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

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

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

CALDER

AIR

Effective Like Me? Does Having a More Productive Mentor Improve the Productivity of Mentees?

Dan Goldhaber

John Krieg

Roddy Theobald

WORKING PAPER No. 208-1118-1 • November 2018

Effective Like Me? Does Having a More Productive Mentor Improve the Productivity of Mentees?

Dan Goldhaber
American Institutes for Research/CALDER
University of Washington

John Krieg
Western Washington University

Roddy Theobald
American Institutes for Research/CALDER

Contents

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.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A180023 to the American Institutes for Research. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. This research 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.

The collection and cleaning of the TELC data was funded by the Bill and Melinda Gates Foundation (grant #OPP1128040) and an anonymous foundation. Finally, we wish to thank 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 • American Institutes for Research
1000 Thomas Jefferson Street NW, Washington, DC 20007
202-403-5796 • www.caldercenter.org

Effective Like Me? Does Having a More Productive Mentor Improve the Productivity of Mentees?

Dan Goldhaber, John Krieg, Roddy Theobald
CALDER Working Paper No. 208-1118-1
November 2018

Abstract

We use a novel database of the preservice apprenticeships ("student teaching placements") of teachers in Washington State to investigate the relationship between mentor effectiveness (as measured by value added) and the future effectiveness of their mentees. We find a strong, positive relationship between the effectiveness of a teacher's mentor and their own effectiveness in math and a more modest relationship in English Language Arts. The relationship in math is strongest early in a teacher's career, decays significantly over time, and would be positive and statistically significant even in the presence of nonrandom sorting on unobservables of the same magnitude as the sorting on observables. Put together, this suggests that at least some of this relationship reflects a causal relationship between mentor effectiveness and the future effectiveness of their mentees in math.

1. Introduction

Does working with a more effective mentor improve the productivity of mentees? This very basic question has received little empirical attention despite the prevalence of mentoring across a variety of educational and occupational settings. A 2002 publication by the Bureau of Labor Statistics, for instance, reports that there are over 800 apprenticeable occupations (Crosby, 2002). Formalized apprenticeships in which prospective labor market participants are mentored as part of their preparation are an occupational licensing requirement prior to workforce entry in large occupations such as nursing, medicine, clinical social work, and teaching (Bureau of Labor Statistics, 2018). There are about half a million individuals being trained each year in these occupations alone.[1] In a study of occupational licensing, Kleiner and Krueger (2013) report that nearly 30% of employees were licensed and that about half of these require apprenticeships. Thus, it is clear that knowing more about what constitutes a high-quality apprenticeship should inform the training of a large segment of the U.S. workforce.

This paper explores whether a key aspect of apprenticeships, the effectiveness of the mentor who supervises the apprenticeship, is predictive of the labor market productivity of mentees. We use a novel database of the preservice apprenticeships ("student teaching placements") of teachers in Washington State to address the question of whether assignment to a more effective mentor teacher during these apprenticeships impacts the effectiveness of student teachers who become teachers themselves. There are a number of reasons to focus on the connection between mentor and mentee productivity in the case of teaching. First, teachers are

[1]This includes about 175,000 to 300,000 teacher candidates (Cowan et al., 2016); about 8,000 to 19,000 medical school graduates (American Association of Medical Colleges, 2017), 60,000 to 155,000 nursing graduates (U.S. Department of Health and Human Services, 2014), and nearly 25,000 Masters of Social Work (MSW) graduates (Council on Social Work Education, 2015).

the single largest college-educated profession—there are over three million public school teachers—and education is a major industry, with K–12 public school education expenditures in the United States comprising approximately 4% of GDP. Teachers have also been shown to play a critical role in the creation of future human capital.[2] Finally, and importantly for the purposes of the study, there is a well-established measure of labor market productivity for teachers—the "value added" that teachers contribute toward student achievement test scores (discussed more extensively below)—permitting a direct link between the productivity of mentors and mentees.[3]

We find evidence of a strong and positive relationship between value-added measures of mentor effectiveness and mentees' value-added effectiveness in math and more modest relationships in English Language Arts (ELA). Specifically, across a variety of specifications, apprenticing with mentors whose value added is one standard deviation higher is associated with roughly 10–20% of a standard deviation higher value added of mentees in math and (an inconsistently statistically significant) 5–12% of a standard deviation higher value added in ELA.[4] The increase in math value added associated with a one standard deviation increase in mentor quality is roughly equivalent to the difference in average value added between a second-year and novice teacher; in other words, the expected gain in teacher effectiveness from

[2] Differences between teachers are estimated to account for 7–10% in the overall variation in student test achievement (Goldhaber et al., 1999; Nye et al., 2004; Rivkin et al., 2005), and these differences are found to have important impacts on student test scores (Aaronson et al., 2007; Goldhaber et al., 2013).

