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

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

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

AIR

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

Dan Goldhaber
John M. Krieg
Roddy Theobald

WORKING PAPER 149 • JANUARY 2016

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

Dan Goldhaber
American Institutes for Research
University of Washington-Bothell

John M. Krieg
Western Washington University

Roddy Theobald
American Institutes for Research

Contents

Acknowledgements

This research was made possible in part by generous support from the the Bill and Melinda Gates Foundation (grant #OPP1035217), the National Center for the Analysis of Longitudinal Data in Education Research (CALDER), and an anonymous foundation. Research supported by CALDER is funded through Grant R305C120008 to the American Institutes for Research from the Institute of Education Sciences, U.S. Department of Education.

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. We also wish to thank Trevor Gratz for outstanding research assistance and James Cowan, Cap Peck, Matt Ronfeldt, and participants at the 2015 APPAM Fall Conference for helpful comments.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007
202-403-5796 • www.caldercenter.org

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

Dan Goldhaber, John M. Krieg, Roddy Theobald

CALDER Working Paper No. 149
January 2016

Abstract

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 and probability of attrition. We find that teachers who student taught in schools with lower teacher turnover are less likely to leave the state's teaching workforce, and 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, these findings suggest that the school context in which student teaching occurs has important implications for the later outcomes of teachers and their students.

I. 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, it is only weakly related to readily quantifiable teacher attributes like licensure status, degree, and experience levels (Aaronson et al., 2007; Goldhaber, 2002; Rivkin et al. 2005). Moreover, the empirical literature about policies designed to affect teacher quality—such as pay for performance (e.g., Fryer et al., 2012; Glazerman and Seifullah, 2010; Goldhaber and Walch, 2012; Springer et al., 2013) and professional development (e.g., Hill and Ball, 2004; Jacob and Lefgren, 2004)—focuses almost exclusively on in-service teachers (and the findings are mixed).

This focus on in-service interventions ignores the reality that the majority of our country's investment in teacher workforce development occurs before teachers enter the workforce. 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 and Sass, 2011), but there is a great deal of speculation that teacher education—and student teaching experiences in particular (Levine, 2006; NCATE, 2010; Wilson et al., 2001)—has a powerful influence on a teacher's later success.

The theory of action connecting student teaching to teacher outcomes 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 the National Council for Accreditation of Teacher

[1] Estimates suggest, for example, that a one standard deviation increase in teacher quality raises student achievement in reading and math between 10–25% of a standard deviation (see Aaronson et al. [2007] and Hanushek and Rivkin [2010] for estimates of the effect size associated with changes in teacher quality). 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 et al., 2005).

Education (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, 2009; Ronfeldt et al., 2014) and characteristics of the school in which student teaching occurs (Ronfeldt 2012, 2015) are predictive of teacher effectiveness and attrition from the profession.

In this paper we investigate the relationship between candidates' student teaching experiences—both where student teaching occurs and which teachers mentor student teachers (cooperating teachers)—and their later teaching effectiveness and probability of attrition. 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 two broad research questions: (1) What student teaching experiences are predictive of value-added estimates of teacher effectiveness?; and (2) What student teaching experiences are predictive of the probability that a teacher leaves the teacher workforce in Washington State?

We find evidence that candidates assigned to cooperating teachers with an advanced degree are less effective once they enter the workforce, while teachers who student taught in schools with less teacher turnover are less likely to leave the state's teaching workforce. We also find 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. These findings are 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 are ultimately descriptive due to the non-random sorting of teachers to student teaching positions and their first teaching jobs. 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 paper proceeds as follows. In Section II, we review the existing empirical evidence linking student teaching experiences to teacher attrition and effectiveness, and then introduce the data set that allows us to build on this prior work in Section III. We present our analytic approach in Section IV, review our findings in Section V, and offer some conclusions in Section VI.

II. Background

Five years ago, a National Research Council (2010) 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.[2] Since then a number of studies have investigated differences in effectiveness associated with acquiring a teaching credential from a specific TEP (Gansle et al., 2012; Goldhaber et al., 2013; Koedel et al., 2015; Lincove et al., 2013; Mihaly et al., 2013). While there is some variation across studies in how much of teacher effectiveness can be attributed to TEPs, these studies show that TEP indicators themselves explain less than 1% of the variation of student achievement (Goldhaber et al., 2013); i.e., the vast majority of the variation in teacher effectiveness is within rather than between TEPs. But an important caveat about these estimates is that they do not isolate the impact of the TEPs themselves from various selection mechanisms (discussed below) associated with prospective teachers' enrollment in training programs, entrance into the workforce, or employment in specific schools and classrooms.[3]

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

[2] At the time there were only two large-scale quantitative studies (Boyd et al., 2006, 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 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.

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

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 et al., 2014b; Krieg et al., 2015) focuses on the relationship between student teaching experiences and workforce entry. Goldhaber et al. (2014b) find that characteristics of individual prospective teachers (such as 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. (2015), 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 et al., 2005) may also operate through student teaching assignments.

Our study builds closely on recent work by Matt Ronfeldt and colleagues that connects student teaching experiences to teacher effectiveness and retention. Ronfeldt (2012) suggests that student teaching experiences may be linked to teacher attrition and effectiveness—especially in underserved student populations. The evidence from this study challenges positions like those of Haberman and Post (1998) and Haberman (1995) who propose that teachers should be trained "in the worst schools and under the poorest conditions of practice" (Haberman, 1995). The findings in Ronfeldt (2012) in fact suggest the opposite: the effectiveness of novice teachers is significantly higher if they student taught in "high-functioning" schools with lower teacher turnover.[4]

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 and are less likely to leave the profession than other teachers. 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. Ronfeldt (2015) also quantifies the "match" between internship schools and first-job

[4] Ronfeldt quantifies the level of teacher turnover based on the "stay ratio," a term we define in the next section.

schools by computing the absolute difference between internship and first-job characteristics, and finds some 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, although this finding does not hold for differences in school racial composition.[5]

We contribute to the sparse literature relating teacher education experiences to the in-service experiences of teachers. In particular, 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 workforce outcomes. Teacher candidates are non-randomly 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 non-randomly assigned to internship experiences (Krieg et al., 2015), so stronger or weaker teacher candidates might be systematically matched with particular internship schools or mentor teachers. Finally, the possibility that selection into particular teaching assignments might influence teacher effectiveness estimates has received a good deal of attention in the literature (Chetty et al., 2014; Kane et al., 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.

III. Data and Summary Statistics

Data

The analytic data set we utilize combines information about teacher candidates and their

[5] Ronfeldt also experiments with other measures of the "match", including variants of Euclidean distance, but also finds little relationship between these measures and teacher effectiveness (Personal Communication, July 2015).

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.[6] TEPs also provided the academic year of the internship, the building and district in which the internship occurred, and the name of the cooperating teacher supervising the internship.[7]

The earliest individuals considered in this study completed their student teaching in 1998, while the most recent student taught in 2010.[8] **Figure 2** shows the frequency of observations by student teaching year as well as the years each TEP provided data for their student teachers. 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).[9]

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 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.[10] We

[6] The longest span provided by a university was every intern between the years 1998 and 2011 and the shortest span was 2006–2011.
[7] Internships in this paper 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.
[8] When representing years, this paper uses the convention of listing the first year of the academic year. Thus, 1994 represents the 1994–1995 academic year.
[9] There are a total of 21 TEPs in Washington (see Goldhaber et al. [2013] for a full list.) Approximately 15% of the state's public school teachers were trained outside the state. See http://program.pesb.wa.gov/reports/reporting_progress/clinicallocation for detailed maps on where Washington teachers tend to do their student teaching.
[10] We combine specific endorsement information into four categories: elementary education, special education, STEM, and other. These data also contain the birth year of each intern, which allows us to calculate the age of each intern during his or her internship year.

merge the above 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 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 the percent of students eligible for free or reduced priced lunch, the percent of under-represented minority students,[11] and, following Ronfeldt (2012), we also calculate the "stay ratio" of each school.[12] The stay ratio is a school-level measure of teacher turnover, which we define in each school year as the percent of the school's non-retirement-age teachers who return to the school in the following year.[13] Therefore, schools with *less* teacher turnover have a *higher* stay ratio.[14]

We use this analytic data set to investigate predictors of teacher attrition. 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

[11] We define under-represented minority as American Indian, Black, or Hispanic. And, we standardize this variable and free and reduced priced lunch (FRL) within year and school level to ensure that these measures are comparable for all interns in the sample.

