



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



Exploring the Impact of Student Teaching Apprenticeships on Student Achievement and Mentor Teachers

Dan Goldhaber
John Krieg
Roddy Theobald

Exploring the Impact of Student Teaching Apprenticeships on Student Achievement and Mentor Teachers

Dan Goldhaber
American Institutes for Research/ CALDER
University of Washington

John Krieg
Western Washington University

Roddy Theobald
American Institutes for Research/ CALDER

Contents

Contents.....	i
Acknowledgment.....	ii
Abstract.....	iii
1. Introduction.....	1
2. Background.....	3
3. Data and Summary Statistics.....	6
4. Analytic Approach.....	10
5. Results.....	15
6. Discussion and Conclusions.....	19
References.....	21
Tables & Figures.....	24

Acknowledgments

The research presented here would not have been possible without the administrative data provided by the Washington Office of Superintendent of Public Instruction through data-sharing agreement 2015DE-030 or without the student teaching data provided by TEPs from the following institutions participating in the Teacher Education Learning Collaborative (TELC): Central Washington University (CWU), City University, Evergreen State College, Gonzaga University, Northwest University, Pacific Lutheran University, St. Martin’s University, Seattle Pacific University, Seattle University, University of Washington Bothell, University of Washington Seattle, University of Washington Tacoma, Washington State University, Western Governors University, and Western Washington University. The research presented here utilizes confidential data from CWU. The views expressed here are those of the authors and do not necessarily represent those of CWU or other data contributors. Any errors are attributable to the authors.

This research was funded by the Bill and Melinda Gates Foundation (grant #OPP1128040) and an anonymous foundation, and was also supported by the National Center for the Analysis of Longitudinal Data in Education Research (CALDER), which is funded by a consortium of foundations. For more information about CALDER funders, see www.caldercenter.org/about-calder. All opinions expressed in this paper are those of the authors and do not necessarily reflect the views of our funders or the institutions to which the authors are affiliated.

Finally, we wish to thank James Cowan, Cap Peck, and Jim Wyckoff for comments that improved the manuscript, Trevor Gratz and Wezi Phiri for outstanding research assistance, and Nate Brown, Jessica Cao, Elliot Gao, Andrew Katz, Tony Liang, Arielle Menn, Natsumi Naito, Becca Ortega, Cameron Thompson, Stacy Wang, Malcolm Wolff, Hilary Wu, and Yunqi Zhang for their support with data collection and cleaning.

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

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

Exploring the Impact of Student Teaching Apprenticeships on Student Achievement and Mentor Teachers

Dan Goldhaber, John Krieg, Roddy Theobald
CALDER Working Paper No. 207-1118-1
November 2018

Abstract

We exploit within-teacher variation in the years that teachers host an apprentice (“student teacher”) in Washington State to estimate the causal effect of these apprenticeships on student achievement, both during the apprenticeship and afterwards. The average causal effect of hosting a student teacher on student performance in the year of the apprenticeship is precisely estimated and indistinguishable from zero in both math and reading, though effects are large and negative in math when ineffective teachers host an apprentice. Hosting a student teacher is also found to have modest positive impacts on student math and reading achievement in a teacher’s classroom in following years.

1. Introduction

Every year there are more than 125,000 student teachers who complete apprenticeships in K-12 public schools.¹ These apprenticeships occur in the classrooms of (and are supervised by) inservice teachers known as mentor teachers (or “cooperating teachers” in Washington, the setting of this study). Does hosting these teacher candidates affect student test performance, either during the apprenticeship or in the classrooms of mentor teachers after they have hosted a student teacher? As we describe below, there is a good deal of speculation about this, but no published quantitative exploration of the impacts on students in the classrooms where student teaching has taken place.²

The lack of information about how student teaching impacts K-12 students is problematic. States and localities make decisions about key contextual aspects of student teaching that influence whether there are positive or negative effects on students and the quality of the apprenticeship. While specific state-level requirements for mentor teachers are relatively rare, state laws occasionally mandate aspects of the field placements in which student teaching occurs,³ such as the diversity of the school in which student teaching occurs or the effectiveness or qualifications of the mentor teachers.⁴ Nevertheless, teacher education programs (TEPs) often have trouble finding student teacher placements for their candidates because of the perception

¹ This figure is based on the number of completers of “traditional” teacher preparation programs, as reported in federal Title II reports (<https://title2.ed.gov/Public/DataTools/Tables.aspx>). This is a lower bound on the total number of preservice teaching internships because many alternative programs (e.g., Teach for America) require some form of preservice teaching apprenticeship prior to workforce entry.

² There is some qualitative evidence suggesting that student teaching benefits K-12 students (e.g., Field and Philpott, 2000; Kerry and Shelton Mayes, 1995).

³ For instance, only 20% of states required that a mentor teacher hold a minimum level of professional experience or demonstrate mentoring skills in 2011 (Greenberg et al. (2011).

⁴ Similarly, locally negotiated memorandums of understanding between TEPs and school systems tend to focus on the broad nature of the student teaching experience without getting into the specifics of whether internships ought to take place in specific types of schools or be overseen by specific types of mentor teachers (Goldhaber et al., 2014).

that student teaching may be disruptive in ways that negatively impact students (St. John et al., 2018).

In this paper we explore the effects of hosting student teachers on the achievement of students in the host classroom and for future students of the mentor teacher. In particular, we utilize a unique, longitudinal database of student teachers from 15 TEPs that place student teachers in Washington State public schools to address three questions: 1) Does hosting a student teacher have an impact on student achievement in the classrooms in which student teaching occurs?; 2) Does hosting a student teacher have an impact on student achievement in the classrooms of mentor teachers in years after student teaching occurs?; and 3) Do these effects vary according to the prior effectiveness of mentor teachers?

Relying on within-teacher estimates of the effects of hosting student teachers, we find little evidence that hosting a student teacher impacts student achievement during the year of student teaching. Specifically, the average causal effect of hosting a student teacher on contemporaneous student performance is precisely estimated and indistinguishable from zero in both math and reading. However, we find modest positive impacts of supervising a student teacher on student math and reading achievement *in subsequent years*. Under our identification strategy, this estimate could be biased by patterns of student assignments after a teacher hosts a student teacher, but we find no evidence of non-random sorting of stronger classrooms to teachers who have previously hosted a student teacher, and this relationship is not different between schools that appear to track students to different classrooms on the basis of prior performance and schools that do not. Together, this supports the argument that these apprenticeships have benefits for mentor teachers that persist into the future.

Finally, we document heterogeneity in the impacts of hosting a student teacher by the prior mentor teacher value-added effectiveness. Most notably, while the *average* effect of hosting a student teacher on concurrent student performance is indistinguishable from zero, the effect of hosting a student teacher on student math performance is in fact relatively large and negative (about -.10 standard deviations of student performance) for mentor teachers with low prior value added. In other words, an ineffective teacher's students do *even worse* in years that the teacher serves as a mentor teacher than in years before the teacher served as a mentor teacher, all else equal. This suggests that the practice of using student teachers to support struggling teachers—as documented in Washington by St. John et al. (2018)—may be counter-productive. Instead, this finding suggests that more effective mentor teachers may be better able to mitigate the impact of hosting a student teacher on student performance, and further supports an emerging empirical research base (Goldhaber et al., 2018; Ronfeldt et al., 2018) supporting the placement of student teachers with more effective mentor teachers.

2. Background

Student teaching is widely regarded as the capstone to a teacher candidate's preparation experience (Anderson and Stillman, 2013). Not surprisingly then, a good deal of academic literature describes the role student teaching plays in the development of teacher candidates (Borko and Mayfield, 1995; Clarke et al., 2014; Ganser, 2002; Graham, 2006; Hoffman et al., 2015; Zeichner, 2009). More recently, researchers have investigated how the attributes of internship schools (Goldhaber et al., 2017; Ronfeldt 2012, 2015) and mentor teachers (Goldhaber et al., 2018; Ronfeldt et al., 2018) influences the later outcomes of student teachers who become public school teachers. Importantly for this study, both Goldhaber et al. (2018) and Ronfeldt et

al. (2018) find that teachers tend to have higher value when the mentor teacher of their student teaching placement has higher value added, all else equal, though Goldhaber et al. (2018) document that this relationship decays somewhat after candidates enter the workforce.

There is only a small academic literature addressing the ways in which schools or classrooms that host student teachers might themselves be affected, though much of this could be classified as speculation (St. John et al., 2018). But there are clear ways in which the role mentor teachers play in the mentorship of student teachers could lead to short- or longer-run effects on student achievement, either because of changes in resources or teaching practices. In the short-run, for instance, student teachers bring internship schools additional human resources, which could allow for more adult attention to the individual needs of students and greater differentiation of instruction; student teachers could also bring to schools more recently adopted practices taught in TEPs (Hurd, 2007).

