American Educational Research Journal
April 2017, Vol. 54, No. 2, pp. 325–359
DOI: 10.3102/0002831217690516

Does the Match Matter? Exploring Whether Student Teaching Experiences Affect Teacher Effectiveness

Dan Goldhaber
American Institutes for Research
John M. Krieg
Western Washington University
Roddy Theobald
American Institutes for Research

We use data from six Washington State teacher education programs to investigate the relationship between teacher candidates' student teaching experiences and their later teaching effectiveness. Our primary finding is that teachers are more effective when the student demographics of their current school are similar to the student demographics of the school in which they did their student teaching. While descriptive, this suggests that the school context in which student teaching occurs has important implications for the later outcomes of teachers and their students and that teacher education programs and school districts should consider placing student teachers in schools that are similar to the schools in which they are likely to teach once they enter the workforce.

Dan Goldhaber serves as the director of National Center for Analysis of Longitudinal Data in Education Research (CALDER) at American Institutes for Research and the director of the Center for Education Data & Research (CEDR) at the University of Washington Bothell. His work focuses on issues of educational productivity and reform at the K–12 level; the broad array of human capital policies that influence the composition, distribution, and quality of teachers; and connections between students' K–12 experiences and postsecondary outcomes.

John M. Krieg is professor of economics and director of the Office of Survey Research at Western Washington University. His research interests involve educational policy, teacher labor markets, and unionization in higher education.

Roddy Theobald is a senior researcher in CALDER at American Institutes for Research, 3876 Bridge Way N., Suite 201, Seattle, WA 98103; e-mail: *rtheobald@air.org*. His research interests are in teacher education, teacher evaluation, special education, and teacher collective bargaining.

KEYWORDS: educational policy, student teaching, teacher education/development

Introduction

It is well documented that teacher quality is the most important school-based factor associated with improving student achievement. Differences in teacher effectiveness (*effectiveness* and *quality* are used interchangeably here) swamp the impact of other in-school investments on student test achievement.[1] Although teacher quality is critically important, the policy mechanisms through which it may be cultivated have proved elusive. For example, teacher effectiveness is only weakly related to readily quantifiable teacher credentials like licensure status and degree level (Aaronson, Barrow, & Sander, 2007; Goldhaber, 2002; Rivkin, Hanushek, & Kain, 2005). Moreover, the empirical literature about policies designed to affect teacher quality—such as pay for performance (e.g., Fryer, Levitt, List, & Sadoff, 2012; Glazerman & Seifullah, 2010; Goldhaber & Walch, 2012; Springer et al., 2011) and professional development (e.g., Hill & Ball, 2004; Jacob & Lefgren, 2004)—focuses almost exclusively on in-service teachers (and the findings are mixed).

This focus on in-service interventions ignores the reality that a considerable amount of our country's investment in teacher workforce development occurs before teachers enter the workforce; our back-of-the-envelope calculation suggests that annual costs of teacher education a decade ago were about \$6.7 billion, while annual professional development costs from the same time period were about \$13.6 billion.[2] For greater than 80% of U.S. teachers (Feistritzer, 2010), this investment occurs in university-based teacher education programs (TEPs). Relatively little quantitative research investigates teacher preservice education (Harris & Sass, 2011), but there is a great deal of speculation that teacher education—and student teaching experiences in particular (Levine, 2006; National Council for Accreditation of Teacher Education [NCATE], 2010; Wilson, Floden, & Ferrini-Mundy, 2001)—has a powerful influence on a teacher's later success.

The theory of action connecting student teaching to teacher effectiveness is simple: For most prospective teachers, the student teaching requirement is the single prolonged experience they will have in an actual classroom before the management and learning of students becomes their primary responsibility. This is reflected in a report by NCATE (2010) that identifies student teaching as the most important aspect of a highly effective clinical program as well as empirical evidence suggesting that specific aspects of student teaching (Boyd et al., 2006; Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Ronfeldt, Schwartz, & Jacob, 2014) and characteristics of the school in which student teaching occurs (Ronfeldt, 2012, 2015) are predictive of teacher effectiveness.

In this article, we investigate a single broad research question:

> *Research Question:* What is the relationship between candidates' student teaching experiences—both where student teaching occurs and teachers mentor student teachers (cooperating teachers)—and their later teaching effectiveness?

Specifically, we utilize detailed information on prospective teachers and their student teaching experiences from six TEPs in Washington State matched with K–12 administrative data about students and teachers to investigate whether any specific student teaching experiences are predictive of value-added estimates of teacher effectiveness.

Our primary finding is that teachers are more effective when the student demographics of their current school are similar to the student demographics of the school in which they student taught. This finding is consistent across various robustness checks, including models that use data on teacher candidates who do not enter the workforce to account for potential bias associated with selection into the public school teacher workforce but is ultimately descriptive due to the nonrandom sorting of teachers into both their student teaching positions and first teaching jobs. With that said, the results lend credence to the hypothesis that the school in which student teaching occurs has important implications for the later outcomes of teachers and their students and that TEPs (and the school systems with which they partner) should consider candidates' future teaching plans in determining student teaching assignments.

The remainder of the article proceeds as follows. Next, we review the existing empirical evidence linking student teaching experiences to teacher effectiveness and then introduce the data set that allows us to build on this prior work. Then, we present our analytic approach, review our findings, and offer some conclusions.

Background

In 2010, a National Research Council report concluded that we know relatively little about how specific approaches to teacher preparation, including student teaching, are related to the effectiveness of teachers in the field.[3] Since then, a number of studies have investigated differences in effectiveness associated with acquiring a teaching credential from a specific TEP (Gansle, Noell, & Burns, 2012; Goldhaber, Liddle, & Theobald, 2013; Koedel, Parsons, Podgursky, & Ehlert, 2015; Lincove, Osborne, Dillon, & Mills, 2013; Mihaly, McCaffrey, Sass, & Lockwood, 2013). While there is some variation across studies in how much of teacher effectiveness can be attributed to TEPs, these studies show that the vast majority of the variation in teacher effectiveness is within rather than between TEPs (Goldhaber et al., 2013). An important caveat regarding these estimates is that they do not isolate the impact of the TEPs themselves from various selection mechanisms (discussed in the following) associated with prospective teachers' enrollment in training programs, entrance into the workforce, or employment in specific schools and classrooms.[4]

A second strand of literature, more closely related to this study, focuses on specific teacher education and student teaching experiences. Harris and Sass (2011) consider several aspects of teacher education (e.g., the number of courses required in different areas) but find practically no evidence of a relationship between these observable aspects of teacher education and future teacher effectiveness. Prior work with the same data set used in this study (Goldhaber, Krieg, & Theobald, 2014; Krieg, Theobald, & Goldhaber, 2016) focuses on the relationship between student teaching experiences and workforce entry. Goldhaber, Krieg, et al. (2014) find that characteristics of individual prospective teachers (e.g., endorsement area and race) are more predictive of whether they find a teaching job than characteristics of their student teaching placement or cooperating teacher. Krieg et al. (2016), by contrast, find that the *location* of prospective teachers' student teaching is far more predictive of the location of their first job than the location of their TEP or high school, suggesting that the well-known "draw of home" phenomenon in teacher hiring (Boyd, Lankford, Loeb, & Wyckoff, 2005) also operates through student teaching assignments.

Our study builds closely on recent work by Matt Ronfeldt and colleagues that connects student teaching experiences to teacher effectiveness. Ronfeldt (2012) suggests that student teaching experiences may be linked to later teacher effectiveness and in particular that the effectiveness of novice teachers is significantly higher if they student taught in "high-functioning" schools with lower teacher turnover. Ronfeldt et al. (2014) find that teachers who completed more hours of student teaching as part of their teacher education have higher self-assessments of teaching preparedness than other teachers. And most recently, Ronfeldt (2015) collected detailed data about internship schools and finds that the level of teacher collaboration in these schools (and to a lesser extent, the amount of teacher turnover in the school) is also predictive of later teacher effectiveness.

We contribute to the sparse literature relating teacher education experiences to the estimated effectiveness of in-service teachers. We pay particular attention to the match between the student demographics in a teacher's current teaching position and the student demographics in the teacher's student teaching school. This focus is motivated both by theory and empirical evidence that such a match matters. Teacher educators have long hypothesized that the setting of student teaching matters, a position exemplified by Haberman and Post (1998) and Haberman (1995) who—because of well-documented trends of new teachers being hired into poorly functioning schools—propose that teachers who are likely to teach in these schools should be trained "in the worst schools and under the poorest conditions of practice" (Haberman, 1995). Though not directly related to student demographics, this argument suggests that the experience of student teaching in a certain kind of school provides candidates with human capital that is specific to that kind of school.

This argument is bolstered by empirical evidence about in-service teachers suggesting that returns to teacher experience depend on the specifics of the experience. For instance, Ost (2014) finds that teachers improve faster early in their careers if they do not switch between grades, while Atteberry, Loeb, and Wyckoff (2016) find similar results for switching between different types of teaching assignments. Similarly, there is evidence that the returns to experience can vary depending on the percentage of minority students (Steele, Pepper, Springer, & Lockwood, 2015), percentage of free/reduced-price lunch (FRL) students (Sass, Hannaway, Xu, Figlio, & Feng, 2012), and supportiveness of the professional environment (Kraft & Papay, 2014) in a teacher's school. Moreover, it is well established that the returns to teacher experience tend to be greatest early in a teacher's career (e.g., Rice, 2013; Rivkin et al., 2005; Rockoff, 2004), and student teaching is, by design, the first time that teacher candidates have significant classroom responsibilities. It is therefore not a great leap to suggest that the nature of this student teaching experience—namely, how specific it is to the context in which teacher candidates will find themselves when they are employed as teachers, or in our terms, *the match*—could have important implications for teacher effectiveness.

