > 1200	2140	0.21			
Family Income Missing	2140	0.09			
Family Income < $60,000	2140	0.17			
Family Income $60,000–$100,000	2140	0.30			
Family Income > $100,000	2140	0.44			
Average Observation Rating (1–4 scale)	2140	3.30	0.35	2.09	4.00
Average Observation Rating (standardized)	2140	0.04	0.97	-3.65	2.49
Certification Score (Pedagogy)	2112	0.06	0.96	-4.28	2.24
Certification Score (Content)	2136	0.15	0.94	-4.74	2.00
Entered Texas Public Schools	2140	0.83			
Immediately Entered Texas Public Schools	2140	0.74			
Entered Same District as Clinical Teaching	1717	0.37			
Entered Same School as Clinical Teaching	1717	0.17			
Advantage Index of Full-Time School	1711	0.47	1.01	-1.71	2.55
Pre-Service School Characteristics					
Total Enrollment	2078	730	372	117	3577
Proportion FRPL Students	2078	0.43	0.27	0.01	1.00
Proportion Asian Students	2078	0.06	0.09	0.00	0.80
Proportion Black Students	2078	0.12	0.11	0.00	0.88
Proportion Hispanic/Latinx Students	2078	0.33	0.19	0.03	0.99
Proportion White Students	2078	0.45	0.21	0.00	0.92
Proportion Other Race/Ethnicity Students	2078	0.03	0.02	0.00	0.13
Elementary School	2086	0.60			
Middle School	2086	0.36			
High School	2086	0.01			
Other Level School	2086	0.02			
Urban School	2090	0.36			
Suburban School	2090	0.30			
Town/Rural School	2090	0.34			

Table 2

Variance Decomposition of Pre-Service Teacher Observation Scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Observation Order		0.495***	0.495***	0.495***	0.493***	0.458***	0.430***
		(0.027)	(0.027)	(0.027)	(0.022)	(0.023)	(0.029)
Observation Order^2		-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
		(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Constant	-0.092***	-1.297***	-1.074***	-1.066***	-1.043***	-0.969***	-0.891***
	(0.016)	(0.033)	(0.039)	(0.064)	(0.064)	(0.065)	(0.092)
Random Effects Parameters (SD)							
Supervisor Intercept ($\sigma_{\zeta_{000s}}$)				0.610***	0.642***	0.601***	0.983***
				(0.044)	(0.046)	(0.045)	(0.068)
Supervisor Slope ($\sigma_{\zeta_{100s}}$)							0.211***
							(0.016)
Correlation ($\zeta_{000s}, \zeta_{100s}$)							-0.846***
							(0.032)
School Intercept ($\sigma_{\mu_{00js}}$)			0.513***	0.222***	0.232***	0.644***	0.359***
			(0.022)	(0.019)	(0.019)	(0.032)	(0.027)
School Slope ($\sigma_{\mu_{10js}}$)						0.213***	0.086***
						(0.008)	(0.009)
Correlation (μ_{00js}, μ_{10js})						-0.940***	-0.793***
						(0.012)	(0.051)
PST Intercept ($\sigma_{r_{0tjs}}$)	0.620***	0.693***	0.447***	0.372***	0.584***	0.328***	0.343***
	(0.014)	(0.012)	(0.010)	(0.010)	(0.020)	(0.022)	(0.021)
PST Slope ($\sigma_{r_{1tjs}}$)					0.216***	0.076***	0.084***
					(0.006)	(0.010)	(0.009)
Correlation (r_{0tjs}, r_{1tjs})					-0.753***	0.112	-0.035
					(0.017)	(0.181)	(0.139)
Residual ($\sigma_{\epsilon_{itjs}}$)	0.775***	0.477***	0.477***	0.477***	0.388***	0.388***	0.388***
	(0.007)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Supervisor Variance Prop.				0.47	0.51	0.42	0.65
School Variance Prop.			0.38	0.06	0.07	0.27	0.09
PST Variance Prop.	0.39	0.68	0.29	0.18	0.24	0.14	0.12
Residual Variance Prop.	0.61	0.32	0.33	0.29	0.18	0.17	0.15
-2LL	21946.2	15928.8	15122.3	14474.1	13795.8	13184.6	12788.2
AIC	21952.2	15938.8	15134.3	14488.1	13813.8	13206.6	12814.2
BIC	21973.2	15974.0	15176.4	14537.3	13877.0	13283.8	12905.6
N	8295	8295	8295	8295	8295	8295	8295

Notes: Standard errors shown in parentheses. Results shown are from hierarchical linear models (see equations 1–4) where observations (standardized) are nested within PSTs (pre-service teachers), nested within schools, nested within field supervisors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3

The Relationship between Observation Scores and Pre-Service Teacher and School Characteristics

	(1)	(2)	(3)	(4)
Observation Order	0.433***	0.433***	0.431***	0.431***
	(0.030)	(0.030)	(0.030)	(0.030)
Observation Order^2	-0.005	-0.005	-0.005	-0.005
	(0.004)	(0.004)	(0.004)	(0.004)
Pre-Service Teacher Characteristics				
Male	-0.257***	-0.201***	-0.195***	-0.194***
	(0.049)	(0.049)	(0.049)	(0.049)
Person of Color	-0.092***	-0.065**	-0.067**	-0.068**
	(0.029)	(0.030)	(0.030)	(0.030)
High School GPA		0.169***	0.168***	0.168***
		(0.039)	(0.039)	(0.039)
Transfer Student		-0.048*	-0.047*	-0.047*
		(0.027)	(0.027)	(0.027)
First-Generation Student		-0.001	-0.001	-0.001
		(0.026)	(0.026)	(0.026)
SAT Composite $<$ 1000		0.034	0.035	0.035
		(0.032)	(0.031)	(0.031)
SAT Composite $>$ 1200		-0.021	-0.019	-0.019
		(0.029)	(0.029)	(0.029)
Family Income $<$ \$60,000		-0.112***	-0.108***	-0.107***
		(0.032)	(0.032)	(0.032)
Family Income $>$ \$100,000		-0.062**	-0.063**	-0.063**
		(0.027)	(0.027)	(0.027)
Certification Score (Pedagogy)		0.056***	0.056***	0.056***
		(0.014)	(0.014)	(0.014)
Certification Score (Content)		-0.015	-0.015	-0.015
		(0.016)	(0.016)	(0.016)
Pre-Service School Characteristics				
Middle School			-0.031	-0.030
			(0.039)	(0.039)
High School			0.040	0.042
			(0.120)	(0.120)
Other Level School			-0.010	-0.013
			(0.101)	(0.102)
Suburban School			-0.020	-0.020
			(0.044)	(0.044)
Town/Rural School			-0.114***	-0.114***
			(0.041)	(0.041)
Proportion FRPL Students			-0.175**	-0.174**
			(0.071)	(0.071)
Supervisor Characteristics				
Male Supervisor				-0.022
				(0.127)
Gender Match with Supervisor				0.020
				(0.048)
-2LL	12500.3	12415.1	12399.6	12399.3
N	8183	8183	8183	8183