There is also some suggestion that hosting a student teacher could confer benefits to mentor teachers. Kerry and Shelton Mayes (1995), for instance, argue that the act of helping student teachers dissect their classroom practices cause mentor teachers to reflect on their own practices in ways that lead to self-improvement.⁵ Field and Philpott (2000) provide survey evidence supporting the hypothesis as “mentors often claimed that they were forced to re-evaluate current practice in light of rationalizing their work to student teachers” (p. 127). There is also some evidence of this type of “peer learning” among inservice teachers (Jackson and Bruegmann, 2009; Papay et al., 2016).⁶

⁵ There is little quantitative evidence in support or opposition to this hypothesis for student teaching, but some evidence suggesting that the matching of higher and lower performing inservice teachers benefits both (Papay et al., 2016)

⁶ Specifically, Jackson and Bruegmann (2009) find that teachers tend to be more effective when they work in schools with more effective peers, while Papay et al. (2016) find large test score gains associated with the pairing of low-performing and high-performing teachers for professional development, with the gains concentrated in the classrooms of the low-performing teachers.

On the other hand, hosting student teachers could require a substantial time and resource commitment by mentor teachers which could divert attention from students, decreasing student achievement. Moreover, given the clear evidence of positive student achievement benefits associated with having teachers with greater experience (e.g. Ladd and Sorensen, 2017; Rivkin et al., 2005; Rockoff, 2004), and the possibility that experienced mentor teachers turn their classrooms over to inexperienced student teachers, we might expect negative effects on test scores in classrooms hosting student teachers.

These concerns are borne out in a qualitative study based on interviews with individuals responsible for student teacher placements in several TEPs, districts, and schools in Washington, the setting of this study (St. John et al., 2018). Specifically, principals and administrators responsible for student teaching placements for schools and districts reported protecting low-performing schools from student teachers—and in one case, declining all student teacher placements during a superintendent’s first year in the district—on the assumption that these placements could be disruptive in these settings. This study also illustrates stark differences in attitudes about the desirability of potential mentor teachers; while many placement coordinators and district administrators stressed the importance of finding highly-effective mentor teachers to support candidate development, some principals reported recruiting teachers with “stagnant” teaching practices to serve as a mentor teacher on the hypothesis that those teachers will benefit from serving as a mentor (St. John et al., p. 16).

In short, the sparse literature touching on how hosting student teachers impacts student achievement does not provide a clear theoretical direction about what we should expect; rather, it suggests that the effects of hosting likely depend on how student teachers are utilized. TEPs often provide guidelines on the length of internships and the hours teacher candidates are

required to be in the classroom, but little systemic information is known regarding the actual time breakdown of mentor-mentee interactions.⁷ It is generally understood that the hours mentor teachers typically spend mentoring, the frequency teacher candidates observe the mentor teacher in instruction, and the time mentor teachers observe instruction by the teacher candidate all vary both within and across TEPs (Greenberg et al., 2011). In some cases, having a teacher candidate intern may be highly interactive with the mentor-mentee relationship akin to a co-teaching environment (e.g, Heck & Bacharach, 2016), whereas in other scenarios mentor teachers may simply “hand off” the classroom and the corresponding responsibilities to the teacher candidate. While this analysis cannot test these differences directly, we can address whether—on average, across these different models of student teaching—hosting a student teacher appears to impact concurrent and subsequent student performance in a mentor teacher’s classroom.

3. Data and Summary Statistics

The data set we utilize combines student teaching data about teacher candidates from institutions participating in Washington State’s Teacher Education Learning Collaborative (TELC) with K–12 administrative data provided by Washington State’s Office of the Superintendent of Public Instruction (OSPI). During the years of this study, the TELC data observes student teaching placements from 15 of the state’s 21 college and university-based Washington State TEPs. This data includes when student teaching occurred, the schools in which teacher candidates completed their student teaching, and the mentor teachers that supervised

⁷ One exception is Bacharach et al. (2010), who find some evidence that students taught by a student teacher in a co-teaching setting have greater learning gains than students taught by a student teacher in a traditional setting.

these internships.⁸

Though many of the institutions in TELC provided student teaching data going back to the mid-2000s and, in one case, to the late 1990s, we focus on student teaching data from 2009-10 to 2014-15 in this analysis for two reasons. First, nearly all TEPs provided complete data about their teacher candidates over this time period.⁹ Second, these years of data correspond with years in which student-level data from OSPI can be linked to teachers through the state's CEDARS data system, introduced in the 2009-10 school year.¹⁰ By connecting the student teaching data from TELC institutions to the student-level data from OSPI, we create a dataset that links student teachers to: the K-12 students they taught during their student teaching placements; the students of the mentor teachers both before and after hosting the student teacher; and the public schools in which student teaching occurred.¹¹

Importantly, this dataset can be further linked to a number of additional variables about students and teachers. Specifically, the student-level data from OSPI includes annual standardized test scores in math and English Language Arts (ELA) and demographic/program participation data

⁸ The institutions participating in TELC and that provided data for this study include: Central Washington University, City University, Evergreen State College, Gonzaga University, Northwest University, Pacific Lutheran University, St. Martin's University, Seattle Pacific University, Seattle University, University of Washington Bothell, University of Washington Seattle, University of Washington Tacoma, Washington State University, Western Governors University, and Western Washington University. The 6 institutions that are not participating in TELC include one relatively (for Washington) large public institution in terms of teacher supply, Eastern Washington University, and five smaller private institutions: Antioch University, Heritage University, University of Puget Sound, Walla Walla University, and Whitworth University. Only two of these, the University of Puget Sound and Antioch University, are west of the Cascades.

⁹ 13 TEPs provided data for all six of these years, while 2 TEPs (both small private institutions) provided data for only three of these six years. Although programs provided data on mentor teachers in a variety of formats, we are able to match 97% of teacher candidates in the TELC data whose program provided mentor teaching information and who did their student teaching in public schools in Washington to a valid mentor teacher observation in the OSPI data.

¹⁰ 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.

¹¹ Note that, while many placements occurred in private schools and out-of-state schools, we do not consider these placements in this analysis because we do not have data about these schools or the students and teachers in these schools.

for all K-12 students in the state, while the OSPI personnel data include information on teachers' years of teaching experience. We standardize student test scores by grade and year and use these as the dependent variable in the analytic models described in the next section, while the other variables are used as control variables in these regressions.

As noted above, the TELC data include only 15 of the 21 TEPs that placed student teachers in Washington public schools during the 2009-10 to 2014-15 period. However, as shown in **Figure 1**, these participating TEPs are distributed unevenly throughout the state. Specifically, Figure 1 shows the percentage of newly-hired, in-state teachers between 2010 and 2015 in each district in the state who graduated from one of the TEPs participating in TELC (and thus included in this study).¹² While 81 percent of all new teachers in Washington State are prepared by TEPs participating in TELC, this figure is 92% for districts west of the Cascade Mountains (the pink line in Figure 1) and just 55% for districts east of the Cascades.¹³ Because our empirical models rely on identifying which and when mentor teachers host student teachers, we include only observations west of the Cascades where we incorrectly mis-identify relatively few mentor teachers as not hosting student teachers.¹⁴

Since we are focused on estimating the effects of hosting a student teacher on student test scores, we further restrict the data to math and ELA teachers in grades 4-8, because students in these teachers' classrooms can be linked both to current and prior test performance in these

¹² We can obtain this estimate because the OSPI data include information the institutions from which teachers (not teacher candidates) receive their teaching credentials. About 22 percent of new teachers come in from out of state (and receive an OSPI credential) (Goldhaber et al., 2013), and are not included in these figures or this analysis.

¹³ The gap in coverage of student teachers is primarily driven by the fact that the three largest TEPs not participating in TELC are all in the eastern half of the state.

¹⁴ Our empirical models rely upon within-CT variation in student test scores. Mis-identifying a CT as not hosting a student teacher when, in fact they do, simply results in that CT being dropped from the analysis (since there is no variation in their hosting status over the time observed). This would result in biasing our empirical models only to the extent that student teachers from the one west-side program not observed in the TELC data have a differential effect on their mentor teachers classroom relative to our observed CTs.

subjects. Finally, we restrict the data to teachers who serve as a mentor teacher at least once in the period over which we have TELC data. Based on these restrictions, our analytical sample includes 1,319 student teachers (1,086 unique mentor teachers) in math classrooms and 1,358 student teachers (1,107 unique mentor teachers) in ELA classrooms.