To our knowledge, only the recent work by Ronfeldt (2015) puts these arguments to the test (although Ronfeldt, 2012, discusses these arguments extensively as justification for considering attributes of a candidate's student teaching school as predictors of effectiveness). Specifically, Ronfeldt quantifies the match between internship schools and first-job schools by computing the absolute difference between internship and first-job characteristics (e.g., the percentage of FRL students and the percentage of students in different race categories) and finds some statistically significant evidence that teachers who student taught in a school with a similar percentage of FRL students as their current school are more effective than other teachers. However, this finding does not hold when controlling for other measures of the match and does not extend to differences in school racial composition. As we discuss in the following, we replicate the approach of Ronfeldt but also extend these models by allowing for a more flexible relationship between the characteristics of a teacher's current and student teaching school.

Further, our study, by virtue of having data on all teacher candidates from select TEPs (not just those who end up employed as public school teachers), is the first to explicitly account for bias associated with selection into the workforce. But it is important to recognize that there are other types of selection that might bias the estimated relationships between student teaching experiences and teaching effectiveness. Teacher candidates are nonrandomly selected into education programs (Goldhaber et al., 2013; Mihaly et al., 2013); for instance, teacher candidates with varying degrees of unobserved teaching potential might systematically sort into particular types of training institutions. Teacher candidates are also nonrandomly assigned to internship experiences (Krieg et al., 2016), so stronger or weaker teacher candidates might be systematically matched with particular internship schools or mentor teachers. Finally, teachers are nonrandomly sorted into different teaching positions, and the possibility that selection into particular teaching assignments might influence teacher effectiveness estimates has received a good deal of attention in the literature (e.g., Chetty, Friedman, & Rockoff, 2014; Kane, McCaffrey, Miller, & Staiger, 2013; Rothstein, 2010).

We attempt to account for these potential sources of bias by including in our models a rich set of covariates and estimating a variety of analytic models. However, given the various selection mechanisms at play, it is important to be cautious about strong causal interpretations of our findings.

Data and Summary Statistics

Data

The analytic data set we utilize combines information about teacher candidates (also called interns, prospective teachers, or student teachers) and their student teaching experiences from six Washington State TEPs that primarily serve the western half of the state (see Figure 1)—Central Washington University (CWU), Pacific Lutheran University (PLU), University of Washington-Bothell (UWB), University of Washington-Seattle (UWS), University of Washington-Tacoma (UWT), and Western Washington University (WWU)—with K–12 data provided by Washington State's Office of the Superintendent of Public Instruction (OSPI). These TEPs provided information on each teacher candidate who completed a student teaching internship during a specific range of years, though the range of years for which data were available varies by TEP. TEPs also provided the academic year of the internship, building and district in which the internship occurred, and name of the cooperating teacher supervising the internship.[5]

The earliest individuals considered in this study completed their student teaching in 1998, while the most recent student taught in 2010.[6] Figure 2 shows the frequency of observations by student teaching year, years each TEP provided data for their student teachers, and proportion of these interns who eventually appear in the "VAM sample" (as described in the following, 23% of hired interns teach math in subjects that allow for the estimation of value-added models, or VAMs). The TEPs in our sample graduate roughly one-third of the teachers who enter the Washington State teaching workforce each year and include three of the four largest TEPs in the state (as measured by the average annual number of workforce entrants from each program).[7]

We merge the TEP data with administrative data provided by OSPI containing annual information about employment, years of experience, race, and educational background for every K–12 public school employee in

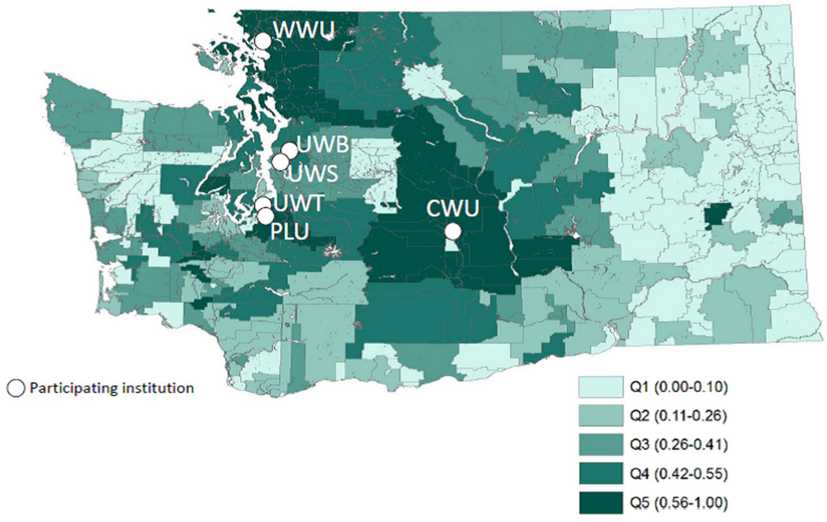

Figure 1. **Proportion of new teachers from participating institutions.**

Note. Figure 1 illustrates the proportion of newly hired teachers in each district over the past 10 years who graduated from one of the six participating institutions in this study. The legend shows how these proportions are binned into five quintiles.

the state between 1994 and 2013 as well as endorsements (the training specialty recognized by the state) for all individuals who are credentialed before 2014. We merge the aforementioned sources of data to both interns and their cooperating teachers, allowing us to consider observable characteristics of cooperating teachers as predictors of outcomes in each year (if any) that interns are observed as teachers in the state's public teaching workforce.

In addition to individual-level information on student teachers and their supervisors, we make use of annual OSPI data on public schools in Washington, which allow us to consider characteristics of both the school in which the student teacher trained and the school in which they were hired (for those who appear in the state's public teaching workforce). In particular, we characterize schools according to two common proxies for the level of student disadvantage in the school—the percentage of students eligible for free or reduced priced lunch and the percentage of American Indian, Black, or Hispanic students (underrepresented minority, or URM),—and following Ronfeldt (2012), we also calculate the "stay ratio" of each school.[8] The stay ratio is a school-level measure of teacher turnover, which we define in each school year as the percentage of the school's non–retirement age teachers who return to the school in the following year.[9] Therefore, schools with *less* teacher turnover have a *higher* stay ratio.

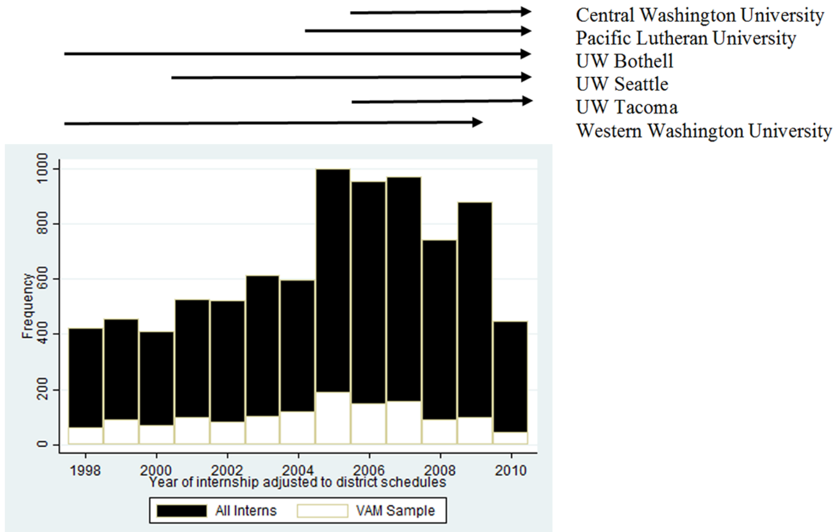

Figure 2. **Student teaching assignments by year of student teaching.**

To investigate predictors of student achievement, we merge these data to student-level data, also provided by OSPI. From 2006 through 2008, students in Grades 4–6 can be linked to their classroom teacher by their proctor on the state exam and to both current year and prior year test scores.[10] Since 2009, the state's Comprehensive Education Data and Research System (CEDARS) longitudinal data system allows all students to be linked to their classroom teachers, and current year and prior year test scores are available for students in Grades 4–8.[11] Our final analytic data set (the "VAM sample") consists only of interns who enter the public teaching workforce and are linked to student-level data in a grade and year in which both current and prior year math test scores are available. We use these test scores and additional demographic information about students to estimate value-added models described in the following.

Summary Statistics

The intern-level data set consists of 8,269 interns, each of whom completed student teaching in a Washington State public school and received a teaching credential and endorsements to teach in Washington K–12 public schools. Of these interns, 2,393 are never observed as teachers in Washington State public schools before 2013, while another 4,571 do enter the public teaching workforce but are never observed in the VAM sample. Although these interns do not factor directly into our analysis (since we do not observe outcomes for any of these interns), they do play an important role in allowing us to account for the potential for sample selection bias. Table 1 compares summary statistics for the three groups of interns: interns who never enter the workforce ("not hired"), interns who enter the workforce but are never in the VAM sample ("hired, non-VAM"), and interns who appear in the VAM sample ("hired, VAM"). Panel A compares variables that are observed before interns enter the workforce (with means calculated in the intern level), while Panel B compares variables that are only observed for hired teachers once they enter the workforce (with means calculated at the teacher-year level). Panel B only considers years in which teachers in the VAM sample are linked to student data; approximately 30% of teachers in the VAM sample are linked to student data for only one year, another 22% are observed for two years, 18% for three years, and the remaining 30% are in the VAM sample for four or more years.