[12] Ronfeldt (2012) shows that a school's stay ratio is correlated with other survey-based measures of school functionality, such as administrative quality, staff support, student behavior, and teacher safety.

[13] 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).

[14] Though we focus on *school-level* variables in this paper, there are good reasons to believe that *classroom-level* variables may be even more important for teacher attrition and effectiveness. A small subset of interns in the sample can be linked to classroom-level data for both their internship and current teaching position, and estimates from models that consider these measures generally support the notion that focusing on the school level masks important variation at the classroom level. However, these estimates are based on such a small subsample that we do not present them in this paper, although they are available from the authors upon request.

their proctor on the state exam and to both current-year and prior-year test scores.[15] Since 2009, the state's new 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.[16] For our investigation of research question #1, we only consider interns who 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 section IV.

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. Although these "not hired" 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 the three internship school characteristics described in the previous section for hired and not hired interns, and demonstrates that hired interns tend to student teach in schools that have more teacher turnover and fewer FRL students than non-hired teachers.[17] The first result may be due to the significant percentage of interns, almost 15%, who are hired into the very schools in which their internships occur (Goldhaber et al., 2014b), and suggests that the internship experiences of hired interns are *not* representative of all interns from the TEPs in our sample, reinforcing the importance

---

[15]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. However, for the 2009 school year, we are able to check the accuracy of these proctor matches using the state's new Comprehensive Education Data and Research System (CEDARS) that matches students to teachers through a unique course ID. Our proctor match agrees with the student's teacher in the CEDARS system for about 95% of students in math and 94% of students in reading.
[16] CEDARS data includes fields designed to link students to their individual teachers, based on reported schedules. However, limitations of reporting standards and practices across the state may result in ambiguities or inaccuracies around these links.
[17] Note that the correlation between school %FRL and school %URM across the state is quite high (r = 0.77), while the correlation between each of these variables and school stay ratio is lower (r = -.07 and r = -.10, respectively). Because of the high correlation between %FRL and %URM, we estimate models that include both variables and models that include each separately.

of accounting for sample selection.[18]

We now turn our attention to the 5,876 hired interns in the data set. 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 each intern's internship (x-axis) and first job (y-axis) school. The estimated linear relationship is positive, 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. Moreover, 60% of interns find their first job in a school with a higher percentage of FRL students than their internship school (i.e., are above the 45° line in Figure 3).

Table 3 provides summary statistics (calculated at the teacher/year level for hired student teachers) for variables of interest in the analytic models described in the next section. These statistics illustrate how the composition of the sample changes between the mobility models (estimated for all 5,876 hired interns) and effectiveness models (estimated only for the 1,351 interns linked to student achievement data). For instance, a much lower percentage of interns are endorsed in elementary education in the mobility data set (64.5%) than the effectiveness data set (88.6%), which is due to elementary teachers being more likely to be linked to student achievement data. Importantly, the effectiveness data set contains *fewer* novice teachers and *more* experienced teachers than the broader 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.

IV. Analytic Approach

Our objective is to investigate the relationship between interns' student teaching experiences and their effectiveness and attrition. We first define Z_j as a vector of student teaching

[18] See Goldhaber et al. (2014b) for more information on predictors of teacher workforce entry.

experiences for intern j (the specific variables in this vector vary across model specifications). To investigate the relationship between each of these variables and the outcomes above, we must consider the implications of four potential sources of bias discussed in Section II. First, individuals are non-randomly selected into different teacher education programs (Goldhaber, 2014). Second, teacher candidates are non-randomly sorted into different student teaching positions (Krieg et al., 2015; Maier and Youngs, 2009). Third, teacher candidates are non-randomly selected into the public teaching workforce (Goldhaber et al., 2014b). And finally, teachers are non-randomly sorted into different teaching positions (Krieg et al., 2015). Each of these sources of non-random variation could lead to bias if they mean that individuals who have different teacher education, student teaching, or teaching experiences are different in other ways that also influence their effectiveness and probability of staying the workforce.

As we describe below, we attempt to address these potential sources of bias in several ways. We estimate models that include a rich set of variables that are intended to 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 and probability of attrition (Goldhaber and Cowan, 2014; Goldhaber et al., 2013; Krieg et al., 2015). 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 that attempt to correct for sample selection bias that could result from only observing outcomes for teacher candidates who ultimately enter the teaching workforce. However, we do not believe that these models fully account for the non-random 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 non-random sorting.

Analytic Models

For the subset of 1,351 interns who enter the workforce and teach math in a grade and year in which current and prior year test scores are available, we can 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:[19]

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

In (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 3) and current school characteristics (including school fixed effects in some specifications) for teacher *j* in year *t*[20]*;* 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 just as 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 OLS, we cluster the standard errors at the teacher level to account for dependence between observations associated with the same teacher.

[19] We also estimate all models for student reading performance, and these estimates are available from the authors upon request.
[20] Recent evidence (e.g., Chetty et al., 2014) suggests that teacher quality is portable across school settings, so our preferred models do not include school fixed effects. However, we also estimate all models with school fixed effects and report these estimates as a robustness check.

To investigate predictors of teacher attrition for the full sample of 5,876 hired interns, we estimate discrete-time hazard models that predict the log odds that each intern we observe in the state workforce decides to leave the workforce at the end of each year:[21]

$$\log\left(\frac{p_{jkt}}{(1-p_{jkt})}\right) = \beta_0 + \beta_1 S_{kt} + \beta_2 T_{jt} + \beta_3 Z_j + \beta_t + \varepsilon_{jkt}^{\beta} \qquad (2)$$

In (2), p_{jkt} is the probability that intern *j* in school *k* leaves the state workforce at the end of year *t*. The log odds of this probability is modeled as a function of school and district characteristics S_{kt}; T_{jt} is a vector of characteristics for teacher *j* in year *t*; Z_j is the vector of preservice teacher education experiences for intern *j*; and β_t is a fixed effect for year *t*. The estimated coefficients in the vector β_3 can be interpreted as the expected increase in log odds of teacher attrition associated with changes in each student teaching experience, holding all other covariates constant. As in the effectiveness models, we cluster the standard errors at the teacher level to account for dependence between observations associated with the same teacher.

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., percent 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 $|C_{jt} - I_j|$ between the 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 attrition or effectiveness:

$$\gamma_1 C_{jt} + \sum_{k=1}^{3} \gamma_{k+1}(C_{jt} - I_j)^k + C_{jt} \sum_{k=1}^{3} \gamma_{k+4}(C_{jt} - I_j)^k \qquad (3)$$

[21] We also estimate equation 2 as a probit model to facilitate a comparison with the Heckit model estimates described later in this section.

The first term in equation (3) 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 non-linear relationship with the outcomes in equations 1 and 2, while the third allows this relationship to vary depending on the characteristic of the intern's current school. When we estimate "match" models, we just add the polynomial in equation 3 as additional predictors in the models in equations 1 and 2.