Table 1 reports summary statistics for K-12 students who are in the classrooms of teachers who identify our estimates of the impact of student teaching (i.e., teachers who host a student teacher in at least one but not all years of observed data).¹⁵ Importantly, these are *not* the only students and teachers who are included in our analytic models, as other students and teachers help identify the relationships between other variables in our analytic models (e.g., prior student performance and teacher experience) and student achievement.

Because the analysis described in the next section relies on within-teacher comparisons between years before, during, and after student teaching placements in that teacher’s classroom, we create indicators for these three different periods for each mentor teacher and present summary statistics separately for each period. In this and subsequent analyses, the “Before student teaching year” period is defined as all years before the *first* year a teacher hosted a student teacher; the “Student teaching year” period is defined as any years in which the teacher hosted a student teacher; and the “After student teaching year” period is defined as all years the teacher did not host a student teacher after the *first* year a teacher hosted a student teacher.

A few interesting findings from Table 1 are relevant for the analysis described in the next section. Mentor teachers, for instance, have students who enter their classrooms with higher baseline tests in the year they host a student teacher than before and even higher in years afterward. On the other hand, by some measures, such as the proportion of students with Limited

¹⁵ For brevity, we do not show ELA descriptive statistics because they are qualitatively similar to the math descriptive statistics shown in Table 1. The ELA descriptive statistics are in Appendix Table A1.

English Proficiency, it appears they have more disadvantaged classrooms in this year. This may reflect the possibility, discussed in St. John et al. (2018), that teachers are more likely to host a student teacher in years in which they have a more “difficult” classroom.

While we are able to control for these observable differences between the composition of the teacher’s classroom during the student teaching year and other years in the analytic models described in the next section, a key identifying assumption of these models is that teachers are not more likely to host a student teacher when they have a more difficult classroom *along unobserved dimensions*, conditional on the observed variables in Table 1. It is concern about unobserved differences in student assignment by mentor teacher status that motivates robustness and a falsification test that we describe in the next section.

4. Analytic Approach

Our analytic approach follows Taylor and Tyler (2012), who rely on within-teacher variation in teacher evaluations to estimate the causal effect of these evaluations on concurrent and future student achievement. We begin by estimating a teacher fixed effects model predicting student performance for student i in the classroom of teacher j in subject s and year t , Y_{ijst} , as a function of prior student performance, $Y_{i(t-1)}$, additional student variables in Table 1, S_{it} , binary indicators for each year of teacher experience, Exp_{jt} , and an indicator for whether teacher j hosted a student teacher in year t , ST_{jt} .¹⁶

$$Y_{ijst} = \alpha_0 + \alpha_1 Y_{i(t-1)} + \alpha_2 S_{it} + \alpha_3 ST_{jt} + \sum_k \beta_k I_k(Exp) + \tau_{js} + \rho_t + \varepsilon_{ijst} \quad (1)$$

¹⁶ The experience and year effects in this model can be separately identified because of teachers who have gaps in their teaching experience. However, this raises the possibility that this source of identification may lead to imprecise controls for returns to teacher experience. We therefore experiment with models that omit the year controls and find that the results are very similar.

The parameter of interest in equation 1 is α_3 , which can be interpreted as the average difference in student performance, all else equal, between years in which a teacher hosted a student teacher and years in which the *same* teacher did not host a student teacher. We cluster all standard errors at the teacher level.

We argue that identifying the impact of student teaching based on within-teacher over-time variation is preferable to identification in the cross-section. Evidence shows that mentor teachers are likely to be slightly more effective teachers than non-mentor teachers (Krieg et al., 2018), though qualitative research also points to the possibility that some apprenticeship assignments are made to help give mentor teachers “a break,” (St. John et al., 2018) and thus may not be based on the instructional quality of the mentor. Either way, an unobserved correlation between teacher effectiveness and assignment as a mentor teacher would lead to a biased finding in the cross-section on the impact of hosting a student teacher.

The teacher fixed effect accounts for time-invariant differences between teachers, but this assignment may also have a dynamic component (Rothstein, 2010), which may also lead to bias. We see, for instance, in Table 1 that teachers appear to be assigned somewhat more disadvantaged students in the year in which they serve as a mentor teacher. We account for observable student characteristics in (1), but unobserved student ability that is correlated with the mentor teacher assignment status would also lead to biased estimates of α_3 .

We address this possibility in two ways. First, we follow Clotfelter et al. (2006) and Horvath (2015) and estimate models restricted to schools in which students are distributed relatively equitably across classrooms according to prior performance, on the assumption that these schools are also the least likely to non-randomly sort students to classrooms along unobserved dimensions. While this data restriction reduces the power of our test, it also limits

the possibility of biasing α_3 by limiting the scope of unobserved sorting of students into (or out of) mentor teacher’s classrooms. Second, we perform a falsification test in which we replace the dependent variable in equation 1, Y_{ijst} , with students’ *prior* test scores $Y_{is(t-1)}$, and control for twice-lagged test scores $Y_{i(t-2)}$ on the right-hand side. If teachers are systematically assigned to “better” classrooms in years they either do or do not host a student teacher, we would observe an “effect” of student teaching on these lagged test scores in these models attributable to unobserved differences in student background.

Another concern is that equation 1 makes comparisons between the student teaching year and *all other years* a teacher did not host a student teacher (before and after). However, the literature on student teaching discussed in Section 2—as well as evidence about peer learning in K-12 schools (Jackson and Bruegmann, 2009; Papay et al., 2016; Taylor and Tyler, 2012)—suggests that we should treat the years *after* a teacher hosts a student teacher differently than the years before hosting. We therefore extend the model in equation 1 to include an additional term, AF_{jt} , indicating whether year t is after teacher j hosted a student teacher for the first time (and is not itself a student teaching year):

$$Y_{ijst} = \alpha_0 + \alpha_1 Y_{i(t-1)} + \alpha_2 S_{it} + \alpha_3 ST_{jt} + \alpha_4 AF_{jt} + \sum_k \beta_k I_k(Exp) + \tau_{js} + \rho_t + \varepsilon_{ijst} \quad (2)$$

The parameters of interest in equation 2, α_3 and α_4 , compare student performance in years during and after student teaching placements (respectively) to student performance in years before the teacher hosted a student teacher, all else equal.

Finally, as we noted in Section 2, there is reason to believe that there could be heterogeneous effects of serving as a mentor teacher associated with a mentor teachers prior effectiveness as a teacher. To investigate this possibility, we split the six-year sample into two periods: we use the data from 2009-10 and 2010-11 (dropping years in which teachers host a

student teacher) to generate a measure of prior teacher value added; and then use the data from 2011-12 through 2014-15 to estimate an extension of the model in equation 2 with additional interactions using prior mentor teacher value added.¹⁷ Specifically, we first estimate teacher value added from 2009-10 and 2010-11 using the model in equation 3, where all terms are defined the same as above:

$$Y_{ijst} = \gamma_0 + \gamma_1 Y_{i(t-1)} + \gamma_2 S_{it} + \sum_{k=1}^{50} \beta_k I_k(Exp) + \tau_{js} + \rho_t + \varepsilon_{ijst} \quad (3)$$

From this equation, we interact the Bayesian-adjusted value added estimate $\hat{\tau}_{js}$, which we refer to as “Prior VA,” with the student teaching terms of interest in models estimated from data between 2011-2012 through 2014-15¹⁸:

$$Y_{ijst} = \alpha_0 + \alpha_1 Y_{i(t-1)} + \alpha_2 S_{it} + \alpha_3 ST_{jt} + \alpha_4 ST_{jt} \times \hat{\tau}_{js} + \alpha_5 AF_{jt} + \alpha_6 AF_{jt} \times \hat{\tau}_{js} + \sum_k \beta_k I_k(Exp) + \rho_t \times \hat{\tau}_{js} + \tau_{js} + \varepsilon_{ijst} \quad (4)$$

There are four parameters of interest in equation 4: α_3 is the average difference in student performance between the student teaching year and years prior to hosting student teacher for teachers with *average* prior value added; α_4 describes how this relationship changes as prior value added increases; α_5 represents the average difference in student performance between years *after* hosting a student teacher and years prior to hosting student teacher for teachers with average prior value added; and α_6 describes how this relationship changes as prior value added increases. In our preferred specifications of the model in equation 4, we include both linear and squared terms of prior value added to capture non-linear relationships between prior mentor teacher effectiveness

¹⁷ We chose these samples to maximize the number of teachers who have an estimate of prior value added from the first period and host a student teacher in the second period.