Notes: Standard errors shown in parentheses. Results are from HLM models predicting standardized observation scores that correspond to the specification shown in Table 2 Column 7, with random slopes and intercepts for PST, school, and supervisor. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4
Predicting Certification Scores

	Pedagogy Score (SD)					Content Score (SD)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Average Observation Score		0.080***	0.066***				0.059***	0.035**		
		(0.022)	(0.022)				(0.022)	(0.018)		
Supervisor BLUP				-0.030	0.017				-0.028	0.017
				(0.024)	(0.025)				(0.024)	(0.020)
School BLUP				0.019	-0.014				0.085**	0.044
				(0.039)	(0.035)				(0.037)	(0.029)
Pre-Service Teacher BLUP				0.181***	0.107***				0.095***	0.021
				(0.031)	(0.028)				(0.031)	(0.023)
Pre-Service Teacher Characteristics										
Male	-0.171**		-0.156*		-0.139*	0.311***		0.319***		0.318***
	(0.081)		(0.081)		(0.079)	(0.076)		(0.076)		(0.077)
Person of Color	-0.079		-0.072		-0.068	-0.092*		-0.088*		-0.086*
	(0.058)		(0.058)		(0.061)	(0.048)		(0.048)		(0.051)
High School GPA	0.668***		0.656***		0.638***	0.730***		0.723***		0.720***
	(0.071)		(0.071)		(0.070)	(0.059)		(0.059)		(0.062)
Transfer Student	-0.194***		-0.186***		-0.181***	-0.285***		-0.280***		-0.281***
	(0.052)		(0.052)		(0.049)	(0.045)		(0.045)		(0.048)
First-Generation Student	0.070		0.063		0.067	0.006		0.002		0.004
	(0.050)		(0.050)		(0.052)	(0.044)		(0.044)		(0.043)
SAT Composite < 1000	-0.544***		-0.544***		-0.547***	-0.789***		-0.789***		-0.789***
	(0.058)		(0.058)		(0.057)	(0.049)		(0.049)		(0.050)
SAT Composite > 1200	0.432***		0.437***		0.435***	0.526***		0.529***		0.526***
	(0.042)		(0.042)		(0.045)	(0.033)		(0.034)		(0.034)
Family Income < $60,000	-0.019		-0.007		0.003	0.019		0.025		0.027
	(0.062)		(0.062)		(0.060)	(0.053)		(0.053)		(0.051)
Family Income > $100,000	-0.070		-0.064		-0.053	0.006		0.009		0.013
	(0.044)		(0.044)		(0.046)	(0.037)		(0.037)		(0.037)
N	2051	2051	2051	2051	2051	2075	2075	2075	2075	2075
R^2	0.230	0.006	0.234	0.023	0.237	0.412	0.004	0.414	0.013	0.414

Notes: Results are from OLS models that predict standardized certification scores for pedagogy and content exams. Models also include school characteristics, which are omitted for brevity. Full results are shown in Appendix Table A4. In columns 1–3 and 6–8, robust standard errors are shown in parentheses. In columns 4–5 and 9–10, we report bootstrapped standard errors (500 repetitions) to account for the fact that the BLUPs are model-based estimates. BLUPs (best linear unbiased predictions) are the estimated random effects (intercepts) from the HLM model shown in Table 2 Column 4. Additionally, we have re-scaled the BLUPs at each level such that they are reported in supervisor-, school-, and PST-level standard deviation units, respectively.
* p < 0.1, ** p < 0.05, *** p < 0.01.

Table 5
Predicting Entry into Texas K–12 Public Schools

	(1)	(2)	(3)	(4)	(5)
Pre-Service Teacher Quality					
Average Observation Score		0.034***	0.034**		
		(0.010)	(0.016)		
Supervisor BLUP				0.035***	0.058***
				(0.011)	(0.019)
School BLUP				0.010	0.044
				(0.010)	(0.046)
Pre-Service Teacher BLUP				0.016*	0.044*
				(0.010)	(0.026)
Certification Score (Pedagogy)		0.018*	0.028*	0.018*	0.018*
		(0.010)	(0.016)	(0.010)	(0.010)
Certification Score (Content)		0.002	-0.001	0.002	0.002
		(0.013)	(0.020)	(0.013)	(0.013)
Pre-Service Teacher Characteristics					
Male	-0.005	0.005	0.044	0.001	0.001
	(0.038)	(0.039)	(0.058)	(0.042)	(0.042)
Person of Color	-0.027	-0.022	-0.010	-0.022	-0.022
	(0.022)	(0.022)	(0.030)	(0.023)	(0.023)
High School GPA	-0.033	-0.054	-0.089*	-0.050	-0.050
	(0.033)	(0.036)	(0.053)	(0.036)	(0.036)
Transfer Student	-0.014	-0.006	-0.032	-0.007	-0.007
	(0.024)	(0.025)	(0.033)	(0.025)	(0.025)
First-Generation Student	0.043**	0.039*	0.024	0.039*	0.039*
	(0.021)	(0.021)	(0.031)	(0.023)	(0.023)
SAT Composite < 1000	0.044*	0.054**	0.041	0.054**	0.054**
	(0.026)	(0.028)	(0.034)	(0.026)	(0.026)
SAT Composite > 1200	-0.048*	-0.055**	-0.043	-0.054**	-0.054**
	(0.025)	(0.026)	(0.035)	(0.026)	(0.026)
Family Income < $60,000	-0.054**	-0.048*	-0.033	-0.049*	-0.049*
	(0.026)	(0.026)	(0.033)	(0.027)	(0.027)
Family Income > $100,000	0.006	0.011	0.005	0.012	0.012
	(0.016)	(0.016)	(0.020)	(0.016)	(0.016)
Pre-Service School Characteristics					
Middle School	-0.006	0.001		0.003	0.003
	(0.025)	(0.024)		(0.025)	(0.025)
High School	-0.084	-0.071		-0.065	-0.065
	(0.050)	(0.052)		(0.067)	(0.067)
Other Level School	-0.014	0.000		0.005	0.005
	(0.060)	(0.060)		(0.072)	(0.072)
Suburban School	0.034	0.024		0.018	0.018
	(0.025)	(0.022)		(0.024)	(0.024)
Town/Rural School	-0.007	-0.003		-0.004	-0.004
	(0.025)	(0.023)		(0.028)	(0.028)
Proportion FRPL Students	-0.040	-0.012		-0.006	-0.006
	(0.025)	(0.027)		(0.029)	(0.029)
Program-by-Cohort FE	✓	✓	✓	✓	✓
School FE			✓		
N	2050	2050	2050	2050	2050
R^2	0.031	0.039	0.286	0.040	0.040