Table 1 illustrates that the 1,349 interns (and 3,385 teacher-year observations) in the VAM sample are not particularly representative of all interns, hired or otherwise, which reinforces the importance of accounting for sample selection. For instance, a much higher percentage of interns are endorsed in elementary education in the VAM sample (84.9%) than in the non-VAM samples, which is due to elementary interns being more likely to be linked to student achievement data and less likely to enter the teaching workforce (Goldhaber, Krieg, et al., 2014). Not-hired interns tend to student teach in schools that have less teacher turnover than hired teachers, which may be due to the significant percentage of interns, almost 15%, who are hired into the very schools in which their internships occur (Goldhaber, Krieg, et al., 2014). Also importantly, the VAM sample contains fewer novice teachers and more experienced teachers than the sample of hired interns who are not in the VAM sample. This is because while we observe student teaching placements going back to 1998, the first year we can observe teachers linked to the students in their classrooms is 2006.

We now turn our attention to the 5,876 hired interns in the data set and more importantly, the 1,349 interns in the VAM sample. One important trend, illustrated in Table 2, is that these interns tend to student teach in schools that are more advantaged (as measured by the percent of URM and FRL students in the school) and have less teacher turnover (as measured by the stay ratio) than the schools in which they find their first teaching jobs. Figure 3 shows a scatterplot of the standardized %FRL of the internship (x-axis) and first job (y-axis) school for each intern in the VAM sample. The estimated linear relationship is positively sloped (r = 0.47), meaning that interns who student teach in disadvantaged schools tend to find first jobs in disadvantaged schools and vice versa, but there are many teachers who student teach in very different schools than the schools in which they begin their teaching careers (i.e., are in the top left corner or bottom right corner of Figure 3). Moreover, 60% of these interns find their first job in a school with a higher

Table 1
Summary Statistics

	1 Not Hired	2 Hired, Non-VAM	3 Hired, VAM
Panel A: Preservice variables			
Age in internship year	28.820	27.797***	28.722
	(8.806)	(7.623)	(8.148)
Male	0.221	0.242	0.231
Non-White	0.098	0.090	0.093
STEM endorsement	0.088***	0.117***	0.204
SPED endorsement	0.050***	0.143***	0.084
Elementary endorsement	0.680***	0.575***	0.849
Internship school stay ratio	–0.144*	–0.202	–0.191
	(0.659)	(0.626)	(0.612)
Internship school %URM	–0.016	–0.029	–0.023
	(0.782)	(0.790)	–(0.825)
Internship school %FRL	–0.110	–0.173	–0.148
	(0.830)	(0.844)	(0.890)
Cooperating teacher experience	15.233	14.947	14.761
	(8.810)	(8.659)	(8.380)
Cooperating teacher advanced degree	0.630	0.617	0.623
Cooperating teacher non-White	0.087	0.098	0.093
Cooperating teacher male	0.219*	0.245***	0.184
Cooperating teacher prior interns	0.530**	0.396	0.414
	(1.145)	(0.948)	(0.922)
Number of unique interns	2,393	4,561	1,349
Panel B: In-service variables			
Time from internship to first job		1.502	1.505
		(1.107)	(1.020)
No years of experience		0.132***	0.074
One year of experience		0.156***	0.105
Two years of experience		0.140**	0.127
Three years of experience		0.123	0.126
Four years of experience		0.104***	0.123
Five plus years of experience		0.345***	0.444
Same school as internship		0.137*	0.123
Number of teacher-year observations		29,952	3,385

Note. Standard deviations are in parentheses. VAM = value-added model; FRL = eligible for free or reduced priced lunch; URM = underrepresented minority.
*p < .05. **p < .01. ***p < .001.

percentage of FRL students than their internship school (i.e., are above the 45° line in Figure 3).

Table 2
Summary Statistics for Internship and First Job School Characteristics

	1	2	3
	All Hired Teachers		
Sample	Internship School	First Job School	Difference
Standardized stay ratio	–0.200	–0.351	–0.151***
	(0.621)	(0.610)	(0.780)
Standardized %URM students	–0.028	0.139	0.166***
	(0.798)	(0.995)	(0.930)
Standardized %FRL students	–0.168	–0.016	0.154***
	(0.855)	(0.973)	(0.978)
Number of unique teachers	5,876	5,876	5,876
	All Hired VAM Teachers		
Sample	Internship School	First Job School	Difference
Standardized stay ratio	–0.191	–0.392	–0.201***
	(0.612)	(0.576)	(0.746)
Standardized %URM students	–0.023	0.165	0.188***
	(0.825)	(1.012)	(0.927)
Standardized %FRL students	–0.148	0.037	0.185***
	(0.890)	(0.987)	(1.008)
Number of unique teachers	1,349	1,349	1,349

Note. Standard deviations are in parentheses. VAM = value-added model; FRL = eligible for free or reduced priced lunch; URM = under-represented minority.
***$p < .001$ (p values from two-sided t test of difference between columns).

Analytic Approach

Our objective is to investigate the relationship between interns' student teaching experiences and their effectiveness. We first define Z_j as a vector of student teaching experiences for intern j (the specific variables in this vector vary across model specifications). To investigate the relationship between each of these variables and teacher effectiveness, we must consider the implications of four potential sources of bias discussed previously. First, individuals are nonrandomly selected into different teacher education programs (Goldhaber, 2014). Second, teacher candidates are nonrandomly sorted into different student teaching positions (Krieg et al., 2016; Maier & Youngs, 2009). Third, teacher candidates are nonrandomly selected into the public teaching workforce (Goldhaber, Krieg, et al., 2014). And finally, teachers are nonrandomly sorted into different teaching positions (Goldhaber, Lavery, & Theobald, 2015). Each of these sources of nonrandom variation could lead to bias if they mean that individuals who have different

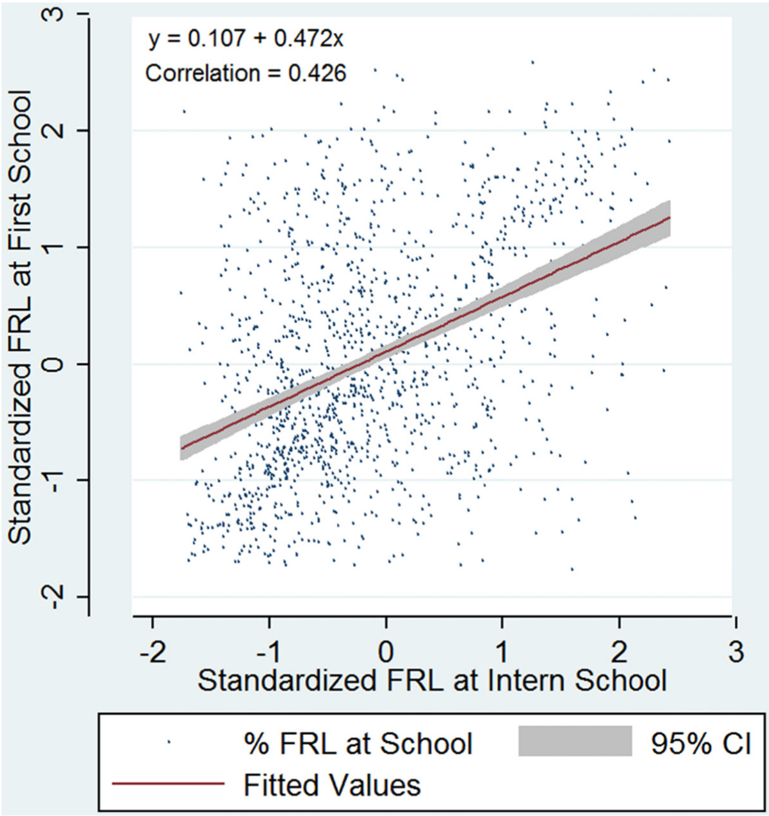

Figure 3. **Standardized percent free or reduced price lunch (FRL) at internship and first job school.**

teacher education, student teaching, or teaching experiences are different in other ways that also influence their effectiveness.

As we describe in the following, we attempt to address these potential sources of bias in several ways. We estimate VAMs that include a rich set of variables that control for sorting into teacher education programs, student teaching assignments, and teaching positions along observable dimensions. Specifically, all models control for candidate-level variables shown to be predictive of a candidate's internship and first job assignments and future teaching effectiveness (Goldhaber et al., 2013; Krieg et al., 2016). And because we know the TEPs from which teacher candidates graduated, we also include program-level fixed effects, which account for candidate selection into programs. Finally, we estimate two-stage "Heckit" models (described in the next subsection) that attempt to correct for sample selection bias that could result from only observing outcomes for teacher candidates who ultimately enter the teaching workforce and appear in the VAM sample. However, because of the number of unobserved factors at play, we do not believe that these models fully account for the nonrandom sorting of individuals to student teaching schools and their teaching positions, so we interpret our results as descriptive estimates that represent a combination of causal effects and the influence of nonrandom sorting.

Analytic Models

For the subset of 1,351 interns in the VAM sample, we investigate whether student teaching experiences are related to teacher effectiveness by estimating value-added models. Specifically, we predict student achievement on Washington State's standardized math exams as a function of lagged student achievement, other student covariates that are correlated with student test performance, and the vector of student teaching experiences of the student's teacher:

$$Y_{ijt} = \alpha_0 + \alpha_1 Y_{i(t-1)} + \alpha_2 X_{it} + \alpha_3 T_{jt} + \alpha_4 Z_j + \varepsilon_{ijt}. \tag{1}$$

In Equation 1, Y_{ijt} is the state math test score for each student i with teacher j in year t, normalized within grade and year; $Y_{i(t-1)}$ is a vector of the student's scores the previous year in both math and reading, also normalized within grade and year; X_{it} is a vector of student attributes in year t (race, gender, program participation, and eligibility for FRL); T_{jt} is a vector of individual characteristics including teacher experience dummies (summarized in Table 1) and current school characteristics (including school fixed effects in some specifications) for teacher j in year t; and Z_j is the vector of student teaching experiences for teacher j. The coefficients in the vector α_4 can be interpreted as the expected increase in student math performance associated with changes in each student teaching experience, holding all other student and teacher covariates constant. Importantly, the specification in Equation 1 assumes that the relationship between student teaching experiences and student achievement is equally strong for first-year teachers and for more experienced teachers, but we relax this assumption in extensions described in the next section. When we estimate the model in Equation 1 by ordinary least squares (OLS), we cluster the standard errors at the teacher level to account for dependence between observations associated with the same teacher (both within the same year and across different years).