Sample Selection Correction

As we describe above, we can use data on interns who do not enter the workforce to account for potential sample selection bias associated with participation in the teacher labor market. "Heckit" models that account for sample selection (e.g., Heckman, 1979) require instrumental variables (IVs) that are predictive of the probability that interns enter the workforce, but are otherwise unrelated with teacher effectiveness and attrition. We utilize two IVs: 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 and 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 predictive of workforce entry, all else equal (both separately and jointly).[22] These estimates come from a first-stage probit regression:

$$\Pr(O_j = 1) = \Phi(\nu_0 + \nu_1 IV_j + \nu_2 T_j + \nu_3 I_j + \varepsilon_j) \qquad (4)$$

[22] 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.

In (4), $O_j = 1$ if intern *j* is observed in the public 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 also 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 teacher 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.[23]

Under the exclusion restriction, we can use the first-stage coefficients estimated from equation (4) to form a selection correction term that can be included an additional covariate in a Heckit model that accounts for non-random sample selection (Vella, 1998):[24]

$$\pi_j = \frac{\varphi(v_0 + v_1 I_j + v_2 T_j + v_3 I_j + \varepsilon_j)}{\Phi(v_0 + v_1 I_j + v_2 T_j + v_3 I_j + \varepsilon_j)} \quad (5)$$

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 et al. (2012) and Goldhaber et al. (2014a).[25]

V. Results

We discuss the estimates from the model in equation 1 in the first subsection, and the estimates from the model in equation 2 in the second subsection. Within each section, we first discuss specifications of these models that only consider internship school characteristics (as in Ronfeldt, 2012), then consider additional specifications that consider characteristics of the

[23] 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 non-random sorting along unobserved dimensions.
[24] We estimate equation 2 as a probit model in these specifications.
[25] 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.

cooperating teacher, and conclude with specifications that include measures of the "match" between each intern's student teaching and current school.

Effectiveness Models

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 above, earlier work (Goldhaber et al., 2013) finds that TEP indicators 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 TEPs 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*.[26] 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.[27]

We now turn to the coefficient estimates from different specifications of equation 1. **Table 4** presents estimates from specifications of the model in equation 1 that replicate the specifications

[26] Note that the student teaching variables include measures of the "match" described later in this section.
[27] 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.

reported in Ronfeldt (2012). We focus on the estimates associated with the three internship school characteristics discussed in Section III, 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).[28] These findings differ from those of Ronfeldt (2012), 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 5**, we extend the specifications from Table 4. In all columns, we include an indicator for whether the teacher is employed in the same school where they student taught,[29] 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 et al., 2014b), they are missing an opportunity to use this process to identify and hire more effective teachers.[30]

In the even columns of Table 5, we consider characteristics of the intern's cooperating teacher (including measures of the gender and racial "match" between the cooperating teacher and intern). The only consistent finding from these specifications is that interns whose cooperating

[28] 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 %URM and internship school %FRL are both highly predictive of teacher effectiveness across all specifications. This previews the "match" findings discussed later in the section.
[29] We also experiment with models that consider hiring into the same district and find similar results.
[30] This still leaves the possibility that schools are using the *selection* of student teachers as a screening process in hiring.

teacher held an advanced degree are less effective than other teachers, all else equal.[31] Given that these specifications also control for cooperating teacher experience, it is difficult to know what to make of this finding. One plausible explanation is that the cooperating teachers with advanced degrees are systematically different from cooperating teachers without advanced degrees. For instance, while there is little empirical evidence that advanced degrees are predictive of teacher effectiveness in general (Goldhaber and Brewer, 1997; Monk and King, 1994), this may not be the perception amongst 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.

In **Table 6,** 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 3 is added to the model in equation 1). We first let I_j and C_{jt} be the %FRL of the teacher's internship school and current school (respectively), and then estimate similar models that consider %URM as the school measure. From column 1, we can see that, all else equal, a one standard deviation increase in current school %FRL is associated with a decrease in student performance of about .025 standard deviations.

The coefficients of interest, though, are for the terms that consider the "match" between a teacher's internship and current school. Because these estimates are difficult to interpret, we use the estimates from column 1 to calculate average predicted test scores at each combination of internship school %FRL and current school %FRL and plot the resulting estimates in the contour plot in **Figure 5**.[32] In the contour plot, red regions indicate higher test scores (where a "+" indicates a region in which the average predicted test score is significantly greater than zero), and blue regions indicate lower test scores (where a "-" indicates a region in which the average predicted test score is

[31] This estimate is also negative and statistically significant in models that consider student reading performance. Moreover, we find that the effect is strongest for more experienced cooperating teachers but still statistically significant and negative for a cooperating teacher of average experience, and is statistically significant and negative for interns from three of the six TEPs in our sample.

[32] 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 with all variables other than internship school %FRL and current school %FRL kept at their observed values.

significantly less than zero).[33] Importantly, because our goal here is to assess the importance of the *match* between internship and current school, we do *not* use the coefficient on current school %FRL 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 6 suggest that the student's test scores will vary as a function of the %FRL of his or her teacher's internship school.[34]

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 in a school with a very different %FRL).[35] 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 statistically significant and positive regions of Figure 5 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 (2 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 (2 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

[33] These hypothesis tests are performed using the "margins" command in STATA, and simply test the average predicted test score at each combination of internship school characteristic and current school characteristic against zero.

[34] 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 5) is incorporated into these estimates.

[35] These contour plots look quite similar when we use non-standardized (i.e., 0%-100%) measures of school %FRL.

teachers are in a considerably *more* disadvantaged school than where they student taught.

Figure 6 plots average predicted scores from column 2 of Table 6, 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 (3 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 (2 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 above, the relationships in Figures 5 and 6 (and the estimates in Table 6) 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 subsection on robustness checks, we do not find evidence that teachers with better observable *preservice* 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 non-random 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.

Mobility Models

We now turn to predictors of teacher attrition (recall that we can estimate these models for the full "mobility sample" summarized in Table 3, not just the subsample who are linked to student test scores). **Tables 7–9** are direct parallels of Tables 4-6, except they present estimates from the specifications of the discrete-time hazard model in equation 2. The primary conclusion from the estimates in Table 7 is that, like Ronfeldt (2012), we find strong evidence that interns who student taught in schools with less teacher turnover (i.e., a higher stay ratio) are less likely to leave the state's public teaching workforce. This is true even when we include all three internship school measures in the same model (column 4), and is robust to the inclusion of school fixed effects.

In **Table 8**, we find that individuals who are teaching in the same school where they student taught are no less likely to leave the teaching workforce, all else equal. We also find little evidence that any cooperating teacher characteristics are predictive of the probability that an intern leaves the teaching workforce. Finally, in **Table 9**, we present estimated coefficients from the polynomial of the difference between various internship school and current school characteristics (see equation 3) as predictors of teacher attrition. The coefficients of interest are for the difference terms and the difference terms interacted with current school characteristics, and of these 12 estimates, only one is statistically significant at the .10 level. We therefore avoid interpreting these estimates more broadly and instead conclude that there is little evidence that these measures of the "match" between internship and current school are predictive of teacher attrition.

Sample Selection Correction and Robustness Checks

As we discuss at the outset of Section IV, there are several potential sources of bias for the estimates from Tables 4–9. We account for one of these sources of bias (non-random selection into the teaching workforce) directly by estimating sample selection (or Heckit) models that rely on data about teacher candidates who never enter the public teaching workforce. **Table 10** shows that the two IVs we have identified for the first stage of these models are predictive of workforce entry, both separately and jointly, and in the expected direction (i.e., teacher candidates are more likely to enter the workforce, 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). Therefore, we use these IVs to produce Heckit estimates presented in the even columns of **Table 11** (for the effectiveness models) and **Table 12** (for the mobility models).

The broad conclusion from these tables is that the sample selection correction makes very little difference in our estimates of the relationship between student teaching experiences and student math performance. Specifically, in Table 11 the negative and statistically-significant relationship between cooperating teacher advanced degree and student math achievement is nearly identical in the OLS and Heckit models.[36] Likewise, in Table 12 the relationship between internship school stay ratio and attrition is *even more* negative and statistically significant when we account for sample selection, although the difference between the estimates in the probit and Heckit models is not statistically significant. We view this as additional support to the overall conclusion from Ronfeldt (2012) that interns from higher-functioning schools (as proxied by the level of teacher turnover) are better prepared for the challenges of teaching and thus less likely to leave the workforce.