Notes: Results are from linear probability models that predict entry into Texas K–12 public schools. In columns 1–3, we report standard errors clustered by school district in parentheses. In columns 4 and 5, we report bootstrapped standard errors (500 repetitions, clustered by school district) to account for the fact that the BLUPs are model-based estimates. BLUPs (best linear unbiased predictions) are the estimated random effects (intercepts) from the HLM model shown in Table 2 Column 4. In column 4, we have re-scaled the BLUPs at each level such that they are reported in supervisor-, school-, and PST-level standard deviation units, respectively. Column 5 shows the same model as Column 4, except the BLUPs are not re-scaled by level and instead remain in observation score units.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6

Predicting Alternative Entry Outcomes

	Immediate Entry			Entry into Same School		
	(1)	(2)	(3)	(4)	(5)	(6)
Pre-Service Teacher Quality						
Average Observation Score	0.046***	0.051**		0.059***	0.062***	
	(0.011)	(0.022)		(0.010)	(0.021)	
Supervisor BLUP			0.046***			0.049***
			(0.014)			(0.016)
School BLUP			-0.003			-0.007
			(0.013)			(0.023)
Pre-Service Teacher BLUP			0.034***			0.053***
			(0.013)			(0.016)
Certification Score (Pedagogy)	0.028***	0.042***	0.028***	0.010	0.019	0.009
	(0.010)	(0.015)	(0.010)	(0.014)	(0.018)	(0.014)
Certification Score (Content)	0.012	0.006	0.013	-0.007	-0.012	-0.006
	(0.013)	(0.021)	(0.014)	(0.014)	(0.021)	(0.014)
Pre-Service Teacher Characteristics						
Male	0.005	0.056	0.003	0.015	-0.014	0.016
	(0.043)	(0.062)	(0.046)	(0.039)	(0.044)	(0.042)
Person of Color	-0.040*	-0.032	-0.041	-0.026	0.005	-0.028
	(0.023)	(0.034)	(0.025)	(0.024)	(0.038)	(0.024)
High School GPA	-0.118***	-0.144**	-0.115***	-0.018	-0.007	-0.017
	(0.043)	(0.062)	(0.041)	(0.033)	(0.048)	(0.034)
Transfer Student	-0.007	-0.015	-0.008	-0.005	0.024	-0.005
	(0.031)	(0.042)	(0.032)	(0.023)	(0.030)	(0.025)
First-Generation Student	0.047*	0.032	0.047*	0.044	0.040	0.044*
	(0.026)	(0.038)	(0.028)	(0.027)	(0.033)	(0.027)
SAT Composite < 1000	0.081**	0.066	0.081**	0.008	-0.040	0.008
	(0.031)	(0.040)	(0.032)	(0.027)	(0.033)	(0.028)
SAT Composite > 1200	-0.060***	-0.029	-0.059***	0.003	-0.021	0.005
	(0.022)	(0.028)	(0.023)	(0.025)	(0.025)	(0.026)
Family Income < $60,000	-0.053*	-0.054	-0.054*	0.020	0.017	0.020
	(0.030)	(0.042)	(0.031)	(0.031)	(0.038)	(0.031)
Family Income > $100,000	0.003	-0.012	0.005	0.019	0.036	0.020
	(0.021)	(0.031)	(0.022)	(0.029)	(0.034)	(0.029)
Pre-Service School Characteristics						
Middle School	-0.014		-0.012	0.033		0.035
	(0.032)		(0.032)	(0.034)		(0.035)
High School	-0.104*		-0.097	-0.180***		-0.175***
	(0.057)		(0.074)	(0.028)		(0.036)
Other Level School	-0.075		-0.070	-0.058		-0.056
	(0.091)		(0.101)	(0.044)		(0.055)
Suburban School	0.047		0.038	0.078**		0.070*
	(0.031)		(0.034)	(0.036)		(0.038)
Town/Rural School	0.009		0.003	0.014		0.007
	(0.033)		(0.039)	(0.031)		(0.038)
Proportion FRPL Students	-0.039		-0.035	-0.035		-0.038
	(0.032)		(0.037)	(0.043)		(0.048)
Program-by-Cohort FE	✓	✓	✓	✓	✓	✓
School FE		✓			✓	
N	2050	2050	2050	1702	1702	1702
R^2	0.047	0.296	0.048	0.063	0.549	0.065

Notes: Results are from linear probability models that predict immediate entry (i.e., in the year immediately following graduation) into Texas K–12 public schools (column 1–3) and entry into the same school as clinical teaching (columns 4–6), conditional on ever entering. In columns 1–2 and 4–5, we report standard errors clustered by school district in parentheses. In columns 3 and 6, we report bootstrapped standard errors (500 repetitions, clustered by school district) to account for the fact that the BLUPs are model-based estimates. BLUPs (best linear unbiased predictions) are the estimated random effects (intercepts) from the HLM model shown in Table 2 Column 4. We have re-scaled the BLUPs at each level such that they are reported in supervisor-, school-, and PST-level standard deviation units, respectively.
* p < 0.1, ** p < 0.05, *** p < 0.01.

Table 7
Predicting Advantage Index of Full-time School Placement

	(1)	(2)	(3)	(4)	(5)
Pre-Service Teacher Quality					
Average Observation Score			0.011	0.047	
			(0.028)	(0.041)	
Supervisor BLUP					-0.054
					(0.035)
School BLUP					0.127**
					(0.050)
Pre-Service Teacher BLUP					0.007
					(0.029)
Certification Score (Pedagogy)			0.048*	0.055	0.047*
			(0.027)	(0.042)	(0.028)
Certification Score (Content)			-0.058*	-0.102**	-0.062*
			(0.034)	(0.050)	(0.033)
Pre-Service Teacher Characteristics					
Male	-0.380***	-0.360***	-0.325***	-0.223	-0.310***
	(0.093)	(0.093)	(0.089)	(0.139)	(0.091)
Person of Color	-0.327***	-0.311***	-0.310***	-0.172**	-0.302***
	(0.069)	(0.060)	(0.059)	(0.082)	(0.058)
High School GPA	0.219***	0.193**	0.199**	0.280**	0.186**
	(0.081)	(0.081)	(0.083)	(0.114)	(0.081)
Transfer Student	0.135**	0.026	0.023	-0.061	0.025
	(0.064)	(0.060)	(0.059)	(0.091)	(0.064)
First-Generation Student	-0.082	-0.068	-0.074	-0.085	-0.068
	(0.055)	(0.054)	(0.054)	(0.073)	(0.054)
SAT Composite < 1000	-0.012	0.024	0.006	0.028	0.004
	(0.053)	(0.055)	(0.069)	(0.098)	(0.066)
SAT Composite > 1200	0.033	0.038	0.048	0.027	0.039
	(0.066)	(0.063)	(0.066)	(0.096)	(0.066)
Family Income < $60,000	-0.216**	-0.150	-0.149	-0.133	-0.144
	(0.094)	(0.099)	(0.098)	(0.147)	(0.099)
Family Income > $100,000	0.124**	0.072	0.075	0.117*	0.076
	(0.051)	(0.049)	(0.048)	(0.060)	(0.052)
Pre-Service School Characteristics					
Middle School		-0.110*	-0.105		-0.116*
		(0.066)	(0.066)		(0.069)
High School		0.022	0.037		-0.012
		(0.252)	(0.254)		(0.243)
Other Level School		0.124	0.119		0.075
		(0.281)	(0.275)		(0.264)
Suburban School		0.187**	0.186**		0.238***
		(0.076)	(0.078)		(0.083)
Town/Rural School		0.054	0.054		0.083
		(0.088)	(0.088)		(0.089)
Proportion FRPL Students		-1.176***	-1.165***		-1.197***
		(0.164)	(0.168)		(0.164)
Program-by-Cohort FE	✓	✓	✓	✓	✓
School FE				✓	
N	1702	1702	1702	1702	1702
R^2	0.085	0.203	0.205	0.488	0.212