Investigating the Match

In some models, we consider measures of the match between an intern's current school and student teaching school. For example, let C_{jt} be a characteristic of the current school of intern j in year t (e.g., percentage FRL), and let I_j be the comparable characteristic of the internship school of intern j. We experiment with a number of different measures of the similarity between C_{jt} and I_j, including (following Ronfeldt, 2015) the absolute difference $\left|C_{jt} - I_j\right|$ between the standardized characteristic of the intern's current and internship schools. However, in our primary results, we present estimates from more flexible models that include a polynomial of the difference between C_{jt} and I_j as a predictor of student achievement:

$$\gamma_1 C_{jt} + \sum_{k=1}^{3} \gamma_{k+1}\left(C_{jt} - I_j\right)^k + C_{jt} \sum_{k=1}^{3} \gamma_{k+4}\left(C_{jt} - I_j\right)^k. \tag{2}$$

The first term in Equation 2 is the main effect of the current school characteristic, the second term is a polynomial of the match between current and internship school characteristics, and the third term interacts this polynomial with the main effect of the current school characteristic. In other words, the first term controls directly for the characteristic of the intern's current school, the second allows the difference between the characteristic of the intern's current and internship school to have a nonlinear relationship with teaching effectiveness, while the third allows this relationship to vary depending on the characteristic of the intern's current school. Note that since both current and internship school characteristics are standardized within grade and year, the interaction effects can be interpreted as the interaction for teachers in the "average" school in the state. When we estimate match models, we add Equation 2 as additional independent variables to Equation 1.

Sample Selection Correction

As we describe previously, we can use data on interns who are not in the VAM sample to estimate Heckit models that account for sample selection (e.g., Heckman, 1979). There are two potential sources of sample selection bias: bias associated with the nonrandom selection of teacher candidates into the public teaching workforce and bias associated with the nonrandom selection of teachers into grades and subjects in which we can estimate value-added models (i.e., the VAM sample). While both sources of bias are potentially problematic, we are primarily concerned with the first (selection into the workforce) for two reasons. First, our prior work (Goldhaber, Krieg, et al., 2014) has demonstrated that interns with specific student teaching experiences (e.g., student teaching in a school with higher teacher turnover) are more likely to enter the workforce, and we hypothesized in that article that this nonrandom selection could bias the estimates from existing studies (e.g., Ronfeldt, 2012) relating student teaching experiences to teacher effectiveness. Second, the nonrandom selection of teachers into "tested" grades and subjects is endemic to the value-added literature, and we do not intend to generalize our estimates beyond these grades and subjects.

With this in mind, we identify two instrumental variables (IVs) that are predictive of the probability that interns enter the public teaching workforce but (we assume) are otherwise unrelated with teacher effectiveness: the number of new teachers hired in the intern's internship school in the year immediately following his or her internship and an indicator for whether the principal at the intern's internship school graduated from the same TEP as the intern. The first IV is motivated by the observation that many interns are hired into the same school where they student taught, so interns may be more likely to be hired as teachers if there are "slots" available in their internship school right after their internship ends. The second IV is motivated by the importance of social networks in teacher hiring (e.g., Maier & Youngs, 2009), on the assumption that that interns who attended the same institution as the principal of their internship school may have access to a better social network for finding a permanent position.

In the next section, we show that these IVs are highly predictive of entrance into the public teaching workforce.[12] These estimates come from a first-stage probit regression:

$$\Pr(O_j=1)=\Phi(\nu_0+\nu_1 IV_j+\nu_2 T_j+\nu_3 I_j+\varepsilon_j). \tag{3}$$

In the primary version of Equation 3 that we use in the Heckit models, $O_j = 1$ if intern j is observed in the teaching workforce, IV_j is the vector of IVs for intern j, and T_j and Z_j are vectors of individual and student teaching variables, respectively.

We must make the (untestable) assumption that the IVs are not otherwise correlated with the outcomes of the analytic models (the exclusion restriction). This assumption could be violated for each IV. For example, if schools that know they will need to hire a teacher in the following year are more able to get motivated and/or high-quality student teachers than other schools, then the exclusion restriction is violated for the first IV. Likewise, if TEPs are more likely to send their most motivated and/or high-quality student teachers to schools where the principal graduated from that TEP, then the exclusion restriction is violated for the second IV.[13]

Under the exclusion restriction, we can use the first-stage coefficients estimated from Equation 3 to form a selection correction term that can be included as an additional covariate in a Heckit model that accounts for nonrandom sample selection (Vella, 1998):

$$\pi_j=\frac{\varphi(\nu_0+\nu_1 IV_j+\nu_2 T_j+\nu_3 I_j+\varepsilon_j)}{\Phi(\nu_0+\nu_1 IV_j+\nu_2 T_j+\nu_3 I_j+\varepsilon_j)}. \tag{4}$$

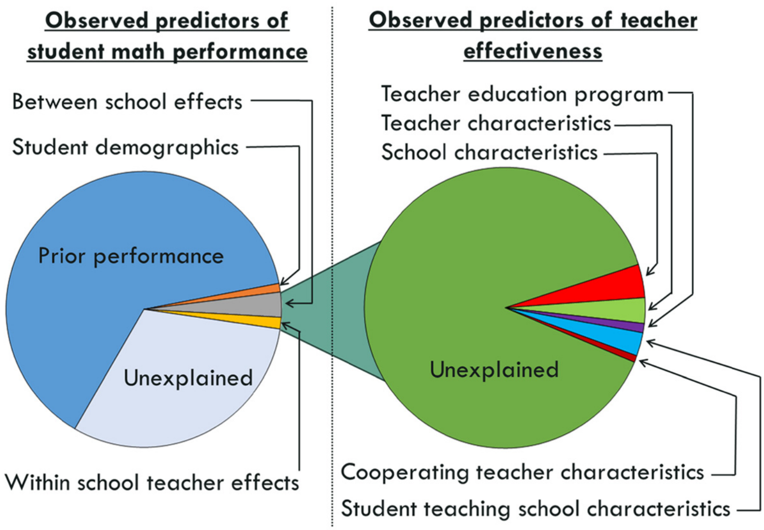

Figure 4. **Summary of ANOVA of student test performance.**

In Equation 4, φ denotes the standard normal density function, while Φ denotes the standard normal cumulative distribution function. Because the same correction term is used for each observation associated with the same intern, we calculate the standard error of each coefficient using a bootstrap procedure described in Winters, Dixon, and Greene (2012) and Goldhaber, Grout, and Huntington-Klein (2014).[14]

Results

Before discussing estimates from our analytic models, we first investigate the overall importance of student teaching experiences and the match between student teaching and current schools in terms of the overall variation in student math performance. As discussed previously, earlier work (Goldhaber et al., 2013) finds that TEP institutions themselves explain relatively little of the variation in teacher effectiveness in Washington State (less than 1%). To put the findings from this section in a similar context, we estimate Equation 1 as an ANOVA model that sequentially removes variation in student math performance due to different kinds of student-, teacher-, and school-level variables. Figure 4 summarizes our conclusions from this exercise. The pie chart on the left shows that similar to other studies (e.g., 16% in Rivkin et al., 2005), teachers and schools explain about 12% of the variation in student math performance that is not explained by prior test scores.

When we decompose this portion of the variation into parts associated with different teacher and school variables in the second pie chart in Figure 4, we see that—as in Goldhaber et al. (2013)—differences across TEP institutions in our sample explain about only about 1% of the variation in teacher effectiveness. On the other hand, the student teaching and cooperating teacher variables in our models explain over 3% of the variation, even after removing all variation at the TEP level.[15] This suggests, sample selection issues aside, that differences in student teaching experiences between teacher candidates within the same program are far more predictive of future effectiveness than average differences between teacher candidates from different programs[16] and motivates the focus in this study on within-institution variation in student teaching experiences.

We now turn to the coefficient estimates from different specifications of Equation 1. Table 3 presents estimates from specifications of the model in Equation 1 that replicate the specifications reported in Ronfeldt (2012). We focus on the estimates associated with the three internship school characteristics discussed previously but control for all the variables listed at the bottom of the table. In general, we find little evidence that characteristics of the school where an intern student taught are uniformly predictive of future teaching effectiveness, although interns who student taught in schools with higher percentages of FRL and URM students are more effective when compared to other teachers within the same school (i.e., in the school fixed effect models).[17] These findings differ from those of Ronfeldt, who finds a positive and statistically significant relationship between the internship school stay ratio and teacher effectiveness. It is not clear what is driving these differences, but one possibility is that it relates to differences across study sites. For instance, Ronfeldt's study uses data from New York City Public Schools, which has over 15 times as many schools as the largest school district in Washington State (Seattle). Given the correspondingly greater opportunity for teachers to move between schools in New York City, the stay ratio may be a better measure of the "functionality" of a school in New York City than in Washington State.

In Table 4, we extend the specifications from Table 3. In all columns, we include an indicator for whether a student's teacher is employed in the same school where they student taught and find a small positive (but not statistically significant) relationship between having one of these teachers and student performance. This suggests that even if schools are using student teaching as a "screening process" in teacher hiring (as suggested in Goldhaber, Krieg, et al., 2014), they are missing an opportunity to use this process to identify and hire more effective teachers.