¹⁸ Empirical Bayes (EB) methods shrink the value added estimates back to the grand mean of the value-added distribution in proportion to the standard error of each estimate. EB shrinkage does not account for the uncertainty in the grand mean, suggesting that estimates may shrink too much under this procedure (McCaffrey et al., 2009); this approach, however, ensures that estimates in the tail of the distribution are not disproportionately estimated with large standard errors. An appendix on Empirical Bayes shrinkage is available from the authors upon request.

and student performance during and after student teaching placements. Importantly, these models also interact prior value added with the school year to control for regression to the mean (discussed in the next section). It also should be noted that since we use two of our six years of our data to compute prior VA, models that estimate equation 4 have significantly fewer observations than those used in equations 1 and 2.¹⁹

A potential drawback of the model in equation 4 has to do with its reliance on a mentor teacher's measure of prior value added to explain future student learning. As shown by Atteberry et al. (2015), it is likely that mentor teachers who did particularly well (poorly) in one year, will do worse (better) in subsequent school years; as discussed in Goldhaber and Hanson (2013), this will be true even when value added is shrunken using EB methods (described above), as EB shrinkage is unlikely to provide an accurate estimate of the “permanent component” of teacher performance. This regression to the mean can be seen in Figure 2 which shows the evolution of future value added as a function of measured value added in prior years. From this figure, it is clear that mentor teachers who have high value added in 2009-10 and 2010-11, on average, perform closer to the mean in future years, while mentor teachers with low measures of value added in the early period tend to score better. Given that equation 4 uses prior value added to explain future performance, the potential regression to the mean may bias α_3 downwards. We correct for this possibility by interacting prior value added with the time dummies, ρ_t . Including these terms removes any systematic change in expected value added caused by regression to the mean over time.

¹⁹ In Appendix Table A3, we illustrate the implications of estimating a model that: a) estimates value added with a “leave-one-out” specification that includes all years other than the student teaching year; and b) does not directly account for regression to the mean in interacting this estimate with an indicator for whether the teacher hosted a student teacher. Estimates from this model lead to the conclusion that there is a positive effect of hosting a student teacher for ineffective teachers and a negative effect of hosting a student teacher for effective teachers, even when the student teaching indicator is a “placebo” that is randomly assigned to teachers in the sample.

A second potential drawback of the model in equation 4 is that, because prior value added is collinear with the teacher fixed effect, we cannot control for prior value added directly in the model. We therefore estimate an alternative to this model that swaps out the teacher fixed effect for a direct control for prior value added and its square:

$$Y_{ijst} = \alpha_0 + \alpha_1 Y_{i(t-1)} + \alpha_2 S_{it} + \alpha_3 ST_{jt} + \alpha_4 ST_{jt} \times \hat{t}_{js} + \alpha_5 AF_{jt} + \alpha_6 AF_{jt} \times \hat{t}_{js} + \alpha_7 \hat{t}_{js} + \alpha_8 \hat{t}_{js}^2 + \rho_t \times \hat{t}_{js} + \sum_k \beta_k I_k(Exp) + \varepsilon_{ijst} \quad (5)$$

While the model in equation 4 is our preferred specification for investigating heterogeneity by prior value added, we present findings from the models in both equations 4 and 5 to test the robustness of our findings to these different specifications.

5. Results

Does hosting a student teacher have an impact on student achievement during or after the student teaching placement?

The estimates of the parameters of interest from equations 1 and 2 are presented in Table 2. The first through fourth columns present results from math, while columns 5 through 8 present results for ELA. Columns 1 and 5 of Table 2 contain estimates from the model in equation 1, in which student performance in the student teaching year is compared to student performance in the same teacher's classroom both before and after the student teaching year. Relative to years without a student teacher, the estimate for math (column 1) implies that a mentor teacher's students score 0.018 standard deviations lower in math, all else equal. The comparable estimate for ELA is very close to zero and not statistically significant.

As described by equation 2, the second and sixth columns of Table 2 include AF , the identifier for years after a mentor teacher hosts a student teacher. The results with AF

demonstrates that students perform as well in a mentor teacher’s class when that mentor teacher hosts a student teacher as they would have in prior years. Specifically, when student performance in the student teaching year is compared to the student performance in years before student teaching (the estimated coefficients on “Year Hosting Student Teacher” in columns 2 and 6 of Table 2), the estimates in both math and ELA are indistinguishable from zero. These estimates are quite precisely estimated—i.e., the standard error of each estimate is less than 0.01 standard deviations of student performance—meaning that we can rule out relatively small average impacts in both subjects. We therefore conclude that student teaching placements have minimal impact on student performance in the year the mentor teacher hosts a student teacher.

However, the comparisons between years *after* the mentor teacher hosts a student teacher and the years prior to student teaching placement reveal modest, positive effects in both math and ELA. The estimates in each subject suggest that a teacher’s students score 0.02-0.03 standard deviations higher in years after they host a student teacher, all else equal, than in years before the teacher hosts a student teacher.²⁰ These estimates control for returns to teacher experience, so cannot be attributed to teachers gaining experience over time. Our interpretation is that these estimates support the notion that mentor teachers benefit from hosting a student teacher and improve their performance in subsequent years. We stress, though, that these effects are modest; for example, the estimated effect in both subjects is less than half of the returns to the first year of teaching experience in that subject (estimated from our models).

The third and sixth columns of Table 2 restrict the sample to schools that demonstrate minimal sorting between classrooms. This restriction is intended to eliminate observations in

²⁰ We refer to these effects as modest because they are roughly half the estimated effect of a one standard increase in average peer quality (Jackson & Bruegmann, 2009) and less than a fourth of the estimated effect of assignment of a low-performing teacher to work with a higher-performing peer in the school (Papay et al., 2016).

which principals can purposely sort students across classrooms, for instance by placing rewarding teachers with easier (along unobserved dimensions) to educate students after serving as a mentor teacher. This type of behavior would mean that the positive, post-hosting effects could be attributed to student composition not controlled for by our student measures rather than improved teaching on the part of mentor teachers. Eliminating the schools where it appears that students are non-randomly placed in classrooms across years does little to change the coefficients on AF , in both Math and ELA the coefficients are statistically similar to those of columns 2 and 3. However, neither is statistically significant which we attribute to the larger standard errors caused by restricting the sample size.²¹

The final columns in Table 2, columns 4 and 8, present results from the falsification exercise. In this exercise, equation 2 is estimated using the test score performance earned by students during the year prior to enrollment in the class taught by the mentor teacher. If teachers are systematically assigned to “better” (in the sense that students have unobserved characteristics associated with better test outcomes) classrooms in years they either do or do not host a student teacher, this would be reflected in the relationships between these years and the prior performance of their students. However, we do not find significant relationships between these years and student prior performance.

Does the impact of hosting a student teacher vary by prior mentor teacher effectiveness?

Table 3 explores heterogeneity in these effects by the prior value added of mentor teachers, estimated from those years these teachers did not host a student teacher (i.e., equations 4 and 5). Columns 1 and 4 present estimates from our preferred mentor teacher fixed effect specification on the full sample. The second and fifth columns repeat this specification on the

²¹ Note the decrease in the sample size by 15% increases the standard errors by 9%, which is nearly exactly the growth in the standard errors between columns 2 and 3 of Table 2.

sample restricted to buildings that do not appear to sort students between classrooms. The final set of columns, the third and sixth, present results using the prior value-added models of equation 5 on the entire sample. The variables of interest are the interaction of prior value added with the binary variables indicating the presence of a student teacher and the identifier for years after hosting a student teacher. In both cases, we include quadratics in prior value added which makes the interpretation of these coefficients difficult. For ease of interpretation, we translate Table 3 into Figures 3 and 4 and focus on these figures in our discussion.

Figure 3 presents graphical representation of the impact on students of contemporaneously hosting a student teacher derived from Table 3. Math (first column) and ELA (second column) findings are presented from each of the three models in this table: the first row of Figure 3 presents our preferred teacher effects model; the second row restricts the sample to schools with little selection of students between classrooms; and the third row replaced the teacher fixed effects with measures of lagged value added. The first conclusion we draw from this figure is that under all three model specifications the impact of hosting a student teacher on *current math performance* (Panels A and B) is negative for teachers with very low prior value added (i.e., two standard deviations below the mean in the sample) and statistically significant for the two full-sample results. This is perhaps surprising given that, as discussed in St. John et al. (2018), principals sometimes report using student teaching placements to support struggling teachers. These results suggest that this approach may be counter-productive, as these teachers' students perform *even worse* in years they host a student teacher, all else equal. The ELA results show the opposite pattern except no estimates are statistically significant suggesting that there is no heterogeneity by prior value added within this subject.