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 above, this could be a spurious correlation that is due to non-random 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 6, except at the teacher-year level and with teacher credential test scores as the outcome variable.[37] Credential test scores are an imperfect measure of teacher quality, but empirical evidence shows that they are modestly

[36] OLS and Heckit estimates from the match models (Table 6) are also very similar. Results are available from the authors upon request.
[37] We omit all student-level variables in these models.

predictive of future teacher effectiveness (Clotfelter et al., 2010; Goldhaber, 2007). 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 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), none of the measures of the match from equation 3 are statistically significant predictors of credential test scores (for either %URM or %FRL). While this suggests that there is not dramatic non-random sorting along observable dimensions, it is of course still possible that there is non-random sorting along unobserved dimensions of more effective teachers to schools that are a good "match" with their student teaching experience.[38]

Finally, the match results discussed in Section IV 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 characteristic of the intern's current and internship schools and report the estimates in **Table 13**. The first column of Table 13 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.[39] 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 13 show that, as hypothesized, the match effect is almost

[38] We also find no evidence that teachers with better credential test scores are any more or less likely to be assigned to a cooperating teacher with an advanced degree, which does little to cast doubt on the estimated negative relationship between cooperating teacher advanced degree and future teacher effectiveness.

[39] 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.

three times larger for novice teachers (with two or fewer years of experience) than it is more experienced teachers.[40] 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 13), similar patterns emerge; the relationship between the match and teacher effectiveness is stronger for novice teachers than more experienced teachers.[41] Thus the estimates in Table 6 (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.

A final question is whether the benefit of a good "match" is unique to student teaching or could also be observed in the broader teacher labor market (e.g., a second-year teacher in a disadvantaged school may be more effective if he or she has already taught for one year in a similar school). There is little empirical evidence on this topic, but early estimates from Washington State (available from the authors on request) suggest that, among teachers who switch schools after their first year of teaching, the relationship between the "match" between their second school and their first school and teacher effectiveness is very similar to the relationships in Table 13 (i.e., teachers who switch to a similar school are more effective than teachers who switch to a very different school). This early finding is also subject to potential bias—teachers who switch to a very different schooling environment after one year of teaching may be less effective than other teachers for unobserved reasons—but suggests that the story of these results may *not* be just about student teaching. Instead, the primary "match" results in this section may simply illustrate the importance of prior experience in a similar schooling environment (in this case, through student teaching).

VI. Conclusions

This paper contributes to a small but growing empirical evidence base about where future

[40] This difference is statistically significant at the .10-level.
[41] 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 0.300 (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.

teachers should do their student teaching. We replicate the finding from Ronfeldt (2012) that relatively low levels of teacher turnover in internship schools are associated with a reduced likelihood of attrition for teacher candidates who enter the workforce. But 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. Although these results could reflect non-random 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, they do suggest that TEPs should consider placing teacher candidates into student teaching schools that look like the schools they are likely to be hired into.

Our findings 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 *more* advantaged setting than their current school, so students in disadvantaged schools are less likely to have a teacher whose student teaching "matches" their school setting than students in advantaged schools. This suggests that TEPs that are committed to educating teachers who will be successful in disadvantaged schools need to be placing more student teachers in these settings.[42]

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

[42] This comports with our policy recommendation from Krieg et al. (2015) that placing better (or just more) student teachers in disadvantaged schools could have important impacts on the distribution of new teacher quality.

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.

References

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

Boyd, D., Lankford, H., Loeb, S., and 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.

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.

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. National Center for Education Information.

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

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

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. (2014). What do value-added measures of teacher preparation programs tell us? Carnegie Knowledge Brief 12.

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., & Cowan, J. (2014). Excavating the teacher pipeline: Teacher preparation programs and teacher attrition. *Journal of Teacher Education*, 65(5), 449–462.

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

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

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

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.

Greenberg, J., Walsh, K, & McKee, A. (2014). 2014 Teacher prep review: A review of the nation’s teacher preparation programs. National Council on Teacher Quality.

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. (1986). The economics of schooling - production and efficiency in public-schools. *Journal of Economic Literature* 24: 1141–1177.

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

Krieg, J., Theobald, R., & Goldhaber, D. (2015). A foot in the door: Exploring the role of student teaching assignments in teachers' initial job placements. CEDR Working Paper 2015-2. Seattle, WA.

Levine, A. (2006). Educating school teachers. Report for *The Education Schools Project*, September 2006.

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.

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*. 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, Washington D.C.

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

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

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

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

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

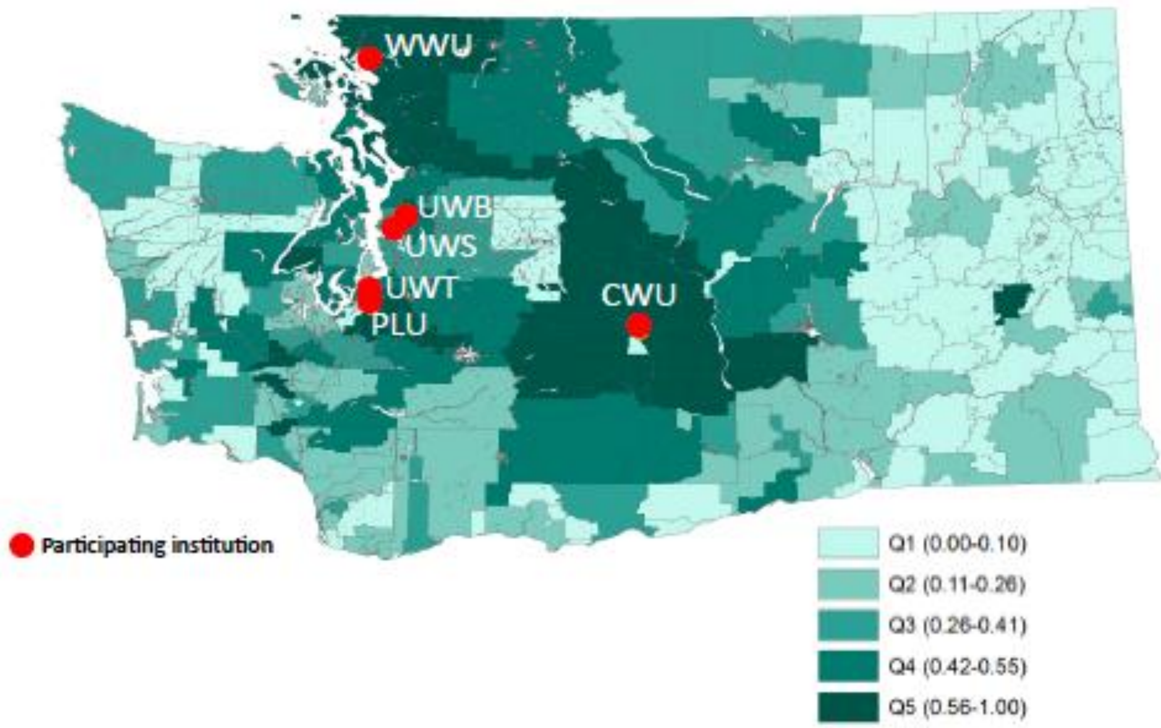
Vella, F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources*, 33(1), 127–169.

Wilson, S., Floden, R., & Ferrini-Mundy, J. (2001). Teacher preparation research: Current knowledge, gaps, and recommendations. Center for the Study of Teaching and Policy, University of Washington. Retrieved November 20, 2010, 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.

Figures

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 ten years who graduated from one of the six participating institutions in this study. The legend shows how these proportions are binned into five quintiles.