In the even columns of Table 4, we consider characteristics of the intern's cooperating teacher (including measures of the gender and racial match between the cooperating teacher and intern; note that once we add interactions, the main effect for cooperating teacher male is the effect for

Table 3
Math Effectiveness Models (Value-Added Model of Student Math Performance)

	1	2	3	4	5	6	7	8
Internship school stay ratio	0.001			0.003	–0.005			–0.001
	(0.010)			(0.010)	(0.012)			(0.012)
Internship school %URM		0.007		–0.009		0.020*		0.002
		(0.010)		(0.017)		(0.010)		(0.019)
Internship school %FRL			0.011	0.017			0.020**	0.019
			(0.010)	(0.016)			(0.009)	(0.016)
Current school controls	X	X	X	X				
Current school fixed effects					X	X	X	X
N	112,985	112,985	112,985	112,985	113,024	113,024	113,024	113,024

Note. All models include students' prior test scores interacted with their grade, race, gender, and program participation and teacher internship year, institution, student teaching quarter, experience, endorsements, time to hire, gender, age, minority status, school enrollment, school level, school urbanicity, current school %FRL, %URM, and stay ratio controls. Standard errors are in parentheses and clustered at the teacher level. FRL = eligible for free or reduced priced lunch; URM = underrepresented minority.
*$p < .10$. **$p < .05$ (p values from two-sided t test).

Table 4
Math Effectiveness Model Extensions

	1	2	3	4
Internship school stay ratio	0.003	0.003	0.001	–0.000
	(0.010)	(0.010)	(0.013)	(0.012)
Internship school %URM	–0.010	–0.009	–0.001	–0.003
	(0.017)	(0.017)	(0.019)	(0.019)
Internship school %FRL	0.018	0.019	0.020	0.021
	(0.016)	(0.016)	(0.016)	(0.016)
Same school as internship (relative to different school)	0.027	0.022	0.015	0.019
	(0.021)	(0.022)	(0.021)	(0.021)
Student teacher male (relative to female)	0.014	0.004	0.031**	0.042**
	(0.014)	(0.016)	(0.015)	(0.017)
Student teacher Non-White (relative to White)	–0.032	–0.034	–0.001	0.002
	(0.020)	(0.021)	(0.021)	(0.022)
Cooperating teacher experience		0.000		–0.000
		(0.001)		(0.001)
Cooperating teacher advanced degree		–0.029**		–0.030**
		(0.013)		(0.012)
Cooperating teacher male (relative to female)		–0.026		0.035*
		(0.019)		(0.021)
Cooperating teacher non-White (relative to White)		0.001		0.034
		(0.029)		(0.026)
Cooperating teacher prior interns		0.012		–0.000
		(0.007)		(0.007)
Cooperating Teacher Male × Intern Male		0.033		–0.055
		(0.033)		(0.036)
Cooperating Teacher Non-White × Intern Non-White		0.046		–0.029
		(0.065)		(0.057)
Current school controls	X	X		
Current school fixed effects			X	X
N	112,985	112,985	112,985	112,985

Note. All models include students' prior test scores interacted with their grade, race, gender, and program participation and teacher internship year, institution, student teaching quarter, experience, endorsements, time to hire, gender, age, minority status, school enrollment, school level, school urbanicity, current school %FRL, %URM, and stay ratio controls. Standard errors are in parentheses and clustered at the teacher level. FRL = eligible for free or reduced priced lunch; URM = underrepresented minority.
*$p < .10$. **$p < .05$ (p values from two-sided t test).

male cooperating teachers working with female student teachers, etc.).[18] The only consistent finding from these specifications is that interns whose cooperating teacher held an advanced degree are less effective than other teachers, all else equal. Given that these specifications also control for cooperating teacher experience, it is difficult to know what to make of this finding. Two plausible explanations are that the cooperating teachers with advanced degrees are systematically different from cooperating teachers without advanced degrees or that students assigned to cooperating teachers with advanced degrees are systematically different from students assigned to cooperating teachers without advanced degrees. For instance, while there is little empirical evidence that advanced degrees are predictive of teacher effectiveness in general (Goldhaber & Brewer, 1997; Monk & King, 1994), this may not be the perception among those who are making TEP assignments. If TEPs place a very high priority on placing interns with cooperating teachers who hold advanced degrees, interns may be placed with relatively lower-quality cooperating teachers with advanced degrees. Alternatively, TEPs might place relatively weaker interns with cooperating teachers with advanced degrees in the hope that these cooperating teachers will provide more support for these interns.

In Table 5, we present estimates from specifications that consider the match between internship and current school characteristics as predictors of teaching effectiveness (in which the polynomial in Equation 2 is added to the model in Equation 1). We first let I_j and C_{jt} be the standardized %FRL of the teacher's internship school and current school (Column 1) and then estimate similar models that consider standardized %URM as the school measure (Column 2). Given the presence of the interaction terms in Equation 2, the main effect of %FRL in Column 1 can be interpreted as the expected change in student performance associated with a one standard deviation increase in current school %FRL, all else equal, for students whose teachers experience a perfect match between their current and internship school (i.e., $C_{jt} - I_j = 0$). From Column 1, we can see that all else equal, a one standard deviation increase in current school %FRL for these students is associated with a decrease in student performance of .023 standard deviations.

The coefficients of interest, though, are for the terms that consider the match between a teacher's internship and current school. In Table 5 and the subsequent discussion, we refer to this term as *difference*, which in Column 1 refers to the difference between %FRL of the teacher's current school and internship school (i.e, $C_{jt} - I_j$). In the case of %FRL, the only coefficient of marginal statistical significance is the linear difference measure; its negative value suggests that the greater the difference between the %FRL of a teacher's current and internship school, the lower the achievement of the teacher's students.

As discussed previously, one can imagine difference having a differential impact on student achievement depending on the characteristics of the school that currently employs the teacher. For instance, a one standard deviation difference in FRL may be large for a teacher trained in a school with very few FRL students, but it might be hardly noticeable for a teacher trained in a high FRL school. To account for this possibility, we interact the current

Table 5
Math Effectiveness Match Models

School Measure	1 %FRL	2 %URM
Current school	–0.023*	–0.015
	(0.012)	(0.013)
Difference	–0.023*	–0.024
	(0.013)	(0.015)
Difference^2	–0.012	–0.018**
	(0.008)	(0.008)
Difference^3	0.004	0.002
	(0.005)	(0.004)
Current School × Difference	0.029**	0.040***
	(0.012)	(0.011)
Current School × Difference^2	–0.000	–0.003
	(0.008)	(0.007)
Current School × Difference^3	–0.004**	–0.001
	(0.002)	(0.001)
Current school controls	X	X
N	113,001	112,985

Note. All models include students' prior test scores interacted with their grade, race, gender, and program participation and teacher internship year, institution, student teaching quarter, experience, endorsements, time to hire, gender, age, minority status, school enrollment, school level, and school urbanicity. Standard errors are in parentheses and clustered at the teacher level. FRL = eligible for free or reduced priced lunch; URM = underrepresented minority.
$^*p < .10$. $^{**}p < .05$. $^{***}p < .01$ (p values from two-sided t test).

school FRL with a polynomial in difference and report the associated coefficients in the bottom half of Table 5. A number of the resulting coefficients are statistically significant, suggesting that the differences between the %FRL of a teacher's current and internship school are related to student achievement differently for individuals in advantaged versus disadvantaged schools.

The goal of the models reported in Table 5 is to evaluate the impact of the match between the current teaching environment and the environment from where a teacher student taught. However, as the preceding discussion indicates, the match is a function of both difference and the interaction of difference with the current teaching environment. Because we have estimated both of these with polynomials, it is difficult to understand the match solely by focusing on the regression coefficients in Table 5. As an alternative, we use the coefficients from Column 1 of Table 5 to calculate the average predicted test scores across combinations of standardized internship school FRL and standardized current school FRL. We plot the resulting estimates in

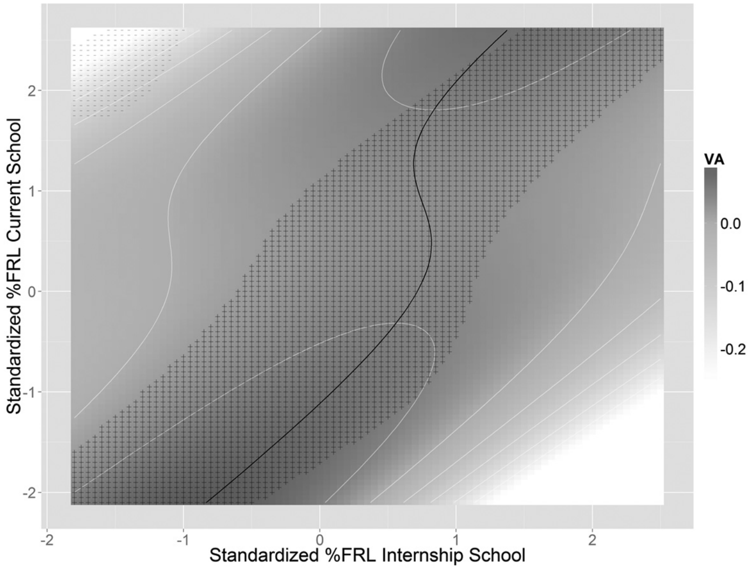

Figure 5. **Predicted math performance by internship and current school %FRL.**
Note. + denotes region that is significantly greater than zero ($p < .05$); – denotes region that is significantly less than zero. FRL = eligible for free or reduced price lunch.

the contour plot in Figure 5.[19] A similar contour plot using the URM coefficients from Column 2 of Table 5 is shown in Figure 6. In these contour plots, dark regions indicate higher test scores (where a + indicates a region in which the average predicted test score is statistically significantly greater than zero), and light regions indicate lower test scores (where a – indicates a region in which the average predicted test score is significantly less than zero). Importantly, because our goal is to assess the importance of the match between internship and current school, we do not use the coefficient on current school %FRL or %URM to produce these predicted values. Figure 5 can therefore be interpreted as follows: For a student in a school with a given %FRL (any value on the y-axis), the predicted values tell us how the estimates in Table 5 suggest that the student's test scores will vary as a function of the %FRL of his or her teacher's internship school.[20] The range of the x- and y-axis in Figures 5 and 6 represents the range of observed values in the VAM data set.