Figure 4 examines heterogeneity in performance after a mentor teacher hosts a student

teacher. The results displayed in this Figure differ somewhat from Table 2's earlier findings in that these results exclude the first two years of data in order to estimate prior value added. In all three math specifications there is little evidence of post-student teacher heterogeneity. Among the ELA results, the preferred fixed effects model shows that high value added mentor teacher's tend to perform about .15 standard deviations better after hosting a student teacher, though the restricted sample has this impact decrease by about half with a concomitant increase in standard error that leaves this finding statistically insignificant.

6. Discussion and Conclusions

The primary findings we present provide evidence that apprenticeships in teaching are not generally harmful to the achievement of students in the classrooms in which those apprenticeships take place. This is a somewhat surprising finding given that experienced teachers are, in some cases, turning over their classrooms to truly novice mentees, though perhaps can be explained by the growing prevalence of "co-teaching" arrangements in the student teaching apprenticeship (e.g, Heck & Bacharach, 2016). It is also an important finding since schools facing accountability pressures may be reluctant to host student teachers fearing adverse impacts on student achievement, and our null findings could help assuage these concerns.

The heterogeneity of the estimated effects by the effectiveness of the mentor teacher is also an important finding. There is currently wide variation in the effectiveness of teachers assigned as a mentor teachers (Goldhaber et al., 2018); a related literature suggests that mentees are better off in the long-run (in terms of their future productivity) when they have a more effective mentor (Goldhaber et al., 2018; Matsko et al., 2018; Ronfeldt et al., 2018a,b), and our findings suggest that short-run benefits to the students in the classrooms hosting student teachers

when apprenticeships are supervised by more effective mentors. A clear implication of this for policy and practice is that school districts and teacher education programs might use measures of effectiveness for consideration of who serves as a mentor. Given that about 3% of teachers host a student teacher in a given year—at least in Washington, the setting of this study (Goldhaber et al., 2018), there is wide scope for school administrators to engage many teachers along these lines.

More generally, our findings support a small but growing literature showing that peer learning (Jackson & Bruegmann, 2009; Papay et al., 2016) appears to be an important means of improving incumbent teachers as teachers are found to be more effective *after* having the experience of having served as a mentor. This finding is robust to a number of specification and falsification checks, and has important implications for the schools and districts in which student teaching apprenticeships take place.²² In fact, taken together, our findings suggest that school districts should be far more invested in the student teaching process than they often are (St. John et al., 2018), as there appear to be few short-term costs and modest long-term benefits to the schools and districts that host these apprenticeships.

²² As one example, given that teachers report that serving as a mentor is time consuming and there is little financial reward—by one estimate, the average compensation that mentor teachers receive in 2012-13 is \$232, far lower than the nearly \$1600 (adjusted for inflation) that was typical back in 1959 (Fives et al., 2016)—there is a compelling argument for additional rewards for serving in this role.

References

- Anderson, L. M., & Stillman, J. A. (2013). Student teaching's contribution to preservice teacher development: A review of research focused on the preparation of teachers for urban and high-needs contexts. *Review of Educational Research*, 83(1), 3-69.
- Atteberry, A., Loeb, S., & Wyckoff, J. (2015). Do first impressions matter? Predicting early career teacher effectiveness. *AERA Open*, 1(4), 2332858415607834.
- Bacharach, N., Heck, T. W., & Dahlberg, K. (2010). Changing the face of student teaching through coteaching. *Action in Teacher Education*, 32(1), 3-14.
- Borko, H., & Mayfield, V. (1995). The roles of cooperating teacher and university supervisor in learning to teach. *Teaching & Teacher Education*, 11(5), 501-518.
- 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.
- Clark, M. A., Chiang, H. S., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). The effectiveness of secondary math teachers from Teach For America and the Teaching Fellows programs. Princeton, NJ: Mathematica Policy Research.
- Clarke, A., Triggs, V., & Nielsen, W. (2014). Cooperating teacher participation in teacher education: A review of the literature. *Review of Educational Research*, 84(2), 163-202.
- Field, K. & Philpott, C. (2000). The impact of hosting student teachers on school effectiveness and school improvement. *Journal of In-Service Education*, 26(1), 115-137.
- Fives, H., Mills, T. M., & Dacey, C. M. (2016). Cooperating teacher compensation and benefits: Comparing 1957-1958 and 2012-2013. *Journal of Teacher Education*, 67(2), 105-119.
- Ganser, T. (2002). How teachers compare the roles of cooperating teacher and mentor. In *The educational forum* (Vol. 66, No. 4, pp. 380-385). Taylor & Francis Group.
- Goldhaber, D., Krieg, J., & Theobald, R. (2014). Knocking on the door to the teaching profession? Modeling the entry of prospective teachers into the workforce. *Economics of Education Review*, 43, 106-124.
- Goldhaber, D., Krieg, J. M., & Theobald, R. (2017). Does the match matter? Exploring whether student teaching experiences affect teacher effectiveness. *American Educational Research Journal*, 54(2), 325-359.
- Goldhaber, D., Krieg, J., & Theobald, R. (2018). Effective like me? Does having a more productive mentor improve the productivity of mentees? CALDER Working Paper 208-1118-1.

- 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, 29-44.
- Graham, B. (2006). Conditions for successful field experiences: Perceptions of mentor teachers. *Teaching and teacher education*, 22(8), 1118-1129.
- Greenberg, J., Pomerance, L., & Walsh, K. (2011). Student Teaching in the United States. *National Council on Teacher Quality*.
- Greenberg, J., McKee, A., & Walsh, K. (2013). Teacher prep review: A review of the nation's teacher preparation programs.
- Hamman, D., Fives, H., & Olivarez, A. (2007). Hamman, D., Fives, H., & Olivarez, A. (2007). Efficacy and pedagogical interaction in cooperating and student teacher dyads. *The Journal of Classroom Interaction*, 41/42, 55-63.
- Heck, T. W., & Bacharach, N. (2016). A better model for student teaching. *Educational Leadership*, 73(4), 24-29.
- Hoffman, J. V., Wetzel, M. M., Maloch, B., Greeter, E., Taylor, L., DeJulio, S., & Vlach, S. K. (2015). What can we learn from studying the coaching interactions between mentor teachers and preservice teachers? A literature review. *Teaching and Teacher Education*, 52, 99-112.
- Hurd, S. (2007). The impact of trainee teachers on school achievement: A review of research. Centre for Research & Development in Teacher Education, The Open University.
- Kerry, T. & Shelton Mayes, A. (Eds) (1995) *Issues in Mentoring*. Buckingham: Open University Press.
- Krieg, J. M., Goldhaber, D., & Theobald, R. (2018). Teacher candidate apprenticeships: Assessing the who and where of student teaching. CALDER Working Paper 206-1118-1.
- Krieg, J. M., Theobald, R., & Goldhaber, D. (2016). A foot in the door: Exploring the role of student teaching assignments in teachers' initial job placements. *Educational Evaluation and Policy Analysis*, 38(2), 364-388.
- Ludwig, L. (2017). Will you be my mentor? Self-selecting a mentor during student teaching. All NMU Master's Theses. 134. <http://commons.nmu.edu/theses/134>
- 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.
- McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy*, 4(4), 572-606.

- Reininger, M. (2012). Hometown disadvantage? It depends on where you're from. Teachers' location preferences and the implications for staffing schools. *Educational Evaluation and Policy Analysis*, 34(2), 127-145.
- 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., Brockman, S., & Campbell, S. (2018). Does cooperating teachers' instructional effectiveness improve preservice teachers' future performance? *Educational Researcher*.
- St. John, E., Goldhaber, D., Krieg, J., & Theobald, R. (2018). How the match gets made: Exploring student teacher placements across teacher education programs, districts, and schools. CALDER Working Paper 204-1018-1.
- Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, 102(7), 3628-51.
- Zeichner, K. M. (2009). *Teacher education and the struggle for social justice*. Routledge, New York, NY.
- Zeichner, K. (2010). Rethinking the connections between campus courses and field experiences in college-and university-based teacher education. *Journal of Teacher Education*, 61(1-2), 89-99.