Figure 2. Student Teaching Assignments by Year of Student Teaching

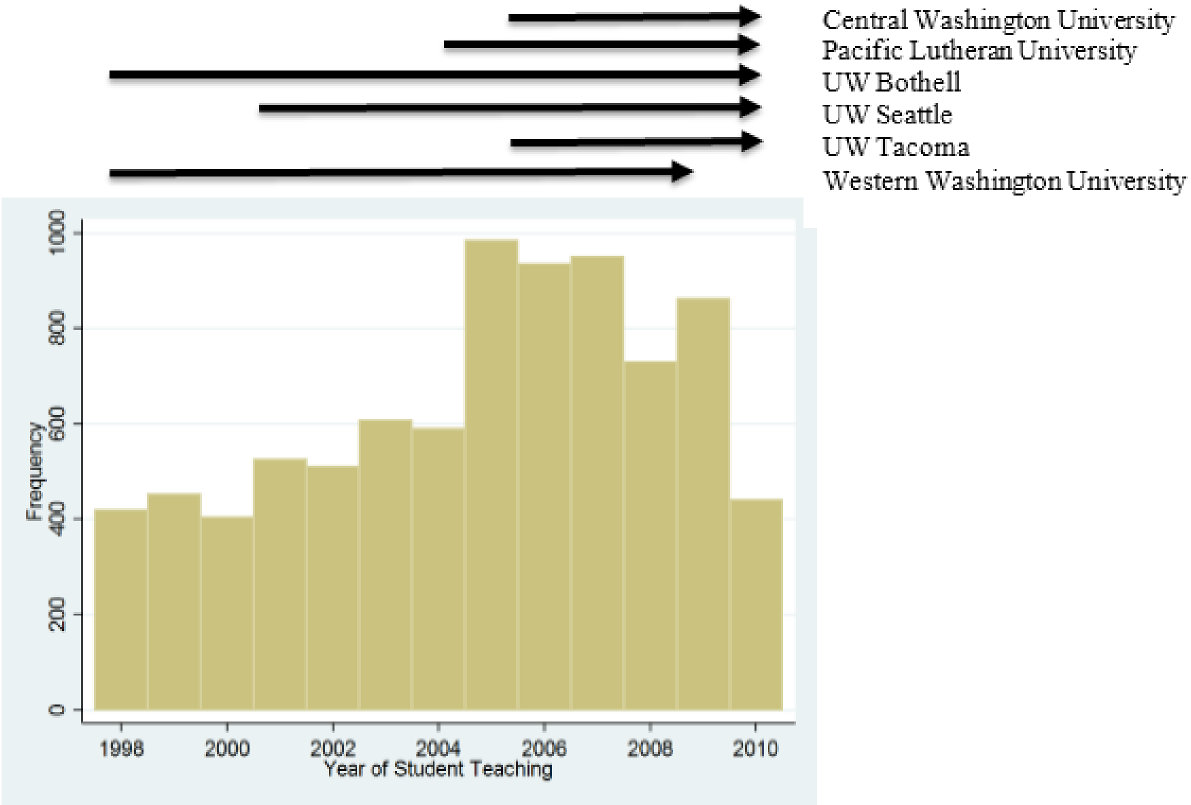


Figure 3. Standardized Percent FRL at Internship and First Job School