The patterns in Figure 5 are striking as for each value of current school %FRL, students tend to score higher when their teacher student taught in a school with a similar %FRL (and lower when their teacher student taught

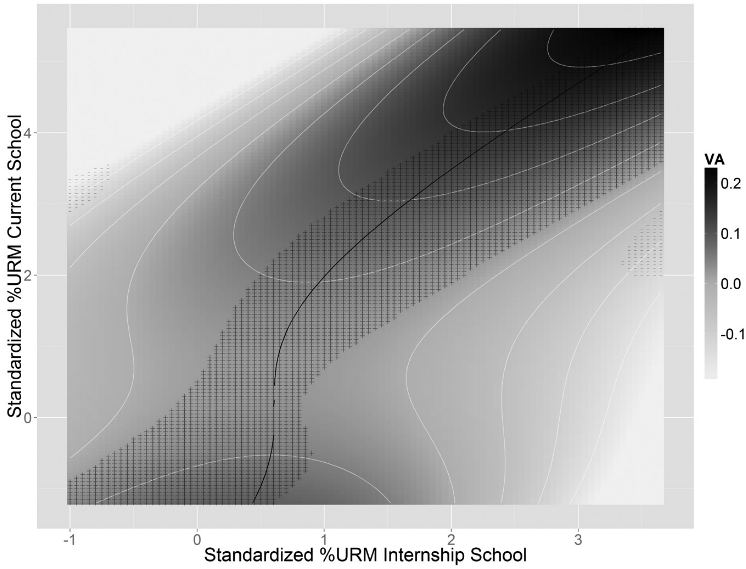

Figure 6. **Predicted math performance by internship and current school %URM.**
Note. + denotes region that is significantly greater than zero ($p < .05$); – denotes region that is significantly less than zero. URM = underrepresented minority.

in a school with a very different %FRL). These relationships are not perfectly linear, as shown by the solid line in Figure 5, which traces the maximum values of predicted student scores for each value of current school %FRL; specifically, students in very low %FRL schools score highest when their teacher student taught in a school with somewhat higher %FRL and vice versa for students in very high %FRL schools. But the large regions of Figure 5 where predicted effectiveness is statistically significant and greater than zero are generally where the %FRL of the internship school is within one standard deviation of the %FRL of the current school, while the large negative region in the top left corner is where the internship school %FRL is over three standard deviations lower than the current school %FRL. The magnitudes of the differences between the "extremes" of the figure are meaningful; for example, students in a high-poverty school (two standard deviations above the mean %FRL) are predicted to score 0.157 standard deviations higher if their teacher student taught in a school with the same %FRL than if their teacher student taught in a low-poverty school (two standard deviations below the mean %FRL). This is suggestive evidence that the match between internship and current school may matter for student achievement and that a mismatch may be particularly detrimental when teachers are in a considerably more disadvantaged school than where they student taught.

Figure 6 plots average predicted scores from Column 2 of Table 5, in which the measure of school disadvantage is %URM students. The patterns in Figure 6 are perhaps even more striking than in Figure 5, particularly in the upper right corner (suggesting that students in high %URM schools benefit greatly from having a teacher who student taught in a high %URM school). Specifically, students in a high URM school (three standard deviations above the mean %URM) are predicted to score 0.167 standard deviations higher if their teacher student taught in a school with the same %URM than if their teacher student taught in a low URM school (two standard deviations below the mean %URM). The general conclusion from Figure 6, then, is the same as for Figure 5; students tend to score higher when their teacher student taught in a school with similar student demographics.

It is important to emphasize that because of the potential sources of bias discussed previously, the relationships in Figures 5 and 6 (and the estimates in Table 5) do not imply a causal relationship between the match between internship and current school characteristics and student achievement. It is possible, for example, that the best teacher candidates are more able to find jobs in schools that are similar to where they student taught (or are more likely to get student teaching positions in schools that are similar to the schools they plan to teach in). As we discuss in the next subsection, we do not find evidence that teachers with better observable pre-service qualifications (i.e., credential test scores) are any more likely to experience a better match between their internship and current school characteristics, but we cannot rule out nonrandom sorting along unobserved dimensions. That said, the descriptive finding that students tend to perform better when their teachers have been trained in similar schooling environments is intuitive.

Sample Selection Correction and Robustness Checks

As we discuss previously, there are several potential sources of bias for the estimates from Tables 3 to 5. We account for one of these sources of bias (nonrandom selection into the teaching workforce and VAM sample) directly by estimating sample selection (or Heckit) models that rely on data about teacher candidates who never enter the public teaching workforce. Table 6 shows that the two IVs we have identified for the first stage of these models are highly predictive of entrance into the public teaching workforce both separately and jointly and in the expected direction (i.e., teacher candidates are more likely to be in the VAM sample, all else equal, if they student teach in a school that will hire more teachers the next year or in a school with a principal from the same TEP). We then use the first stage regression in the third column of Table 6 to estimate Heckit models and present the Heckit estimates in the even columns of Table 7 (alongside the OLS estimates in the odd

Table 6
First-Stage IVs (Probit Predicting Workforce Entry)

	1	2	3
Number of new teachers hired by internship school year after internship	0.040***		0.040***
	(0.012)		(0.012)
Internship principal from same institution		0.069*	0.069*
		(0.036)	(0.036)
Number of unique individuals	8,269	8,269	8,269

Note. All models include teacher internship year, institution, student teaching quarter, endorsements, gender, age, and minority status and internship school %FRL, %URM, and stay ratio controls. Standard errors in parentheses. IVs = instrumental variables; FRL = eligible for free or reduced priced lunch; URM = underrepresented minority.
*$p < .10$. ***$p < .01$ (p values from two-sided t test).

columns). The broad conclusion from this table is that the sample selection correction makes very little difference in our estimates of the relationship between student teaching experiences and student math performance.

The other primary finding in our analysis is that teachers are more effective when they teach in similar schools as where they student taught, but as we discuss previously, this could be a spurious correlation that is due to nonrandom sorting of teacher candidates to student teaching positions and teaching jobs. In particular, teacher candidates who will become more effective teachers regardless of their student teaching experiences may be more likely to experience a match between their student teaching and current school either because they are more likely to find jobs in schools that are similar to where they student taught or because they are more likely to get student teaching positions in schools that are similar to the schools they plan to teach in.

To test for this possibility (and as a falsification test of the conclusions from Figures 5 and 6), we use a pre-training measure of teacher quality available for a subset of teacher candidates in the sample—teacher credential test scores—and investigate whether teacher candidates with higher credential test scores are more likely to experience a match between their current and internship school. Specifically, we estimate similar match models from Table 5 except at the teacher-year level omitting student-level variables and with teacher credential test scores as the outcome variable. Credential test scores are an imperfect measure of teacher quality, but empirical evidence shows that they are modestly predictive of future teacher effectiveness (Clotfelter, Ladd, & Vigdor, 2010; Goldhaber, Gratz, & Theobald, 2016). So, if less qualified teachers are more likely to teach in a school that is very different than their student teaching school, we would expect measures of the match to be predictive of these credential test scores. However, while we

Table 7
Comparison of OLS and Heckit Math Effectiveness Estimates

	1 OLS	2 Heckit	3 OLS	4 Heckit
Internship school stay ratio	0.003	0.004	0.003	0.003
	(0.010)	(0.011)	(0.010)	(0.011)
Internship school %URM	–0.010	–0.005	–0.009	–0.004
	(0.017)	(0.019)	(0.017)	(0.019)
Internship school %FRL	0.018	0.012	0.019	0.014
	(0.016)	(0.017)	(0.016)	(0.017)
Same school as internship	0.027	0.029	0.022	0.025
(relative to different school)	(0.021)	(0.022)	(0.022)	(0.023)
Student teacher male	0.014	0.029	0.004	0.017
(relative to female)	(0.014)	(0.021)	(0.016)	(0.022)
Student teacher non-White	–0.032	–0.032	–0.034	–0.034
(relative to White)	(0.020)	(0.022)	(0.021)	(0.023)
Cooperating teacher experience			0.000	0
			(0.001)	(0.001)
Cooperating teacher advanced degree			–0.029**	–0.028**
			(0.013)	(0.014)
Cooperating teacher male			–0.026	–0.023
(relative to female)			(0.019)	(0.02)
Cooperating teacher non-White			0.001	0
(relative to White)			(0.029)	(0.03)
Cooperating teacher prior interns			0.012	0.012
			(0.007)	(0.007)
Cooperating Teacher Male ×			0.033	0.03
Intern Male			(0.033)	(0.034)
Cooperating Teacher Non-White ×			0.046	0.048
Intern Non-White			(0.065)	(0.069)
Current school controls	X	X	X	X
N	112,985	112,985	112,985	112,985

Note. All models include students' prior test scores interacted with their grade, race, gender, and program participation and teacher internship year, institution, student teaching quarter, experience, endorsements, time to hire, gender, age, minority status, school enrollment, school level, school urbanicity, current school %FRL, %URM, and stay ratio controls. Standard errors are in parentheses and are clustered at the teacher level in the models without sample selection correction and calculated from 1,000 bootstrapped estimates in the models with sample selection correction. OLS = ordinary least squares; FRL = eligible for free or reduced priced lunch; URM = underrepresented minority.
$^{**}p < .05$ (p values from two-sided t test for models without sample selection correction or permutation test for models with sample selection correction).

do find evidence that teachers with low credential test scores are more likely to teach in high %URM schools (which parallels earlier findings in Goldhaber et al., 2015), the relationship between each of the measures of the match from Equation 3 and credential test scores is close to zero and not statistically significant. While this suggests that there is not dramatic nonrandom sorting along observable dimensions, it is of course still possible that there is nonrandom sorting along unobserved dimensions of more effective teachers to schools that are a good match with their student teaching experience.