Tables and Figures

Table 1. Student Summary Statistics (Mentor Teachers in Math Only)

	Before student teaching year	Student teaching year	After student teaching year
Baseline score in math	0.026 (0.976)	0.056 (0.992)	0.092* (0.995)
Baseline score in ELA	0.016 (0.965)	0.048 (0.974)	0.076* (0.968)
Female	0.491	0.491	0.493
American Indian	0.014	0.011**	0.009***
Asian / Pacific Isl.	0.119	0.130+	0.122
Black	0.059	0.068*	0.056
Hispanic	0.144	0.146	0.149
White	0.607	0.584*	0.594
Learning disability	0.057	0.056	0.054
Special Education	0.113	0.115	0.110
Gifted	0.061	0.069	0.074
Limited English	0.050	0.061***	0.061**
Free/Reduced Lunch	0.432	0.429	0.414
Unique teachers	1,086	1,086	1,086
Unique student teachers	0	1,319	0
Unique students	58,338	45,063	69,264

Note. ELA = English Language Arts. P-values from two-sided t-test relative to years before student teaching: +p<0.10; *p<0.05; **p<0.01; ***p<0.001.

Table 2. Effects of hosting a student teacher on standardized student achievement

	Math				ELA			
	1	2	3	4	5	6	7	8
Year Hosting Student Teacher	-0.018* (0.008)	-0.004 (0.010)	-0.005 (0.011)	0.002 (0.012)	0.001 (0.007)	0.013 (0.009)	0.012 (0.011)	0.001 (0.010)
After Hosting Student Teacher		0.025* (0.012)	0.022 (0.014)	0.005 (0.015)		0.022* (0.011)	0.020 (0.012)	-0.006 (0.012)
Students	1,050,090	1,050,090	883,354	853,714	972,201	972,201	773,769	780,206
Sample Years	2010-2015	2010-2015	2010-2015	2010-2015	2010-2015	2010-2015	2010-2015	2010-2015
Model	Full	Full	Minimal Sorting	Falsification	Full	Full	Minimal Sorting	Falsification

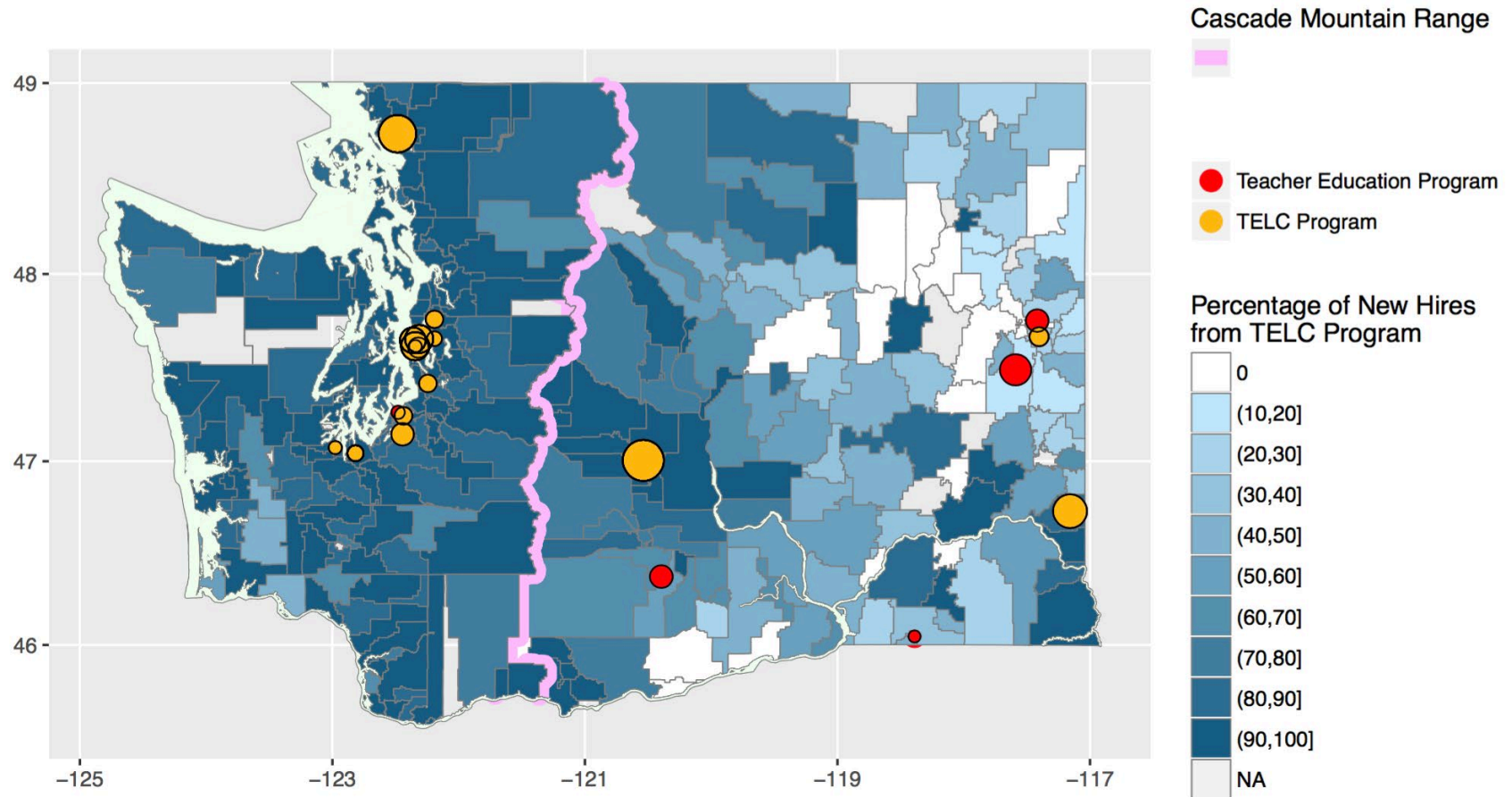
Note. ELA = English Language Arts. All models include a teacher fixed effect, indicators of annual teacher experience and the school year, and also control for the following student control variables interacted by grade: prior performance in math and reading, gender, race/ethnicity, receipt of free or reduced priced lunch, special education status and disability type, Limited English Proficiency indicator, migrant indicator, and homeless indicator. Minimal sorting models are estimated only on the subset of schools in which there is not significant sorting of students across classrooms by prior performance, and falsification models predict lagged tests scores as a function of twice-lagged test scores and the other variables above. Standard errors clustered at the teacher level are in parentheses. P-values from two-sided t-test: +p<0.10; *p<0.05; **p<0.01; ***p<0.001.

Table 3. Effects of hosting a student teacher by prior value added

	Math			ELA		
	1	2	3	4	5	6
Year Hosting Student Teacher (ST)	0.021 (0.018)	0.017 (0.020)	0.003 (0.013)	0.007 (0.018)	0.012 (0.021)	0.019+ (0.011)
After Hosting Student Teacher (After)	0.004 (0.022)	-0.011 (0.025)	0.038** (0.012)	-0.013 (0.023)	-0.007 (0.027)	0.021* (0.010)
Prior VA			0.648*** (0.017)			0.590*** (0.020)
Prior VA Squared			0.116** (0.043)			0.440*** (0.063)
Prior VA * ST	0.126 (0.083)	0.150+ (0.089)	0.085+ (0.049)	-0.038 (0.112)	0.010 (0.123)	-0.053 (0.065)
Prior VA Squared * ST	0.013 (0.108)	-0.341 (0.279)	-0.001 (0.051)	-0.082 (0.140)	-0.119 (0.513)	-0.087 (0.072)
Prior VA * After	-0.449+ (0.267)	0.016 (0.115)	-0.113 (0.146)	0.237 (0.482)	-0.028 (0.151)	-0.162 (0.241)
Prior VA Squared * After	0.149 (0.341)	0.354 (0.366)	0.018 (0.198)	1.282* (0.561)	0.696 (0.588)	0.383 (0.313)
Students	469,868	403,110	469,868	420,243	338,340	420,243
Sample Years	2012-2015	2012-2015	2012-2015	2012-2015	2012-2015	2012-2015
VA Years	2010-2011	2010-2011	2010-2011	2010-2011	2010-2011	2010-2011
Sample	Full	Minimal Sorting	Full	Full	Minimal Sorting	Full
Model	Tch FE	Tch FE	Lagged VA	Tch FE	Tch FE	Lagged VA