Notes: Results are from OLS models that predict advantage index of a PST's full-time teaching school, conditional on entry into Texas K–12 public schools. The advantage index is constructed via principal components analysis based on demographic characteristics. The index is scaled to have a mean of zero and standard deviation of one. Appendix Table A8 shows equivalent models using each individual school characteristic and Appendix Table A9 shows a correlation matrix for the advantage index and school characteristics. In columns 1–4, we report standard errors clustered by school district in parentheses. In column 5, we report bootstrapped standard errors (500 repetitions, clustered by school district) to account for the fact that the BLUPs are model-based estimates. BLUPs (best linear unbiased predictions) are the estimated random effects (intercepts) from the HLM model shown in Table 2 Column 4. We have re-scaled the BLUPs at each level such that they are reported in supervisor-, school-, and PST-level standard deviation units, respectively.
* p < 0.1, ** p < 0.05, *** p < 0.01.

Appendix

Table A1
Sorting Test for Supervisor and School Variance Components

	Supervisor		School		PST	
	BLUP (1)	FE (2)	BLUP (3)	FE (4)	BLUP (5)	FE (6)
Male	-0.038	-0.070	-0.000	0.055	-0.298***	-0.487***
	(0.073)	(0.086)	(0.022)	(0.064)	(0.080)	(0.112)
Person of Color	0.061	0.097	-0.033	-0.152**	-0.094*	-0.098
	(0.076)	(0.098)	(0.020)	(0.060)	(0.049)	(0.070)
High School GPA	0.003	0.001	0.036*	0.096	0.266***	0.300***
	(0.085)	(0.105)	(0.021)	(0.062)	(0.060)	(0.085)
Transfer Student	-0.137	-0.151	-0.009	-0.053	-0.092**	-0.099
	(0.102)	(0.127)	(0.023)	(0.069)	(0.044)	(0.064)
First-Generation Student	0.126	0.139	-0.009	-0.061	0.000	0.022
	(0.083)	(0.101)	(0.022)	(0.058)	(0.041)	(0.058)
SAT Composite Missing	0.090	0.090	0.006	0.030	-0.070	-0.127
	(0.105)	(0.131)	(0.026)	(0.076)	(0.063)	(0.091)
SAT Composite < 1000	0.045	0.067	0.022	0.074	0.063	0.096
	(0.096)	(0.122)	(0.023)	(0.067)	(0.054)	(0.075)
SAT Composite > 1200	-0.123	-0.128	-0.011	-0.012	-0.034	-0.004
	(0.091)	(0.107)	(0.021)	(0.058)	(0.046)	(0.067)
Family Income Missing	-0.108	-0.134	-0.046*	-0.139*	-0.207***	-0.249***
	(0.087)	(0.102)	(0.024)	(0.075)	(0.067)	(0.095)
Family Income < $60,000	-0.144	-0.159	-0.047**	-0.124*	-0.200***	-0.253***
	(0.097)	(0.124)	(0.023)	(0.069)	(0.049)	(0.074)
Family Income > $100,000	0.014	0.031	-0.019	-0.066	-0.135***	-0.220***
	(0.096)	(0.114)	(0.024)	(0.064)	(0.040)	(0.058)
Certification Score (Pedagogy)	0.122	0.192	0.022	0.073	0.081***	0.102***
	(0.125)	(0.173)	(0.027)	(0.070)	(0.022)	(0.033)
Certification Score (Content)	-0.018	-0.032	0.042	0.075	-0.009	-0.027
	(0.106)	(0.141)	(0.031)	(0.091)	(0.028)	(0.037)
Joint Wald chi2 (p)	0.26	0.17	0.01	0.01	0.00	0.00
N	148	148	606	560	2062	1780

Notes: Each model regresses supervisor, school, or PST effects (BLUPs or FEs) on PST characteristics. The BLUPs and FEs are scaled in supervisor-, school-, or PST-level standard deviations. For supervisor and school models, we compute the mean of each characteristic within supervisor/school, then re-scale by the standard deviation of the given mean characteristic, which makes the coefficients equivalent to standardized regression coefficients. We report bootstrapped standard errors (500 repetitions) to account for the fact that the BLUPs and fixed effects are model-based estimates. We report a joint test of significance for all of the PST characteristics at the bottom of each column.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2