Finally, the match results discussed previously assume that differences (or similarities) between a teacher's current school and student teaching school are just as important for an early-career teacher as a more experienced teacher. Some evidence, however, suggests that teacher education effects decay over time (Goldhaber et al., 2013), so there is good reason to believe that the match may be more important for novice teachers. To investigate this, we estimate variants of Equation 1 that include (following Ronfeldt, 2015) the absolute difference $\left|C_{jt} - I_j\right|$ between the standardized characteristic of the intern's current and internship schools and report the estimates in Table 8. The first column of Table 8 shows that a one standard deviation increase in the absolute difference in the percentage FRL between an intern's current and student teaching school is associated with a .018 standard deviation decrease in student performance, pooled across all teachers.[21] This corresponds with the broad conclusion from Figure 5 that teachers are more effective when they teach in a school with a similar percentage of FRL students as their student teaching school.

The next two columns of Table 8 show that, as hypothesized, the match effect is almost three times larger for novice teachers (with two or fewer years of experience) than it is for more experienced teachers. Likewise, when we consider the match between the current school and internship school in terms of the percentage of URM students (Columns 4–6 of Table 8), similar patterns emerge; the relationship between the match and teacher effectiveness is stronger for novice teachers than more experienced teachers.[22] Thus, the estimates in Table 5 (illustrated in Figures 5 and 6) are actually conservative estimates of the relationship between the match and teacher effectiveness for novice teachers as they are pooled across all teachers in the sample.

Conclusions

This article contributes to a small but growing empirical evidence base about where future teachers should do their student teaching. Our main contribution relates to the match between internship schools and first jobs; our findings suggest that what makes a "good" student teaching school appears to vary depending on the type of school a teacher candidate will eventually teach in. Specifically, teachers appear to be more effective when the student demographics of their school are similar to the student demographics of the school in which they did their student teaching.

Table 8
Math Effectiveness Absolute Difference Models

	1	2	3	4	5	6
School Measure	%FRL			%URM		
Sample	All	First and Second Years	3+ Years	All	First and Second Years	3+ Years
Absolute difference	–0.019*	–0.044**	–0.016	–0.017	–0.045**	–0.015
	(0.010)	(0.018)	(0.011)	(0.012)	(0.018)	(0.013)
Unique teachers	1,359	463	1,183	1,359	463	1,183
Teacher-year observations	3,662	588	3,074	3,662	588	3,074
Current school controls	X	X	X	X	X	X

Note. All models include students' prior test scores interacted with their grade, race, gender, and program participation and teacher internship year, institution, student teaching quarter, experience, endorsements, time to hire, gender, age, minority status, school enrollment, school level, and school urbanicity. Standard errors are in parentheses and clustered at the teacher level. FRL = eligible for free or reduced priced lunch; URM = underrepresented minority.
*$p < .10$. **$p < .05$ (p values from two-sided t test).

This finding suggests several potential policy conclusions. First, it suggests that schools should consider hiring more teachers who did their student teaching in schools with similar student demographics. Importantly, this policy recommendation likely holds even if, despite our extensive robustness checks, our findings reflect nonrandom sorting to student teaching schools and teaching positions rather than the causal effect of a good match between a teacher's current job and student teaching experience. For example, suppose that a principal has the choice between hiring two applicants to the school: a candidate with a student teaching background similar to the principal's school and a candidate who student taught in a school with very different student demographics. Our results predict that the first candidate will be more effective, and from the principal's perspective, it does not matter whether this is because of a causal effect of the match, selection effects (e.g., the nonrandom sorting to student teaching positions), or some combination of the two.

A second potential policy recommendation is that TEPs should learn more about the job preferences and opportunities for their graduates and consider placing more teacher candidates into student teaching schools that look like the schools they are likely to be hired into. While this overall conclusion is supported by a number of extensions and robustness checks suggesting that our findings reflect the causal effect of a good match between a teacher's current job and student teaching experiences, this conclusion does not hold if our results reflect nonrandom sorting to student teaching schools and teaching positions (e.g., if teacher candidates who will eventually become better teachers may actively seek out student teaching positions that are similar to the positions they will eventually be hired into). Evidence from experiments that randomize teacher candidates to student teaching positions could help distinguish causal effects of the match from various potential selection effects and further support this conclusion.

If our findings do represent causal effects of the match, they could also have important ramifications in terms of equity. Specifically, teachers in our sample are much more likely to have done their student teaching in a school with a lower percentage of disadvantaged students (i.e., FRL or URM) than their current school, so disadvantaged students are less likely to have a teacher whose student teaching matches their school setting than more advantaged students. This suggests that TEPs that are committed to educating teachers who will be successful in educating disadvantaged students should consider placing more student teachers in schools that serve these students.

Finally, it is worth stressing that this analysis was made possible by linking state administrative databases to data from individual TEPs that are not typically found in these databases and illustrates the potential of similar partnerships that connect the teacher education experiences of teacher candidates to their experiences once they enter the teaching workforce. In particular, while this analysis is based on relatively coarse measures of student teaching (i.e., student teaching school and cooperating teacher), even these coarse measures are far more predictive of student performance than the information about teacher education typically contained in state administrative databases (i.e., the TEPs themselves). We therefore recommend that research continues to move toward considering the *specific teacher education experiences* of teacher candidates (and student teaching experiences in particular) to inform teacher education policies and practice.

Notes

The research presented here utilizes data supplied by the teacher education programs at Central Washington University, Pacific Lutheran University, University of Washington Bothell, University of Washington Seattle, University of Washington Tacoma, and Western Washington University. We gratefully acknowledge the receipt of these data, and we wish to thank Elly Hagen, Cameron Colbo, Kimberly McDaniel, Jim Depaepe, and Joe Koski for their assistance. This work is supported by the National Center for the Analysis of Longitudinal Data in Education Research (CALDER) (Grant No. R305C120008) and the Bill and Melinda Gates Foundation (Grant No. OPP1035217). Finally, we wish to thank Trevor Gratz for outstanding research assistance and James Cowan, Jennifer McCleery, Cap Peck, Matt Ronfeldt, and participants at the 2015 APPAM Fall Conference and 2016 AEFP Conference for helpful comments. The views expressed in this article do not necessarily reflect those of American Institutes for Research, the University of Washington, or Western Washington University. Responsibility for any and all errors rests solely with the authors.

[1]Estimates suggest, for example, that a one standard deviation increase in teacher quality raises student achievement in reading and math between 10% and 25% of a standard deviation (for estimates of the effect size associated with changes in teacher quality, see Aaronson, Barrow, & Sander, 2007; Hanushek & Rivkin, 2010). To put this in perspective, this teacher quality effect size has been found to be equivalent to lowering class size by 10 to 13 students (Rivkin, Hanushek, & Kain, 2005).

[2]Annual per-teacher professional development costs ($4,380) are reported in Miles, Odden, Fermanich, and Archibald (2004), and we estimate the costs of teacher education from the 2003–2004 school year from average college tuition costs reported by National Center for Education Statistics (NCES), weighted by the number of graduates from public and private bachelor's and master's teacher education programs calculated from Integrated Postsecondary Education Data System (IPEDS) data. This understates the per-teacher cost of teacher education since about a third of candidates never enter the teaching workforce (Goldhaber, Krieg, & Theobald, 2014). We estimate the per-teacher costs of teacher education to be about $38,000, compared to $62,000 in professional development costs over the course of an average teacher's career.

[3]At the time, there were only two large-scale quantitative studies (Boyd et al., 2006; Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009) that connected teacher education experiences to in-service teacher workforce outcomes. The first, Boyd et al. (2006), provides evidence that some aspects of student teaching, such as a capstone project where teachers relate curriculum learning to actual practices, are predictive of teacher effectiveness. In the second study, the same authors (Boyd et al., 2009) find differences in effectiveness between teachers who graduated from different teacher education programs (TEPs) and that in terms of students' math achievement in particular, teachers who identify similarities between their student teaching experience and their first-year classroom experiences have greater student achievement gains.

[4]For more detailed discussion of selection issues, see Goldhaber (2014).

[5]Internships in this article are defined as a teacher candidate's primary field experience (many interns complete additional field placements that are primarily for observational purposes). A very small number of interns from Western Washington University

completed two different student teaching internships, and for these interns, we randomly select one internship experience to include in our analytic data set.

[6]When representing years, this article uses the convention of listing the first year of the academic year. Thus, 1994 represents the 1994–1995 academic year.

[7]There are a total of 21 TEPs in Washington (for a full list, see Goldhaber, Liddle, & Theobald, 2013). Approximately 15% of the state's public school teachers were trained outside the state.

[8]We standardize these measures within year and school level to ensure that they are comparable for all interns in the sample.

[9]We follow Ronfeldt (2012) by transforming the stay ratios with an exponential transformation and standardizing within school level (elementary or secondary). Ronfeldt uses an average of each school's stay ratio over the five-year span of his data, and we experiment with several moving averages, including a three-year moving average (the current year and two prior years) and two five-year moving averages (the current year and four prior years, and the current year, two prior years, and two subsequent years). Our results use the five-year moving average calculated over the current year and four previous years, but the results are robust to the choice of average. We also experiment with other measures of the stay ratio that compare attrition in schools to attrition in the same district and Washington as a whole (the findings are little affected by our measure of the stay ratio).

[10]The proctor of the state assessment was used as the teacher-student link for at least some of the data used for analysis. The *proctor* variable was not intended to be a link between students and their classroom teachers, so this link may not accurately identify those classroom teachers.

[11]The Comprehensive Education Data and Research System (CEDARS) data include fields designed to link students to their individual teachers based on reported schedules. However, limitations of reporting standards and practices across the state may result in ambiguities or inaccuracies around these links.