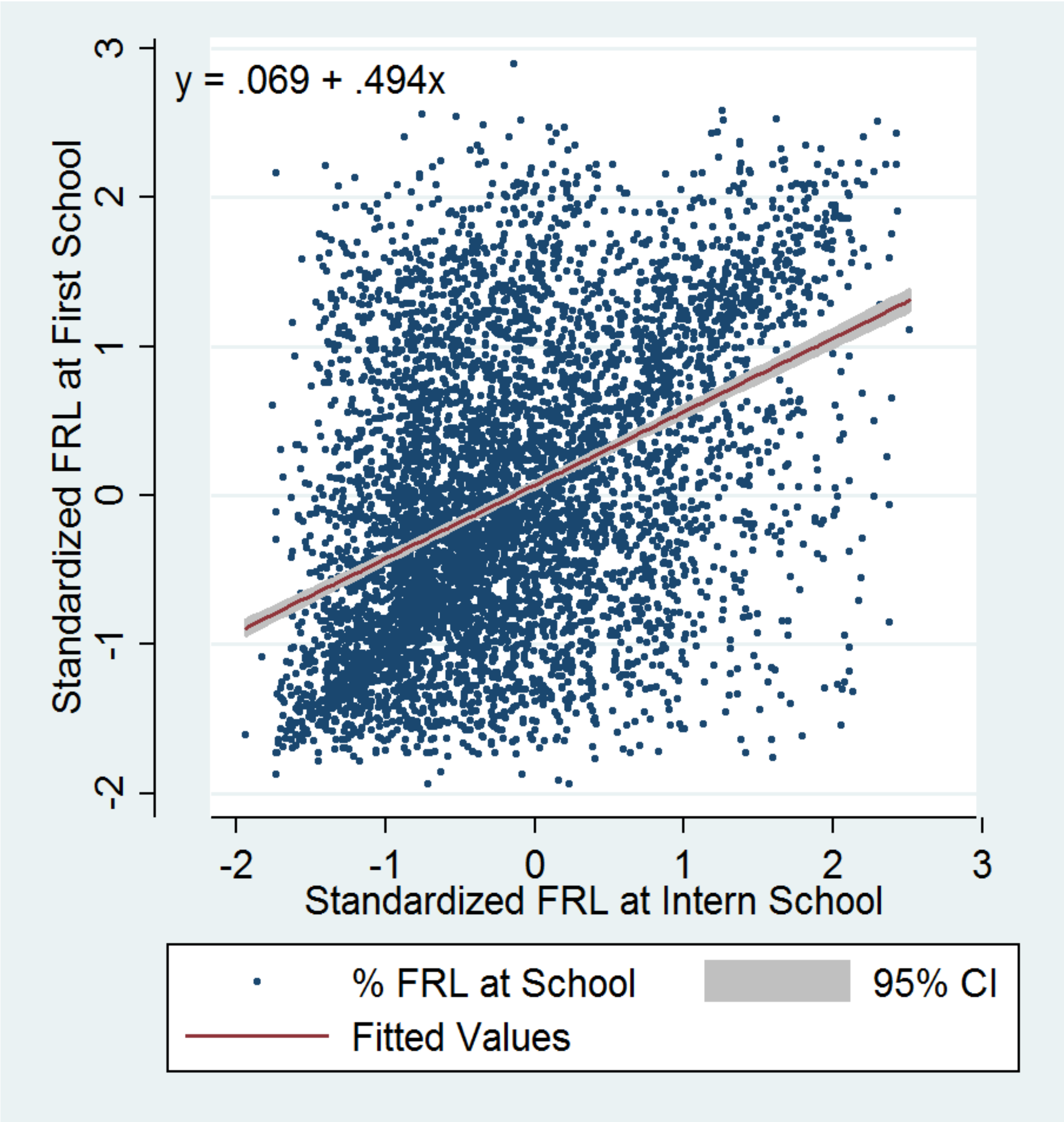


Figure 4. Summary of ANOVA of Student Test Performance

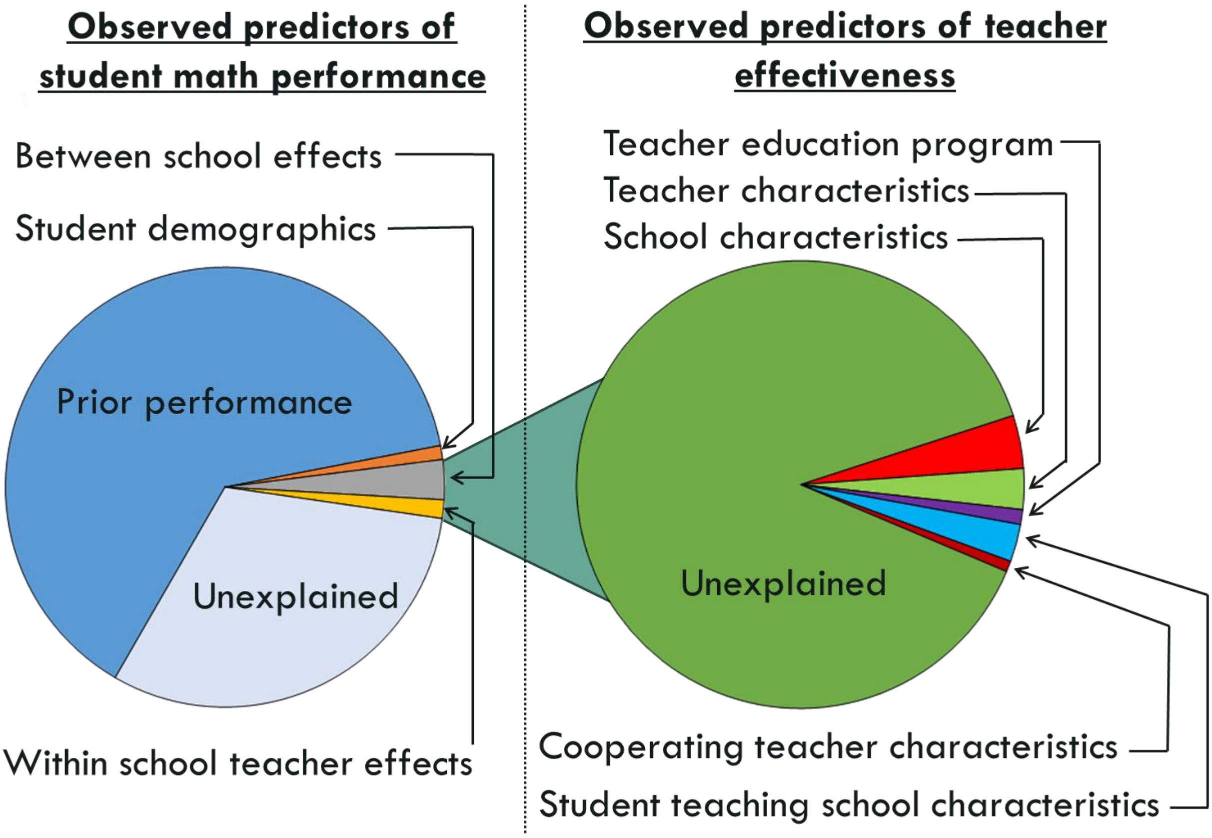
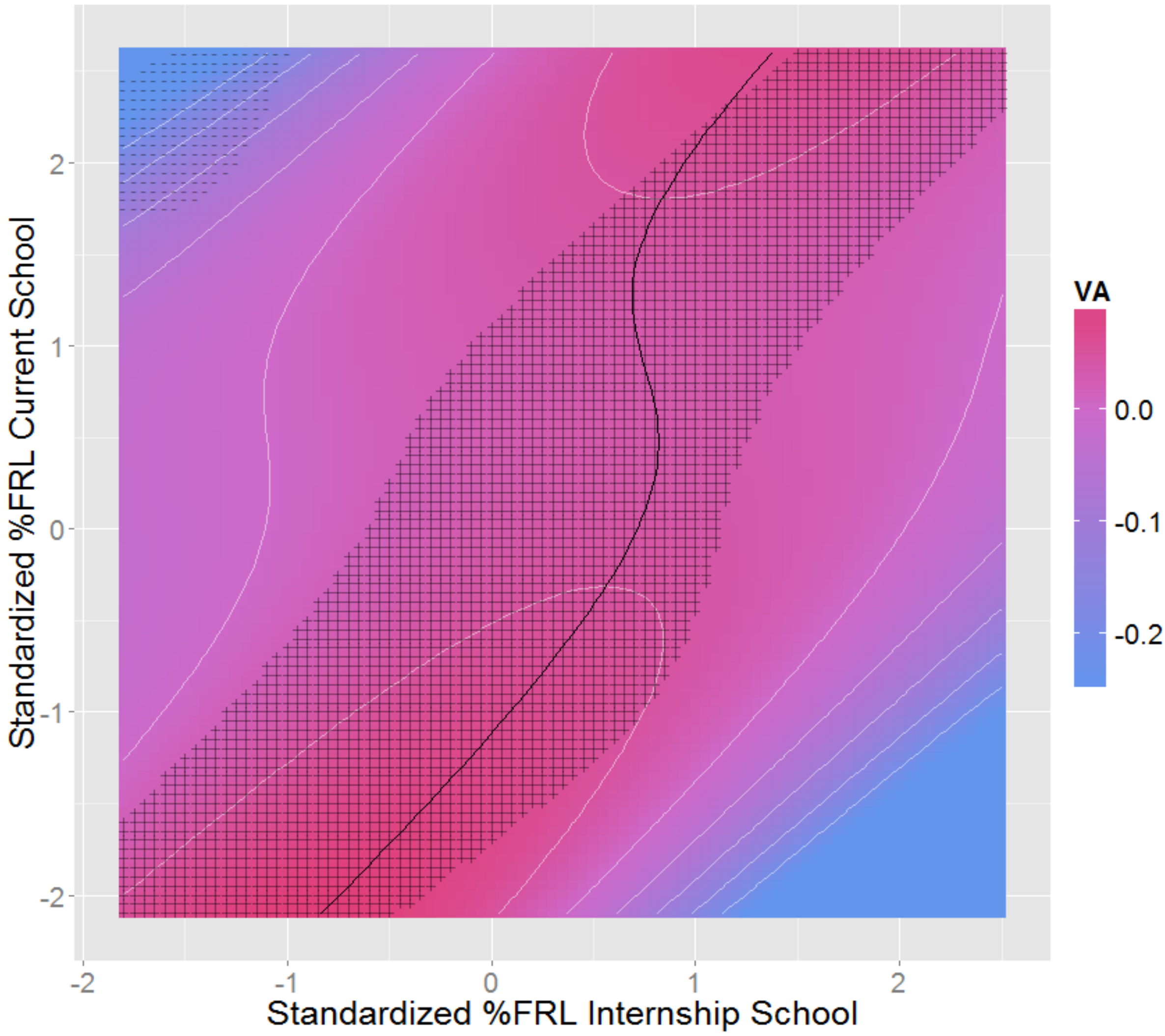
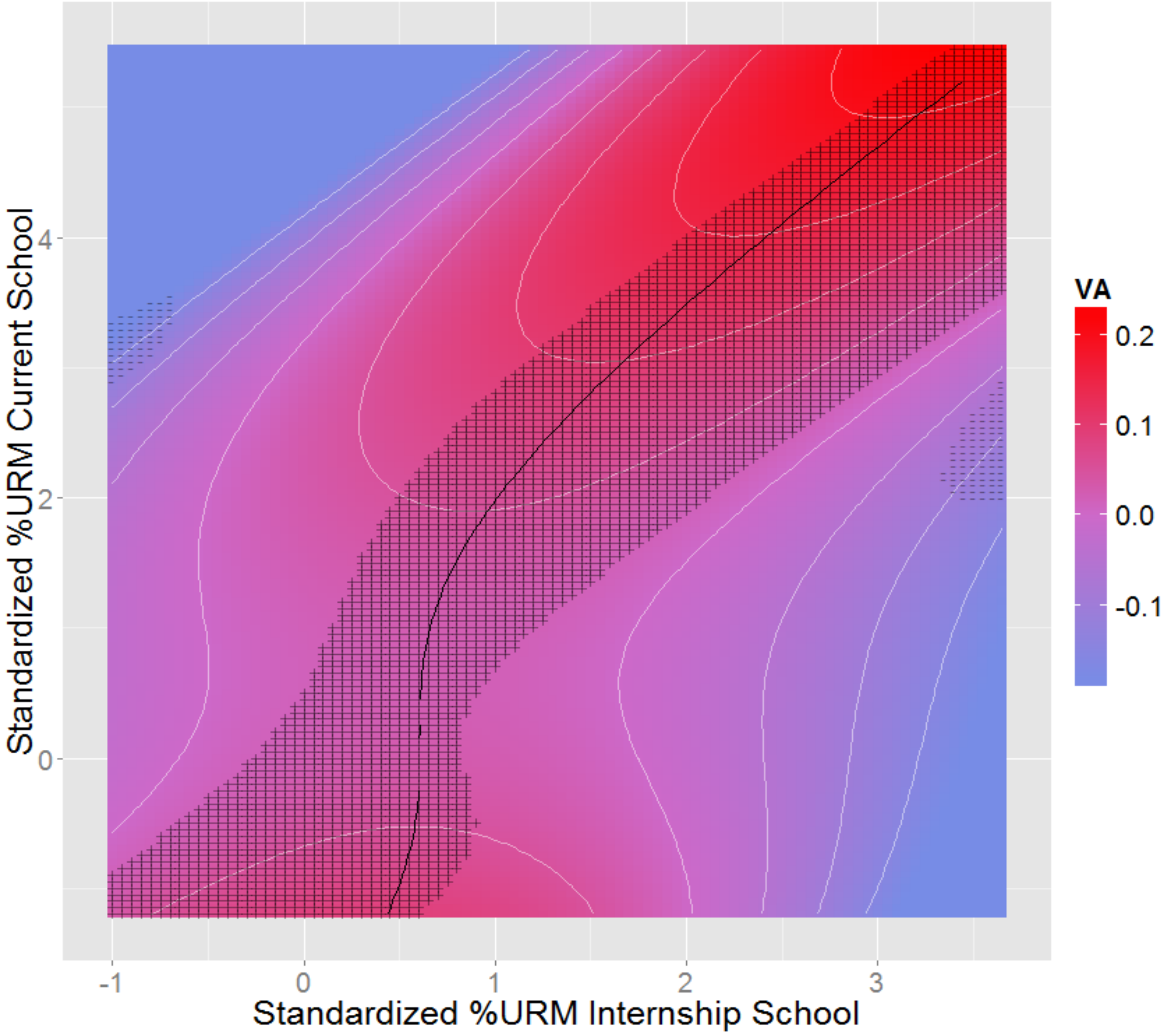