Hausman Tests for Observation Scores

	Supervisor		School		Sup x School	
	FE	RE	FE	RE	FE	RE
	(1)	(2)	(3)	(4)	(5)	(6)
Observation Order	0.493***	0.493***	0.492***	0.492***	0.493***	0.493***
	(0.035)	(0.035)	(0.034)	(0.034)	(0.031)	(0.031)
Observation Order^2	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
Male	-0.193***	-0.195***	-0.240***	-0.225***	-0.296***	-0.266***
	(0.034)	(0.034)	(0.038)	(0.037)	(0.038)	(0.036)
Person of Color	-0.089***	-0.086***	-0.091***	-0.089***	-0.046**	-0.050**
	(0.021)	(0.021)	(0.024)	(0.023)	(0.023)	(0.022)
High School GPA	0.205***	0.205***	0.193***	0.186***	0.224***	0.211***
	(0.025)	(0.025)	(0.029)	(0.028)	(0.029)	(0.028)
Transfer Student	-0.046**	-0.048**	-0.071***	-0.074***	-0.071***	-0.072***
	(0.019)	(0.019)	(0.021)	(0.021)	(0.021)	(0.020)
First-Generation Student	-0.010	-0.007	-0.000	0.005	-0.000	0.005
	(0.018)	(0.018)	(0.021)	(0.020)	(0.021)	(0.020)
SAT Composite < 1000	0.012	0.012	0.034	0.026	0.040*	0.028
	(0.021)	(0.021)	(0.024)	(0.023)	(0.023)	(0.023)
SAT Composite > 1200	0.011	0.008	-0.000	-0.005	-0.028	-0.030
	(0.019)	(0.019)	(0.022)	(0.021)	(0.021)	(0.021)
Family Income < \$60,000	-0.121***	-0.122***	-0.132***	-0.139***	-0.106***	-0.115***
	(0.023)	(0.023)	(0.026)	(0.025)	(0.025)	(0.024)
Family Income > \$100,000	-0.092***	-0.091***	-0.155***	-0.147***	-0.093***	-0.090***
	(0.018)	(0.018)	(0.020)	(0.019)	(0.020)	(0.019)
Middle School	-0.021	-0.019				
	(0.018)	(0.018)				
High School	-0.033	-0.037				
	(0.032)	(0.031)				
Other Level School	0.072	0.066				
	(0.055)	(0.055)				
Suburban School	-0.058**	-0.061**				
	(0.026)	(0.025)				
Town/Rural School	-0.119***	-0.116***				
	(0.020)	(0.020)				
Proportion FRPL Students	-0.142***	-0.142***				
	(0.039)	(0.038)				
Hausman Test (p)		0.59		0.51		0.15
N	8295	8295	8295	8295	8295	8295

Notes: Results show coefficient estimates and an omnibus Hausman test for fixed versus random effects from OLS models predicting standardized observation scores. The panel variable (supervisor, school, or supervisor-by-school) is listed in the column header. p-values for the Hausman test are shown at the bottom of the random effects columns. FE = fixed effects, RE = random effects.

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

Table A3

Robustness of Observation Score Results to Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
Observation Order	0.493***	0.494***	0.493***			
	(0.024)	(0.025)	(0.025)			
Observation Order^2	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
PST Characteristics						
Male	-0.192***	-0.296***	-0.296***	-0.193***	-0.296***	-0.296***
	(0.055)	(0.053)	(0.055)	(0.055)	(0.055)	(0.056)
Person of Color	-0.090***	-0.064*	-0.046	-0.090***	-0.065*	-0.065*
	(0.033)	(0.034)	(0.034)	(0.033)	(0.036)	(0.036)
High School GPA	0.205***	0.209***	0.223***	0.205***	0.209***	0.209***
	(0.037)	(0.039)	(0.039)	(0.038)	(0.041)	(0.041)
Transfer Student	-0.045	-0.072**	-0.071**	-0.045	-0.072**	-0.072**
	(0.028)	(0.030)	(0.030)	(0.028)	(0.031)	(0.031)
First-Generation Student	-0.011	0.015	-0.001	-0.011	0.015	0.015
	(0.027)	(0.028)	(0.028)	(0.027)	(0.029)	(0.029)
SAT Composite < 1000	0.011	0.035	0.041	0.011	0.034	0.034
	(0.032)	(0.033)	(0.034)	(0.032)	(0.035)	(0.035)
SAT Composite > 1200	0.012	-0.007	-0.028	0.012	-0.008	-0.008
	(0.028)	(0.030)	(0.030)	(0.028)	(0.031)	(0.032)
Family Income < $60,000	-0.123***	-0.122***	-0.105***	-0.123***	-0.122***	-0.121***
	(0.033)	(0.034)	(0.035)	(0.034)	(0.036)	(0.036)
Family Income > $100,000	-0.094***	-0.108***	-0.092***	-0.093***	-0.108***	-0.108***
	(0.026)	(0.027)	(0.029)	(0.026)	(0.029)	(0.029)
Placement School Characteristics						
Middle School	-0.027			-0.027		
	(0.026)			(0.026)		
High School	0.079			0.079		
	(0.080)			(0.081)		
Other Level School	0.073			0.073		
	(0.087)			(0.088)		
Suburban School	-0.058			-0.059		
	(0.043)			(0.043)		
Town/Rural School	-0.120***			-0.119***		
	(0.029)			(0.030)		
Proportion FRPL Students	-0.141**			-0.141**		
	(0.056)			(0.056)		
Supervisor Characteristics						
Gender Match with Supervisor	-0.005	0.050	0.036	-0.005	0.050	0.050
	(0.052)	(0.051)	(0.054)	(0.053)	(0.053)	(0.054)
Supervisor FE	✓	✓		✓	✓	✓
School FE		✓			✓	✓
Supervisor-School FE			✓			
Supervisor-Specific Slope				✓		✓
School-Specific Slope					✓	✓
N	8295	8295	8295	8295	8295	8295
R^2	0.619	0.700	0.717	0.665	0.748	0.764

Notes: Results are from OLS models predicting standardized observation scores. Two-way clustered standard errors (supervisor-by-PST) shown in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4

Predicting Certification Scores (Full results, including school characteristics)

	Pedagogy Score (SD)					Content Score (SD)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Average Observation Score		0.080***	0.066***				0.059***	0.035**		
		(0.022)	(0.022)				(0.022)	(0.018)		
Supervisor BLUP				-0.030	0.017				-0.028	0.017
				(0.024)	(0.025)				(0.024)	(0.020)
School BLUP				0.019	-0.014				0.085**	0.044
				(0.039)	(0.035)				(0.037)	(0.029)
Pre-Service Teacher BLUP				0.181***	0.107***				0.095***	0.021
				(0.031)	(0.028)				(0.031)	(0.023)
Pre-Service Teacher Characteristics										
Male	-0.171**		-0.156*		-0.139*	0.311***		0.319***		0.318***
	(0.081)		(0.081)		(0.079)	(0.076)		(0.076)		(0.077)
Person of Color	-0.079		-0.072		-0.068	-0.092*		-0.088*		-0.086*
	(0.058)		(0.058)		(0.061)	(0.048)		(0.048)		(0.051)
High School GPA	0.668***		0.656***		0.638***	0.730***		0.723***		0.720***
	(0.071)		(0.071)		(0.070)	(0.059)		(0.059)		(0.062)
Transfer Student	-0.194***		-0.186***		-0.181***	-0.285***		-0.280***		-0.281***
	(0.052)		(0.052)		(0.049)	(0.045)		(0.045)		(0.048)
First-Generation Student	0.070		0.063		0.067	0.006		0.002		0.004
	(0.050)		(0.050)		(0.052)	(0.044)		(0.044)		(0.043)
SAT Composite < 1000	-0.544***		-0.544***		-0.547***	-0.789***		-0.789***		-0.789***
	(0.058)		(0.058)		(0.057)	(0.049)		(0.049)		(0.050)
SAT Composite > 1200	0.432***		0.437***		0.435***	0.526***		0.529***		0.526***
	(0.042)		(0.042)		(0.045)	(0.033)		(0.034)		(0.034)
Family Income < $60,000	-0.019		-0.007		0.003	0.019		0.025		0.027
	(0.062)		(0.062)		(0.060)	(0.053)		(0.053)		(0.051)
Family Income > $100,000	-0.070		-0.064		-0.053	0.006		0.009		0.013
	(0.044)		(0.044)		(0.046)	(0.037)		(0.037)		(0.037)
Pre-Service School Characteristics										
Middle School	-0.011		-0.004		-0.011	0.017		0.021		0.021
	(0.041)		(0.041)		(0.041)	(0.034)		(0.034)		(0.035)
High School	-0.072		-0.052		-0.068	0.138		0.149		0.142
	(0.199)		(0.201)		(0.209)	(0.161)		(0.161)		(0.163)
Other Level School	-0.048		-0.029		-0.053	-0.217*		-0.205*		-0.212**
	(0.102)		(0.100)		(0.099)	(0.111)		(0.111)		(0.105)
Suburban School	0.007		-0.012		-0.004	0.067		0.057		0.068
	(0.049)		(0.049)		(0.050)	(0.041)		(0.041)		(0.043)
Town/Rural School	0.027		0.035		0.028	0.023		0.027		0.034
	(0.047)		(0.047)		(0.049)	(0.039)		(0.039)		(0.040)
Proportion FRPL Students	0.055		0.112		0.076	0.059		0.090		0.085
	(0.076)		(0.079)		(0.075)	(0.064)		(0.067)		(0.067)
N	2051	2051	2051	2051	2051	2075	2075	2075	2075	2075
R^2	0.230	0.006	0.234	0.023	0.237	0.412	0.004	0.414	0.013	0.414