[12]Note that this model controls for the internship school stay ratio, which averages the amount of teacher turnover over the past five years. Thus, the marginal effect of "number of new teachers" in this model represents the expected increase in the probability of entering the workforce for each additional teacher hired the next year, controlling for the "overall" amount of teacher turnover at the school over the past five years.

[13]We find little evidence that interns with better observable characteristics (e.g., higher credential test scores) are more likely to do their student teaching in schools that will hire a large number of teachers the next year or have a principal from the same TEP, but we cannot rule out nonrandom sorting along unobserved dimensions.

[14]For each bootstrap sample, we estimate the first-stage model only for those individuals and then estimate the second-stage model for all annual observations associated with those individuals.

[15]Note that the student teaching variables include measures of the match described later in this section.

[16]These estimates actually represent a lower bound estimate of the explanatory power of student teaching variables and cooperating teacher characteristics since we enter these variables last in the sequential ANOVA. When these variables are entered before school and teacher variables and TEP indicators, they explain over 7% of the variation in teacher effectiveness, which represents an upper bound estimate of the true explanatory power of student teaching variables and cooperating teacher characteristics.

[17]Interestingly, when we estimate these same specifications only for the subsample of interns who teach in most diverse districts represented in the data set—Seattle, Tacoma, and Kent—internship school percentage underrepresented minority (%URM) and internship school percentage eligible for free or reduced price lunch (%FRL) are both highly predictive of teacher effectiveness across all specifications. This previews the match findings discussed later in the section.

[18]Cooperating teacher race matches are exceedingly rare for non-White interns within the value-added model (VAM) sample; for example, only 1 of the 18 Black interns in the sample worked with a Black cooperating teacher, and only 2 of the 52 Hispanic interns worked with a Hispanic cooperating teacher.

[19]These average predicted scores are calculated using the margins command in STATA and are simply the mean predicted test score for all students in the sample for each combination of C_{jt} and I_j (calculated across the full range of observed values of C_{jt} and I_j at increments of .05 for each). Hypothesis tests are also performed using the margins command in STATA and test the average predicted test score at each combination of internship school characteristic and current school characteristic against zero.

[20]These estimates do not control for whether the internship school is the same as the current school, so the small positive effect of teaching in the same school as student teaching (see Table 4) is incorporated into these estimates.

[21]Given that the range of %FRL in our sample is approximately five standard deviations (see Figure 5), this estimate is roughly comparable to the estimate in Ronfeldt (2015) that a 100 percentage point increase in the absolute difference is correlated with a .093 decrease in student test performance, all else equal.

[22]We also experiment with variants of the "decay model" outlined in Goldhaber et al. (2013) that allows the effect of teacher education experiences to decay with the number of years of teaching experience. We estimate a decay parameter of .300 (meaning that the match effect for a teacher with t years of experience is $e - .3t$ times as large as the match effect of a novice teacher), but the estimate of the decay parameter is not statistically significant, so we do not pursue this model further.

References

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

Atteberry, A., Loeb, S., & Wyckoff, J. (2016). Teacher churning: reassignment rates and implications for student achievement. *Educational Evaluation and Policy Analysis*. Advance online publication. doi:10.3102/0162373716659929

Boyd, D., Grossman, P., Lankford, H., Loeb, S., Michelli, N., & Wyckoff, J. (2006). Complex by design: Investigating pathways into teaching in New York City schools. *Journal of Teacher Education*, *57*(2), 155–166.

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

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2005). The draw of home: How teachers' preferences for proximity disadvantage urban schools. *Journal of Policy Analysis and Management*, *24*(1), 113–132.

Chetty, R., Friedman, J., & Rockoff, J. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcome in adulthood. *American Economic Review*, *104*, 2633–2679.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources*, *45*(3), 655–681.

Feistritzer, E. (2010). *Alternative teacher certification: A state-by-state analysis 2010*. Washington, DC: National Center for Education Information.

Fryer, R. G., Jr., Levitt, S. D., List, J., & Sadoff, S. (2012). *Enhancing the efficacy of teacher incentives through loss aversion: a field experiment* (No. w18237). Cambridge, MA: National Bureau of Economic Research.

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.

Glazerman, S., & Seifullah, A. (2010). *An evaluation of the Teacher Advancement Program (TAP) in Chicago: Year two impact report*. Princeton, NJ: Mathematica Policy Research, Inc.

Goldhaber, D. (2002). The mystery of good teaching. *Education Next*, *2*(1), 50–55.

Goldhaber, D. (2014). *What do value-added measures of teacher preparation programs tell us?* (Carnegie Knowledge Brief 12). Retrieved from http://www.carnegieknowledgenetwork.org/wp-content/uploads/2013/11/CKN-Goldhaber-TeacherPrep_Final_11.7.pdf

Goldhaber, D., & Brewer, D. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources*, *32*(3), 505–523.

Goldhaber, D., Gratz, T., & Theobald, R. (2016). *What's in a teacher test? Assessing the relationship between teacher licensure test scores and student secondary STEM achievement* (CALDER Working Paper 158). Retrieved from http://www.cedr.us/papers/working/CEDR%20WP%202016-4.pdf

Goldhaber, D., Grout, C., & Huntington-Klein, N. (2014). *Screen twice, cut once: Assessing the predictive validity of teacher selection tools* (CEDR Working Paper 2014-9). Seattle, WA: University of Washington.

Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, *44*(5), 293–307.

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

Goldhaber, D., Krieg, J., & Theobald, R. (2014b). 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., & Walch, J. (2012). Strategic pay reform: A student outcomes-based evaluation of Denver's ProComp teacher pay initiative. *Economics of Education Review*, *31*(6), 1067–1083.

Haberman, M. (1995). Selecting star teachers for children and youth in urban poverty. *Phi Delta Kappan*, *76*, 777–781.

Haberman, M., & Post, L. (1998). Teachers for multicultural schools. *Theory into Practice*, *37*(2), 96–104.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *The American Economic Review*, *100*, 267–271.

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

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161.

Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education*, *35*, 330–351.

Jacob, B. A., & Lefgren, L. (2004). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources*, *39*(1), 50–79.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers?* Seattle, WA: Bill and Melinda Gates Foundation.

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*, 508–534.

Kraft, M., & Papay, J. (2014). Can professional development in schools promote teacher development? Explaining heterogeneity in returns to teacher experience. *Educational Evaluation and Policy Analysis*, *36*(4), 476–500.

Krieg, J., 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*, 364–388.

Levine, A. (2006). *Educating school teachers*. Retrieved from http://edschools.org/pdf/Educating_Teachers_Report.pdf

Lincove, J. A., Osborne, C., Dillon, A., & Mills, N. (2013). The politics and statistics of value-added modeling for accountability of teacher preparation programs. *Journal of Teacher Education*, *65*(1), 24–38.

Maier, A., & Youngs, P. (2009). Teacher preparation programs and teacher labor markets: How social capital may help explain teachers' career choices. *Journal of Teacher Education*, *60*(4), 393–407.

Mihaly, K., McCaffrey, D., Sass, T., & 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.

Miles, K. H., Odden, A., Fermanich, M., & Archibald, S. (2004). Inside the black box of school district spending on professional development: Lessons from five urban districts. *Journal of Education Finance*, *30*(1), 1–26.

Monk, D., & King, J. (1994). Multi-level teacher resource effects on pupil performance in secondary mathematics and science: The role of teacher subject matter preparation. In R. G. Ehrenberg (Ed.), *Contemporary policy issues: Choices and consequences in education* (pp. 29–58). Ithaca, NY: ILR Press.

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). Retrieved from http://www.ncate.org/LinkClick.aspx?fileticket=zzeiB1OoqPk%3D&tabid=7

National Research Council. (2010). *Preparing teachers: Building evidence for sound policy*. Washington, DC: National Academies Press.

Ost, B. (2014). How do teachers improve? The relative importance of specific and general human capital. *American Economic Journal: Applied Economics*, *6*(2), 127–151.

Rice, J. K. (2013). Learning from experience: Evidence on the impact and distribution of teacher experience and the implications for teacher policy. *Education Finance and Policy*, *8*(3), 332–348.

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

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, *94*, 247–252.

Ronfeldt, M. (2012). Where should student teachers learn to teach? Effects of field placement school characteristics on teacher retention and effectiveness. *Educational Evaluation and Policy Analysis*, *34*(1), 3–26.

Ronfeldt, M. (2015). Field placement schools and instructional effectiveness. *Journal of Teacher Education*, *66*(4), 304–320.

Ronfeldt, M., Schwartz, N., & Jacob, B. (2014). Does pre-service preparation matter? Examining an old question in new ways. *Teachers College Record*, *116*(10), 1–46.

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

Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, *72*(2), 104–122.

Springer, M. G., Ballou, D., Hamilton, L., Le, V., Lockwood, J. R., McCaffrey, D., … Stecher, B. M. (2011). *Teacher pay for performance: Experimental evidence from the Project on Incentives in Teaching*. Nashville, TN: National Center on Performance Incentives.

Steele, J. L., Pepper, M. J., Springer, M. G., & Lockwood, J. R. (2015). The distribution and mobility of effective teachers: Evidence from a large, urban school district. *Economics of Education Review*, *48*, 86–101.

Vella, F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources*, *33*(1), 127–169.

Wilson, S., Floden, R., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations*. Retrieved from http://depts.washington.edu/ctpmail/PDFs/TeacherPrep-WFFM-02-2001.pdf

Winters, M. A., Dixon, B. L., & Greene, J. P. (2012). Observed characteristics and teacher quality: Impacts of sample selection on a value added model. *Economics of Education Review*, *31*(1), 19–32.

Manuscript received February 3, 2016
Final revision received December 6, 2016
Accepted December 21, 2016