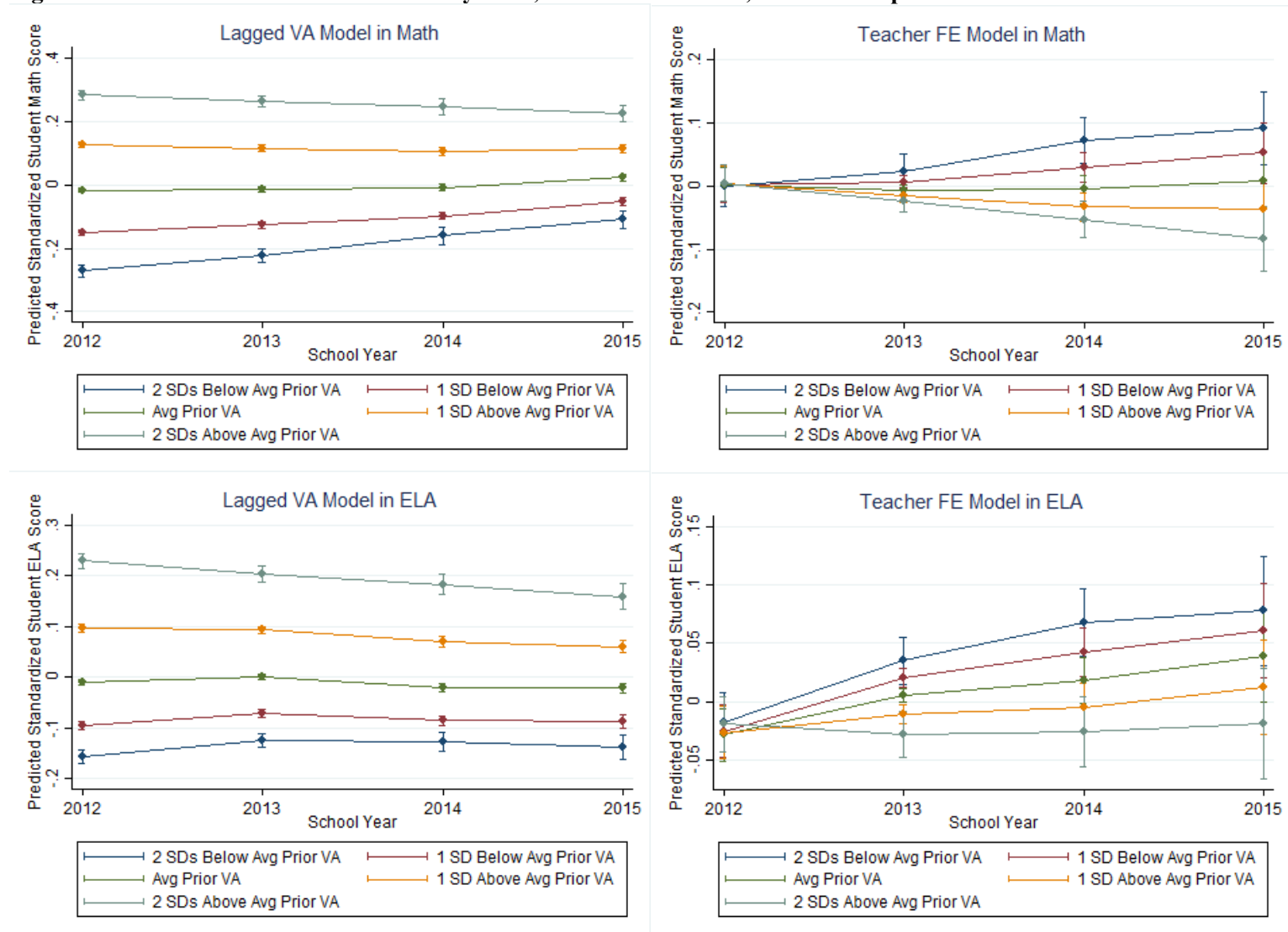
Note. ELA = English Language Arts; VA = value added. All models include a teacher fixed effect, indicators of annual teacher experience and the school year, interactions between prior VA (linear and squared) and school year, and also control for the following student control variables interacted by grade: prior performance in math and reading, gender, race/ethnicity, receipt of free or reduced priced lunch, special education status and disability type, Limited English Proficiency indicator, migrant indicator, and homeless indicator. Minimal sorting models are estimated only on the subset of schools in which there is not significant sorting of students across classrooms by prior performance. Standard errors clustered at the teacher level are in parentheses. P-values from two-sided t-test: +p<0.10; *p<0.05; **p<0.01; ***p<0.001.

Figure 1. Percentage of Newly-Hired, In-State Teachers from Participating TEPs, 2010-2015



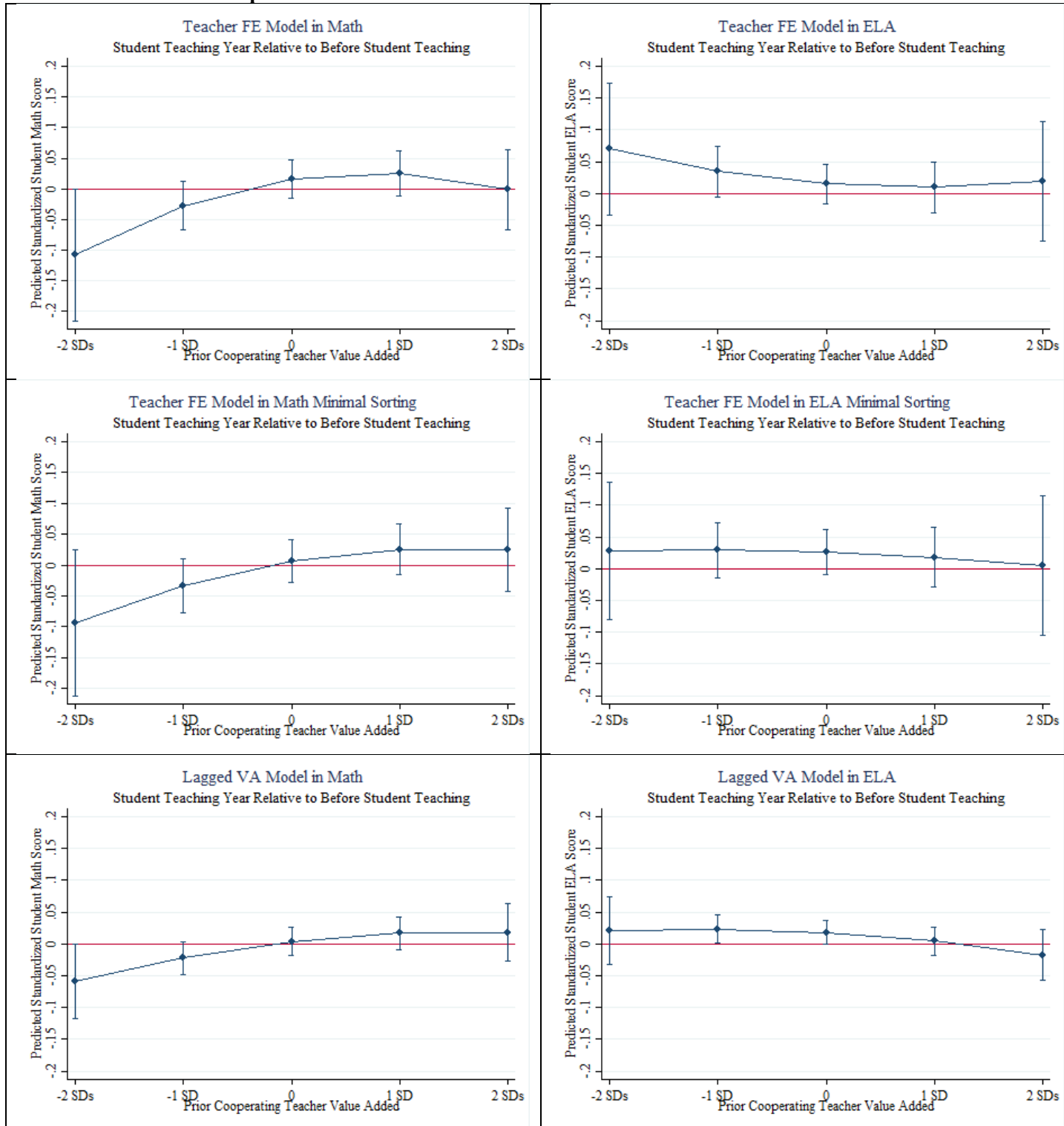
Note. TELC = Teacher Education Learning Collaborative (i.e., participating program); TEP = teacher education program. The diameter of the dot for each TEP is proportional to the number of newly-credentialed teachers from that TEP between 2010 and 2015.