Notes: Results are from OLS models that predict standardized certification scores for pedagogy and content exams. Models also include school characteristics, which are omitted for brevity. In columns 1–3 and 6–8, robust standard errors are shown in parentheses. In columns 4–5 and 9–10, we report bootstrapped standard errors (500 repetitions) to account for the fact that the BLUPs are model-based estimates. BLUPs (best linear unbiased predictions) are the estimated random effects (intercepts) from the HLM model shown in Table 2 Column 4. Additionally, we have re-scaled the BLUPs at each level such that they are reported in supervisor-, school-, and PST-level standard deviation units, respectively.

* p < 0.1, ** p < 0.05, *** p < 0.01.

Table A5

Predicting Certification Scores Using Fixed Effects Instead of BLUPs

	Pedagogy Score (SD)					Content Score (SD)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Average Observation Score		0.068***	0.062**				0.056**	0.037*		
		(0.025)	(0.024)				(0.025)	(0.022)		
Supervisor FE				-0.039	0.015				-0.036	0.021
				(0.031)	(0.034)				(0.029)	(0.027)
School FE				0.046	0.009				0.092***	0.043
				(0.033)	(0.029)				(0.034)	(0.028)
Pre-Service Teacher FE				0.133***	0.080***				0.072***	0.014
				(0.025)	(0.022)				(0.022)	(0.019)
Pre-Service Teacher Characteristics										
Male	-0.155		-0.130		-0.106	0.295***		0.309***		0.305***
	(0.098)		(0.097)		(0.097)	(0.088)		(0.087)		(0.086)
Person of Color	-0.053		-0.041		-0.044	-0.111*		-0.104*		-0.105*
	(0.068)		(0.068)		(0.067)	(0.058)		(0.058)		(0.057)
High School GPA	0.650***		0.636***		0.615***	0.777***		0.768***		0.768***
	(0.082)		(0.082)		(0.081)	(0.070)		(0.070)		(0.071)
Transfer Student	-0.231***		-0.225***		-0.219***	-0.274***		-0.271***		-0.273***
	(0.060)		(0.060)		(0.063)	(0.050)		(0.050)		(0.050)
First-Generation Student	0.090		0.088		0.090	-0.008		-0.010		-0.007
	(0.059)		(0.059)		(0.056)	(0.051)		(0.051)		(0.052)
SAT Composite < 1000	-0.504***		-0.507***		-0.509***	-0.770***		-0.772***		-0.770***
	(0.067)		(0.067)		(0.068)	(0.057)		(0.057)		(0.062)
SAT Composite > 1200	0.433***		0.434***		0.437***	0.512***		0.514***		0.512***
	(0.048)		(0.048)		(0.044)	(0.038)		(0.038)		(0.037)
Family Income < $60,000	0.033		0.044		0.054	0.003		0.009		0.010
	(0.069)		(0.069)		(0.069)	(0.062)		(0.062)		(0.061)
Family Income > $100,000	-0.040		-0.032		-0.024	0.007		0.011		0.015
	(0.049)		(0.050)		(0.048)	(0.041)		(0.041)		(0.040)
N	1613	1613	1613	1613	1613	1633	1633	1633	1633	1633
R^2	0.226	0.005	0.229	0.022	0.233	0.412	0.003	0.414	0.012	0.414

Notes: Results are from OLS models that predict standardized certification scores for pedagogy and content exams. In columns 1–3 and 6–8, robust standard errors are shown in parentheses. In columns 4–5 and 9–10, we report bootstrapped standard errors (500 repetitions) to account for the fact that the fixed effects are model-based estimates. Additionally, we have re-scaled the fixed effects at each level such that they are reported in supervisor-, school-, and PST-level standard deviation units, respectively.

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

Table A6

Predicting Entry (Restricted Samples)