Figure 5. Predicted Math Performance by Internship and Current School %FRL



NOTE: "+" denotes region that is significantly greater than zero ($p < .05$), "-" denotes region that is significantly less than zero

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

Tables

Table 1. Summary Statistics for Hired and Not Hired Teacher Candidates

Sample	**Not Hired**	**Hired**
Internship school stay ratio	-0.144 (0.659)	-0.200*** (0.621)
Internship school %URM	-0.016 (0.782)	-0.028 (0.798)
Internship school %FRL	-0.110 (0.830)	-0.168** (0.855)
Number of Unique Individuals	2,393	5,876

NOTE: *p* values from two-sided *t* test of difference between columns: **p* < .05; ***p* < .01; ****p* < .001

Table 2. Summary Statistics for Internship and First Job School Characteristics

Sample	**All Hired Teachers**		
	Internship School	**First Job School**	**Difference**
Standardized stay ratio	-0.200 (0.621)	-0.350 (0.611)	-0.150*** (0.780)
Standardized %URM	-0.028 (0.798)	0.139 (0.995)	0.166*** (0.930)
Standardized %FRL	-0.168 (0.855)	-0.014 (0.974)	0.154*** (0.979)
Number of Unique Teachers	5,876	5,876	5,876

NOTE: p values from two-sided t test: *p < .05; **p < .01; ***p < .001

Table 3. Teacher/Year-Level Summary Statistics

Models	**Mobility**	**Effectiveness**
Age in internship year	28.293 (7.783)	29.087 (8.240)
Time from internship to first job	1.501 (1.098)	1.430 (0.909)
Male	0.250	0.233
Non-white	0.088	0.093
STEM endorsement	0.130	0.168
SPED endorsement	0.138	0.092
Elementary endorsement	0.645	0.869
No Teaching Experience	0.125	0.102
One Year of Experience	0.151	0.128
Two Years of Experience	0.139	0.128
Three Years of Experience	0.124	0.122
Four Years of Experience	0.106	0.111
Five or More Years of Experience	0.355	0.409
Internship school stay ratio	-0.204 (0.605)	-0.196 (0.603)
Internship school %URM	-0.099 (0.745)	-0.101 (0.764)
Internship school %FRL	-0.236 (0.824)	-0.237 (0.862)
Same school as internship	0.135	0.126
Cooperating Teacher Experience	15.046 (8.500)	14.862 (8.310)
Cooperating Teacher Advanced Degree	0.597	0.605
Cooperating Teacher Non-white	0.094	0.083
Cooperating Teacher Male	0.232	0.182
Cooperating Teacher Prior Interns	0.344 (0.865)	0.365 (0.870)
Number of Unique Teachers	5,876	1,349
Number of Teacher-Year Observations	33,303	3,632

Table 4. Math Effectiveness Models (Value-Added Model of Student Math Performance)

Internship school stay ratio	0.001 (0.010)			0.003 (0.010)	-0.005 (0.012)			-0.001 (0.012)
Internship school %URM		0.007 (0.010)		-0.009 (0.017)		0.020* (0.010)		0.002 (0.019)
Internship school %FRL			0.011 (0.010)	0.017 (0.016)			0.020** (0.009)	0.019 (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: p values from two-sided t test: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. 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 clustered at the teacher level.

Table 5. Math Effectiveness Model Extensions

Internship school stay ratio	0.003 (0.010)	0.003 (0.010)	0.001 (0.013)	-0.000 (0.012)
Internship school %URM	-0.010 (0.017)	-0.009 (0.017)	-0.001 (0.019)	-0.003 (0.019)
Internship school %FRL	0.018 (0.016)	0.019 (0.016)	0.020 (0.016)	0.021 (0.016)
Same school as internship	0.027 (0.021)	0.022 (0.022)	0.015 (0.021)	0.019 (0.021)
Student Teacher Male	0.014 (0.014)	0.004 (0.016)	0.031** (0.015)	0.042** (0.017)
Student Teacher Non-White	-0.032 (0.020)	-0.034 (0.021)	-0.001 (0.021)	0.002 (0.022)
Cooperating Teacher Experience		0.000 (0.001)		-0.000 (0.001)
Cooperating Teacher Advanced Degree		-0.029** (0.013)		-0.030** (0.012)
Cooperating Teacher Male		-0.026 (0.019)		0.035* (0.021)
Cooperating Teacher Non-White		0.001 (0.029)		0.034 (0.026)
Cooperating Teacher Prior Interns		0.012 (0.007)		-0.000 (0.007)
Cooperating Teacher Male * Intern Male		0.033 (0.033)		-0.055 (0.036)
Cooperating Teacher Non-White * Intern Non-White		0.046 (0.065)		-0.029 (0.057)
Current school controls	X	X		
Current school fixed effects			X	X
N	112,985	112,985	112,985	112,985

NOTE: *p* values from two-sided *t* test: **p* < .10; ***p* < .05; ****p* < .01. 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 clustered at the teacher level.

Table 6. Math Effectiveness Match Models

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

NOTE: *p* values from two-sided *t* test: **p* < .10; ***p* < .05; ****p* < .01. 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 clustered at the teacher level.