Figure 2. Predicted Student Achievement by Year, Prior Value Added, and Model Specification



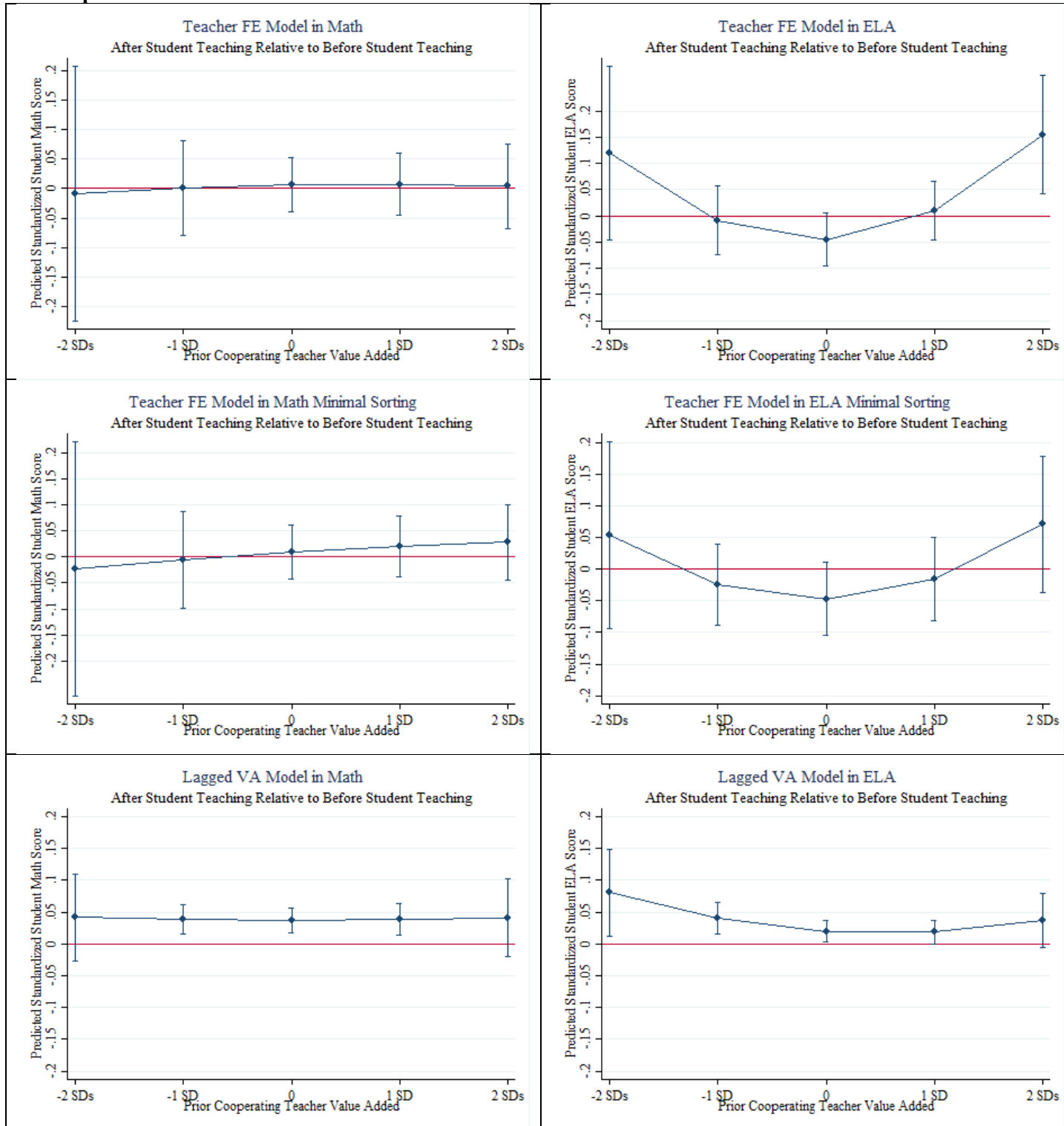
Note. ELA = English Language Arts. Predicted student achievement and associated 90% confidence intervals calculated from estimates in Table 3.

Figure 3. Effects on Contemporaneous Student Achievement of Hosting a Student Teacher by Prior Value Added and Model Specification



Note. ELA = English Language Arts. Estimated effects and associated 90% confidence intervals calculated from estimates in Table 3.

Figure 4. Effects on Future Students of Hosting a Student Teacher by Prior CT Value Added and Model Specification



Note. ELA = English Language Arts. Estimated effects and associated 90% confidence intervals calculated from estimates in Table 3.

Appendix

Table A1. Student Summary Statistics (Mentor Teachers in ELA Only)

	All	Before student teaching year	Student teaching year	After student teaching year
Prior score in math (standardized)	0.076 (0.988)	0.081 (0.988)	0.057 (0.996)	0.087 (0.981)
Prior score in ELA (standardized)	0.084 (0.957)	0.085 (0.955)	0.071 (0.965)	0.093 (0.954)
Female	0.496	0.494	0.496	0.496
American Indian	0.012	0.013	0.012	0.010*
Asian / Pacific Isl.	0.125	0.118	0.139**	0.123
Black	0.065	0.058	0.078***	0.061
Hispanic	0.143	0.133	0.148**	0.150**
White	0.592	0.622	0.565***	0.584***
Learning disability	0.053	0.053	0.055	0.050
Special Education	0.112	0.112	0.116	0.110
Gifted	0.079	0.074	0.075	0.086
Limited English	0.053	0.045	0.058***	0.057**
Free/Reduced Lunch	0.415	0.399	0.436**	0.414
Unique teachers	1,107	1,107	1,107	1,107
Unique student teachers	1,358	0	1,358	0
Unique students	148,525	51,701	41,305	55,519

Note. ELA = English Language Arts. Summary statistics limited to teachers who hosted at least one student teacher in a grade 4-8 ELA classroom between 2009-10 and 2014-15. P-values from two-sided t-test relative to years before student teaching: +p<0.10; *p<0.05; **p<0.01; ***p<0.001.

Table A2. Classroom-Level Summary Statistics (Cooperating Teachers in Math Only)

	All	Before student teaching year	Student teaching year	After student teaching year
Prior score in math (standardized)	0.203 (0.935)	0.151 (0.895)	0.187 (0.941)	0.258* (0.961)
Prior score in ELA (standardized)	0.194 (0.844)	0.156 (0.827)	0.178 (0.842)	0.238* (0.857)
Female	0.064	0.045	0.067	0.077
American Indian	-0.055	-0.033	-0.033	-0.087*
Asian / Pacific Isl.	0.128	0.091	0.152	0.143
Black	0.025	-0.009	0.066*	0.026
Hispanic	-0.006	0.047	-0.015+	-0.043*
White	-0.054	-0.057	-0.077	-0.035
Learning disability	-0.067	-0.061	-0.056	-0.079
Special Education	-0.162	-0.146	-0.126	-0.2
Gifted	0.036	0.008	0.015	0.074
Limited English	-0.017	-0.024	-0.008	-0.016
Free/Reduced Lunch	-0.081	-0.032	-0.048	-0.146*
Unique teachers	1,086	1,086	1,086	1,086
Unique student teachers	1,319	0	1,319	0
Unique students	172,665	58,338	45,063	69,264

Note. ELA = English Language Arts. All variables collapsed to the classroom level and then standardized within years. Summary statistics limited to teachers who hosted at least one student teacher in a grade 4-8 math classroom between 2009-10 and 2014-15. P-values from two-sided t-test relative to years before student teaching: +p<0.10; *p<0.05; **p<0.01; ***p<0.001.

Table A3. Regression to the Mean Falsification

	Math		ELA	
	1	2	3	4
Hosted Student Teacher	0.015 (0.015)	0.052*** (0.016)	0.069*** (0.016)	0.045** (0.017)
Hosted Student Teacher * Q2 Value Added	0.011 (0.022)	-0.054** (0.023)	-0.029 (0.020)	-0.003 (0.023)
Hosted Student Teacher * Q3 Value Added	-0.039* (0.022)	-0.056** (0.022)	-0.064*** (0.020)	-0.061** (0.023)
Hosted Student Teacher * Q4 Value Added	-0.085*** (0.022)	-0.105*** (0.025)	-0.124*** (0.021)	-0.099*** (0.022)
Student teaching	Real	Placebo	Real	Placebo
Sample Years	2010-2015	2010-2015	2010-2015	2010-2015
VAM Years	Leave-one-out	Leave-one-out	Leave-one-out	Leave-one-out
Model	Tch FE	Tch FE	Tch FE	Tch FE

Note: ELA = English Language Arts; Tch FE = teacher fixed effect. VAM = value added model. All models include indicators of annual teacher experience and the school year, and also control for the following student control variables interacted by grade: prior performance in math and reading, gender, race/ethnicity, receipt of free or reduced priced lunch, special education status and disability type, Limited English Proficiency indicator, migrant indicator, and homeless indicator. Standard errors clustered at the teacher level are in parentheses. P-values from two-sided t-test: +p<0.10; *p<0.05; **p<0.01; ***p<0.001.