	Drop Same School				Drop Same District			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pre-Service Teacher Quality								
Average Observation Score		0.029**	0.032*			0.016	0.018	
		(0.011)	(0.018)			(0.014)	(0.021)	
Supervisor BLUP				0.032**				0.019
				(0.013)				(0.016)
School BLUP				0.012				0.018
				(0.011)				(0.015)
Pre-Service Teacher BLUP				0.011				0.001
				(0.012)				(0.014)
Certification Score (Pedagogy)		0.017	0.025	0.018		0.019	0.022	0.019
		(0.012)	(0.018)	(0.011)		(0.014)	(0.020)	(0.014)
Certification Score (Content)		0.004	0.002	0.004		0.006	0.004	0.006
		(0.015)	(0.023)	(0.015)		(0.018)	(0.029)	(0.020)
Pre-Service Teacher Characteristics								
Male	-0.003	0.005	0.050	0.002	0.017	0.022	0.077	0.020
	(0.043)	(0.044)	(0.065)	(0.043)	(0.045)	(0.046)	(0.066)	(0.048)
Person of Color	-0.025	-0.019	-0.012	-0.020	-0.050*	-0.045	-0.047	-0.045
	(0.026)	(0.026)	(0.035)	(0.030)	(0.029)	(0.029)	(0.040)	(0.032)
High School GPA	-0.032	-0.053	-0.090	-0.048	-0.050	-0.070	-0.104	-0.067
	(0.039)	(0.041)	(0.061)	(0.043)	(0.042)	(0.045)	(0.068)	(0.044)
Transfer Student	-0.011	-0.003	-0.025	-0.005	-0.023	-0.016	-0.032	-0.017
	(0.027)	(0.028)	(0.034)	(0.026)	(0.030)	(0.031)	(0.038)	(0.031)
First-Generation Student	0.044*	0.040	0.026	0.039	0.047	0.044	0.035	0.043
	(0.026)	(0.026)	(0.035)	(0.028)	(0.030)	(0.030)	(0.040)	(0.031)
SAT Composite < 1000	0.047*	0.059*	0.037	0.059*	0.052	0.068*	0.040	0.068*
	(0.028)	(0.031)	(0.037)	(0.032)	(0.034)	(0.037)	(0.045)	(0.039)
SAT Composite > 1200	-0.053*	-0.061**	-0.046	-0.061**	-0.067**	-0.079**	-0.066	-0.079**
	(0.027)	(0.029)	(0.038)	(0.028)	(0.031)	(0.035)	(0.046)	(0.036)
Family Income < $60,000	-0.062**	-0.057**	-0.045	-0.059**	-0.080**	-0.077**	-0.062	-0.077**
	(0.029)	(0.029)	(0.036)	(0.029)	(0.032)	(0.033)	(0.041)	(0.034)
Family Income > $100,000	0.007	0.013	0.003	0.013	-0.001	0.004	0.001	0.004
	(0.018)	(0.018)	(0.021)	(0.019)	(0.021)	(0.021)	(0.026)	(0.022)
Pre-Service School Characteristics								
Middle School	-0.012	-0.007		-0.005	0.003	0.007		0.008
	(0.026)	(0.026)		(0.026)	(0.030)	(0.031)		(0.034)
High School	-0.056	-0.047		-0.042	-0.241*	-0.233*		-0.230
	(0.053)	(0.055)		(0.071)	(0.132)	(0.138)		(0.164)
Other Level School	-0.002	0.011		0.016	0.017	0.026		0.029
	(0.064)	(0.064)		(0.085)	(0.066)	(0.066)		(0.087)
Suburban School	0.021	0.011		0.006	-0.003	-0.009		-0.011
	(0.024)	(0.022)		(0.024)	(0.029)	(0.029)		(0.035)
Town/Rural School	-0.010	-0.007		-0.008	0.024	0.025		0.028
	(0.024)	(0.023)		(0.027)	(0.023)	(0.023)		(0.034)
Proportion FRPL Students	-0.034	-0.011		-0.003	-0.017	-0.003		0.003
	(0.027)	(0.031)		(0.037)	(0.037)	(0.040)		(0.044)
Program-by-Cohort FE	✓	✓	✓	✓	✓	✓	✓	✓
School FE			✓				✓	
N	1756	1756	1756	1756	1411	1411	1411	1411
R^2	0.031	0.037	0.312	0.038	0.037	0.041	0.346	0.041

Notes: Results are from linear probability models that predict entry into Texas K–12 public schools. In columns 1–4 (5–8), we drop PSTs who entered the same school (district) as clinical teaching. In columns 1–3 and 5–7, we report standard errors clustered by school district in parentheses. In columns 4 and 8, we report bootstrapped standard errors (500 repetitions, clustered by school district) to account for the fact that the BLUPs are model-based estimates. BLUPs (best linear unbiased predictions) are the estimated random effects (intercepts) from the HLM model shown in Table 2 Column 4. In column 4, we have re-scaled the BLUPs at each level such that they are reported in supervisor-, school-, and PST-level standard deviation units, respectively. Column 5 shows the same model as Column 4, except the BLUPs are not re-scaled by level and instead remain in observation score units.

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

Table A7

Predicting Entry into Same District as Clinical Teaching

				Drop Same School		
	(1)	(2)	(3)	(4)	(5)	(6)
Pre-Service Teacher Quality						
Average Observation Score	0.096***	0.090***		0.062***	0.065***	
	(0.019)	(0.021)		(0.018)	(0.019)	
Supervisor BLUP			0.089**			0.061*
			(0.037)			(0.034)
School BLUP			-0.024			-0.029
			(0.032)			(0.024)
Pre-Service Teacher BLUP			0.077***			0.049***
			(0.015)			(0.015)
Certification Score (Pedagogy)	0.010	0.016	0.010	0.005	-0.002	0.004
	(0.013)	(0.016)	(0.013)	(0.013)	(0.015)	(0.014)
Certification Score (Content)	-0.014	0.007	-0.012	-0.010	0.003	-0.009
	(0.020)	(0.026)	(0.021)	(0.022)	(0.026)	(0.023)
Pre-Service Teacher Characteristics						
Male	-0.041	-0.110**	-0.045	-0.059	-0.117**	-0.066
	(0.040)	(0.048)	(0.040)	(0.044)	(0.051)	(0.046)
Person of Color	0.059*	0.037	0.054	0.093**	0.056**	0.089**
	(0.035)	(0.036)	(0.036)	(0.037)	(0.028)	(0.038)
High School GPA	0.016	-0.015	0.021	0.034	-0.026	0.040
	(0.045)	(0.049)	(0.045)	(0.047)	(0.038)	(0.047)
Transfer Student	0.046*	0.057	0.044*	0.052*	0.038	0.051*
	(0.025)	(0.037)	(0.026)	(0.027)	(0.030)	(0.027)
First-Generation Student	0.040	0.038	0.040	0.013	0.007	0.012
	(0.026)	(0.032)	(0.027)	(0.023)	(0.030)	(0.024)
SAT Composite $<$ 1000	0.014	0.002	0.015	0.016	0.021	0.018
	(0.032)	(0.039)	(0.032)	(0.029)	(0.039)	(0.030)
SAT Composite $>$ 1200	0.035	0.007	0.038	0.036	0.040	0.039
	(0.027)	(0.033)	(0.026)	(0.026)	(0.034)	(0.027)
Family Income $<$ \$60,000	0.052	0.066*	0.049	0.036	0.060	0.032
	(0.032)	(0.038)	(0.033)	(0.034)	(0.046)	(0.036)
Family Income $>$ \$100,000	0.044	0.036	0.045	0.045	0.026	0.045
	(0.032)	(0.036)	(0.030)	(0.031)	(0.035)	(0.032)
Pre-Service School Characteristics						
Middle School	-0.018		-0.015	-0.055		-0.052
	(0.055)		(0.048)	(0.050)		(0.047)
High School	0.208*		0.223*	0.324***		0.338***
	(0.110)		(0.118)	(0.104)		(0.114)
Other Level School	-0.142**		-0.133*	-0.087*		-0.076
	(0.062)		(0.079)	(0.045)		(0.060)
Suburban School	0.124		0.104	0.090		0.071
	(0.096)		(0.091)	(0.103)		(0.106)
Town/Rural School	-0.120		-0.134	-0.154**		-0.169**
	(0.075)		(0.084)	(0.068)		(0.082)
Proportion FRPL Students	-0.083		-0.084	-0.071		-0.069
	(0.080)		(0.082)	(0.075)		(0.077)
Program-by-Cohort FE	✓	✓	✓	✓	✓	✓
School FE		✓			✓	
N	1702	1702	1702	1408	1408	1408
R^2	0.139	0.575	0.141	0.136	0.623	0.138