Table 7. Mobility Models (Discrete Time Hazard Model of Leaving State for the First Time)

Internship school stay ratio	-0.097** (0.044)			-0.093** (0.044)	-0.107** (0.050)			-0.102** (0.051)
Internship school %URM		0.032 (0.038)		0.047 (0.063)		0.042 (0.044)		0.079 (0.075)
Internship school %FRL			0.013 (0.035)	-0.031 (0.058)			0.011 (0.040)	-0.058 (0.068)
Current school controls	X	X	X	X				
Current school fixed effects					X	X	X	X
N	32,923	32,923	32,923	32,923	25,519	25,519	25,519	25,519

NOTE: *p* values from two-sided *t* test: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. All models include teacher internship year, institution, student teaching quarter, experience, endorsements, time to hire, gender, age, minority status, school enrollment, school urbanicity, school level, current school %FRL, %URM, and stay ratio controls. Standard errors are clustered at the teacher level.

Table 8. Mobility Model Extensions

Internship school stay ratio	-0.092** (0.044)	-0.092** (0.044)	-0.112** (0.051)	-0.108** (0.052)
Internship school %URM	0.046 (0.063)	0.047 (0.064)	0.076 (0.075)	0.072 (0.075)
Internship school %FRL	-0.031 (0.058)	-0.032 (0.059)	-0.060 (0.068)	-0.064 (0.068)
Same school as internship	0.021 (0.071)	0.022 (0.072)	-0.030 (0.082)	-0.031 (0.082)
Student Teacher Male	-0.469*** (0.065)	-0.508*** (0.084)	-0.513*** (0.072)	-0.576*** (0.094)
Student Teacher Non-White	-0.144 (0.091)	-0.140 (0.097)	-0.148 (0.102)	-0.078 (0.110)
Cooperating Teacher Experience		-0.001 (0.003)		-0.001 (0.003)
Cooperating Teacher Advanced Degree		0.005 (0.050)		-0.051 (0.058)
Cooperating Teacher Male		0.009 (0.075)		0.046 (0.085)
Cooperating Teacher Non-White		-0.007 (0.089)		-0.034 (0.103)
Cooperating Teacher Prior Interns		0.005 (0.031)		-0.046 (0.033)
Cooperating Teacher Male * Intern Male		0.083 (0.131)		0.128 (0.145)
Cooperating Teacher Non-White * Intern Non-White		-0.010 (0.252)		-0.370 (0.286)
Current school controls	X	X		
Current school fixed effects			X	X
N	32,923	32,923	25,519	25,519

NOTE: p values from two-sided t test: *$p < .10$; **$p < .05$; ***$p < .01$. All models include teacher internship year, institution, student teaching quarter, experience, endorsements, time to hire, gender, age, minority status, school enrollment, school urbanicity, school level, current school %FRL, %URM, and stay ratio controls. Standard errors are clustered at the teacher level.

Table 9. Mobility Match Models

School measure:	% FRL	% URM
Current school	-0.024 (0.042)	-0.007 (0.046)
Difference	-0.001 (0.048)	-0.068 (0.053)
Difference^2	0.038 (0.035)	0.044 (0.035)
Difference^3	0.006 (0.017)	0.006 (0.015)
Current school * Difference	-0.057 (0.048)	-0.047 (0.052)
Current school * Difference^2	-0.019 (0.031)	0.039 (0.033)
Current school * Difference^3	0.008 (0.008)	-0.010* (0.006)
Current school controls	X	X
N	32,933	33,012

NOTE: *p* values from two-sided *t* test: **p* < .10; ***p* < .05; ****p* < .01. All models include 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 clustered at the teacher level.

Table 10. First-Stage IVs (Probit Predicting Workforce Entry)

Number of new teachers hired by internship school year after internship	0.040*** (0.012)		0.040*** (0.012)
Internship principal from same institution		0.069* (0.036)	0.069* (0.036)
Number of Unique Individuals	8,269	8,269	8,269

NOTE: *p* values from two-sided *t* test: *p < 0.10, **p < 0.05, ***p < 0.01. 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.

Table 11. Comparison of OLS and Heckit Effectiveness Estimates

	OLS	Heckit	OLS	Heckit
Internship school stay ratio	0.003 (0.010)	.004 (.011)	0.003 (0.010)	.003 (.011)
Internship school %URM	-0.010 (0.017)	-.005 (.019)	-0.009 (0.017)	-.004 (.019)
Internship school %FRL	0.018 (0.016)	.012 (.017)	0.019 (0.016)	.014 (.017)
Same school as internship	0.027 (0.021)	.029 (.022)	0.022 (0.022)	.025 (.023)
Student Teacher Male	0.014 (0.014)	.029 (.021)	0.004 (0.016)	.017 (.022)
Student Teacher Non-White	-0.032 (0.020)	-.032 (.022)	-0.034 (0.021)	-.034 (.023)
Cooperating Teacher Experience			0.000 (0.001)	0 (.001)
Cooperating Teacher Advanced Degree			-0.029** (0.013)	-.028** (.014)
Cooperating Teacher Male			-0.026 (0.019)	-.023 (.02)
Cooperating Teacher Non-White			0.001 (0.029)	0 (.03)
Cooperating Teacher Prior Interns			0.012 (0.007)	.012 (.007)
Cooperating Teacher Male * Intern Male			0.033 (0.033)	.03 (.034)
Cooperating Teacher Non-White * Intern Non-White			0.046 (0.065)	.048 (.069)
Current school controls	X	X	X	X
N	112,985	112,985	112,985	112,985

NOTE: *p* values from two-sided *t* test (for models without sample selection correction) or permutation test (for models with sample selection correction): *p < .10; **p < .05; ***p < .01. 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 clustered at the teacher level in the models without sample selection correction, and standard errors are calculated from 1,000 bootstrapped estimates in the models with sample selection correction.

Table 12. Comparison of Probit and Heckit Mobility Estimates

	Probit	**Heckit**	**Probit**	**Heckit**
Internship school stay ratio	-0.044** (0.021)	-.064*** (.025)	-0.044** (0.021)	-.064*** (.025)
Internship school %URM	0.018 (0.030)	.059 (.041)	0.019 (0.031)	.06 (.041)
Internship school %FRL	-0.013 (0.028)	-.047 (.036)	-0.013 (0.028)	-.048 (.037)
Same school as internship	0.005 (0.034)	.017 (.034)	0.005 (0.034)	.017 (.034)
Student Teacher Male	-0.227*** (0.030)	-.211*** (.036)	-0.247*** (0.039)	-.23*** (.043)
Student Teacher Non-White	-0.069 (0.043)	-.088* (.052)	-0.070 (0.046)	-.089 (.054)
Cooperating Teacher Experience			-0.001 (0.001)	-.001 (.001)
Cooperating Teacher Advanced Degree			0.003 (0.024)	.004 (.025)
Cooperating Teacher Male			0.003 (0.037)	.006 (.038)
Cooperating Teacher Non-White			-0.006 (0.043)	-.004 (.044)
Cooperating Teacher Prior Interns			0.001 (0.014)	.002 (.015)
Cooperating Teacher Male * Intern Male			0.044 (0.062)	.04 (.064)
Cooperating Teacher Non-White * Intern Non-White			0.013 (0.120)	.015 (.125)
Current school controls	X	X	X	X
N	32,923	32,923	32,923	32,923

NOTE: *p* values from two-sided *t* test (for models without sample selection correction) or permutation test (for models with sample selection correction): **p* < .10; ***p* < .05; ****p* < .01. All models include 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 clustered at the teacher level in the models without sample selection correction, and standard errors are calculated from 1000 bootstrapped estimates in the models with sample selection correction.

Table 13. Math Effectiveness Absolute Difference Models

School measure:	% FRL			% URM		
Sample:	All	1st-2nd yr	3+ yrs	All	1st-2nd yr	3+ yrs
Absolute difference	-0.019* (0.010)	-0.044** (0.018)	-0.016 (0.011)	-0.017 (0.012)	-0.045** (0.018)	-0.015 (0.013)
Current school controls	X	X	X	X	X	X

NOTE: p values from two-sided t test: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$. 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 clustered at the teacher level.