Notes: Results are from linear probability models that predict entry into the same district as clinical teaching, conditional on ever entering. Columns 4–6 also drop PSTs who entered the same school as clinical teaching. In columns 1–2 and 4–5, we report standard errors clustered by school district in parentheses. In columns 3 and 6, we report bootstrapped standard errors (500 repetitions, clustered by school district) to account for the fact that the BLUPs are model-based estimates. BLUPs (best linear unbiased predictions) are the estimated random effects (intercepts) from the HLM model shown in Table 2 Column 4. We have re-scaled the BLUPs at each level such that they are reported in supervisor-, school-, and PST-level standard deviation units, respectively.

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

Table A8

Predicting Characteristics of Full-time School Placement

	FRPL (1)	Black (2)	Hisp (3)	White (4)	Urban (5)	Suburb (6)	Town/Rural (7)
Pre-Service Teacher Quality							
Average Observation Score	-0.009	-0.003	-0.002	0.005	0.020	-0.022*	0.003
	(0.008)	(0.003)	(0.008)	(0.008)	(0.014)	(0.012)	(0.016)
Certification Score (Pedagogy)	-0.013*	-0.000	-0.007	0.007	-0.016	0.019	-0.003
	(0.008)	(0.003)	(0.007)	(0.006)	(0.013)	(0.019)	(0.018)
Certification Score (Content)	0.017*	0.001	0.006	-0.003	0.015	-0.031	0.016
	(0.009)	(0.004)	(0.008)	(0.008)	(0.016)	(0.020)	(0.014)
Pre-Service Teacher Characteristics							
Male	0.065***	-0.003	0.082***	-0.061***	0.049	-0.003	-0.046
	(0.017)	(0.010)	(0.024)	(0.020)	(0.065)	(0.033)	(0.058)
Person of Color	0.059***	0.005	0.085***	-0.085***	0.154***	-0.044*	-0.110***
	(0.017)	(0.008)	(0.017)	(0.016)	(0.026)	(0.027)	(0.022)
High School GPA	-0.063***	-0.011	-0.043**	0.038*	0.027	0.035	-0.062*
	(0.023)	(0.010)	(0.021)	(0.020)	(0.033)	(0.038)	(0.035)
Transfer Student	-0.011	-0.006	-0.003	0.005	0.032	-0.010	-0.022
	(0.017)	(0.007)	(0.016)	(0.013)	(0.032)	(0.030)	(0.031)
First-Generation Student	0.032**	0.002	0.018	-0.004	-0.048	-0.004	0.052*
	(0.014)	(0.006)	(0.013)	(0.011)	(0.033)	(0.026)	(0.027)
SAT Composite < 1000	0.005	0.001	-0.010	0.022	-0.057*	0.017	0.040
	(0.021)	(0.008)	(0.016)	(0.016)	(0.029)	(0.031)	(0.030)
SAT Composite > 1200	-0.014	0.008	-0.009	-0.009	0.013	0.050*	-0.063***
	(0.020)	(0.007)	(0.015)	(0.014)	(0.025)	(0.028)	(0.024)
Family Income < $60,000	0.030	0.010	0.030	-0.030	0.053	-0.030	-0.022
	(0.024)	(0.011)	(0.022)	(0.019)	(0.042)	(0.047)	(0.029)
Family Income > $100,000	-0.029*	-0.006	0.001	-0.001	-0.051	0.077***	-0.026
	(0.015)	(0.006)	(0.013)	(0.012)	(0.033)	(0.025)	(0.026)
Pre-Service School Characteristics							
Middle School	0.006	-0.004	0.025	0.005	-0.046	-0.033	0.079*
	(0.019)	(0.009)	(0.015)	(0.017)	(0.038)	(0.031)	(0.044)
High School	-0.033	-0.027	0.023	0.009	-0.011	-0.135	0.146
	(0.069)	(0.021)	(0.066)	(0.056)	(0.073)	(0.124)	(0.134)
Other Level School	-0.050	-0.018	-0.010	0.016	0.031	-0.109	0.078
	(0.078)	(0.013)	(0.054)	(0.044)	(0.057)	(0.075)	(0.090)
Suburban School	-0.014	0.030***	-0.033	-0.014	-0.319***	0.398***	-0.079**
	(0.019)	(0.010)	(0.026)	(0.024)	(0.046)	(0.042)	(0.031)
Town/Rural School	0.026	-0.014*	0.004	0.034	-0.253***	-0.004	0.256***
	(0.023)	(0.008)	(0.022)	(0.022)	(0.046)	(0.030)	(0.038)
Proportion FRPL Students	0.368***	0.106***	0.191***	-0.222***	0.044	-0.052	0.008
	(0.046)	(0.014)	(0.041)	(0.035)	(0.055)	(0.061)	(0.050)
N	1702	1702	1702	1702	1702	1702	1702
R^2	0.209	0.080	0.137	0.119	0.127	0.173	0.141

Notes: Results are from OLS models that predict characteristics (listed in column header) of a PST's full-time teaching school conditional on entry into Texas K–12 public schools. We report standard errors clustered by school district in parentheses.
* p < 0.1, ** p < 0.05, *** p < 0.01.

Table A9

Pearson Correlations Among Entry School Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Advantage Index	1.00									
(2) Title I SWP Eligible	-0.75	1.00								
(3) Proportion Asian Students	0.53	-0.44	1.00							
(4) Proportion Black Students	-0.33	0.29	-0.08	1.00						
(5) Proportion Hispanic/Latino Students	-0.88	0.56	-0.47	0.03	1.00					
(6) Proportion White Students	0.76	-0.47	-0.00	-0.48	-0.76	1.00				
(7) Proportion FRPL Students	-0.93	0.75	-0.52	0.38	0.80	-0.71	1.00			
(8) Urban	-0.40	0.07	0.00	0.10	0.23	-0.28	0.20	1.00		
(9) Suburban	0.37	-0.19	0.22	-0.02	-0.21	0.10	-0.29	-0.60	1.00	
(10) Town/Rural	0.02	0.14	-0.26	-0.09	-0.01	0.19	0.11	-0.40	-0.49	1.00