[3] Worker productivity clearly depends not only on individual human capital contributions but also on other forms of human capital, but teachers are arguably more isolated from other factors of production than are many other professionals, making the link between mentor and mentee productivity more meaningful. Studies of individual and team production (e.g., Jackson & Bruegmann, 2009) find some evidence of value-added spillover effects perhaps due to peer learning, but these are relatively small, and the empirical evidence of the portability of value added across contexts (grades and schools) also suggests limited team production (Bacher-Hicks et al., 2014; Chetty et al., 2014a).

[4] The estimated relationships in ELA are comparable in magnitude to those found in Tennessee by Ronfeldt et al. (2018b), while the estimated relationships in math are considerably stronger.

assignment to a more effective mentor is equivalent to the well-documented returns to the first year of teaching experience (e.g., Ladd & Sorensen, 2017; Rivkin et al., 2005; Rockoff, 2004).

Our findings are robust to the inclusion of various in-service measures of mentor quality (e.g., experience and degree level) and preservice measures of mentee quality (e.g., teacher preparation program and licensure test scores), but there are several potential threats to the causal interpretation of the above estimates. Most importantly, prior quantitative (Krieg et al., 2016) and qualitative (St. John et al., 2018) evidence from Washington State (the setting of this study) documents considerable nonrandom sorting of teacher candidates to mentor teachers. While we can account for sorting along observable dimensions—for example, the sorting of candidates with higher licensure test scores to mentors with higher value added documented in Krieg et al. (2016, 2018)—it is plausible that positive sorting of teacher candidates who *already would be more effective teachers* to more effective mentors *along unobserved dimensions* may explain at least some of the estimated relationships discussed above.

We address this threat to validity in two ways. First, as we argue more extensively below, if the estimated relationships between mentor effectiveness and future teacher effectiveness are driven by nonrandom sorting of candidates to mentor teachers on the basis of an unobserved, time-invariant measure of candidate productivity, we would expect the resulting bias in our estimates to persist throughout a teacher's career. Instead, we find that the relationship between mentor effectiveness and future teacher effectiveness in math is strongest in a teacher's first year of teaching and decays significantly over a teacher's early-career teaching experience. Second, we follow the approach of Altonji et al. (2005, 2008) and Oster (2017) and calculate that, even if the amount of sorting to more effective mentor teachers along unobserved dimensions is of the same magnitude as the observed sorting to more effective mentors explained by our extensive set of control variables, the relationship between mentor effectiveness and the future effectiveness of their mentees in math would still be positive and statistically significant. Taken together, this suggests that at least some of the estimated relationship between mentor and mentee effectiveness can be explained by more effective mentors having positive causal impacts on the future effectiveness of their mentees.

2. Background Literature on Mentoring and Student Teaching

Mentoring proliferates across a variety of contexts, spanning different occupations and educational and career levels. It serves a variety of purposes: to pass on key skills from mentor to mentee; to engage students and raise their educational and career expectations; and to affect attitudes, expectations, and behaviors toward schooling or jobs. Given the divergent purposes for which mentoring is utilized, it is not surprising that the nature of mentoring relationships and the context in which mentoring occurs are quite varied.[5] Eby et al. (2007) argue that there are three distinct areas of scholarship on mentoring: youth mentoring, academic mentoring, and workplace mentoring. And in their meta-analysis they find positive effects of all three types of mentoring on schooling, behavioral, attitudinal, health, and job/career outcomes.

Here we are focused on workplace mentoring. While there is no formal definition of precisely what this entails, it often is characterized as a hierarchical relationship in which the mentor is more experienced than the mentee and has useful knowledge and skills that can be conveyed to the mentee through role modeling, feedback, and support (Ambrosetti & Dekkers,

[5] It may, for instance, be adults mentoring children or students, peer to peer, or senior to junior in a particular occupation or job. And mentoring occurs informally and through formalized programs. Because of the varied contexts and ways in which mentoring occurs, it is often difficult to distinguish mentoring from more general types of job training and socialization. For more on this and the theory behind different types of mentoring, see Bozeman and Feeney (2007).

2010). But while there are hundreds of studies on the potential and self-reported (e.g., Aryee et al., 1996) benefits of being mentored, the empirical evidence connecting workplace mentorship of early-career employees to their later labor market outcomes is much scarcer.

Rockoff (2008) takes advantage of the implementation of a mandatory teacher mentoring program in New York City (NYC) in 2004 to study the effects of mentoring on teacher retention and student achievement. He exploits the fact that teachers hired into NYC with prior experience were much less likely to be assigned a mentor than novices to implement a difference-in-difference identification strategy and finds that mentoring has little impact on teacher absences, retention, or student achievement.[6] More recently, Papay et al. (2016) find more encouraging evidence based on an experiment in which high- and low-performing teachers in a randomized set of schools are paired with one another. They find larger student test score gains in the schools with the pairing treatment relative to schools in the control group and particularly large gains in the lower performing teachers' classrooms, suggesting that assignment to an effective partner teacher can impact teacher productivity.

In the case of some occupations, mentoring is either strongly encouraged or a requirement for occupational licensure (i.e., mentoring that occurs prior to entering an occupation). Research on this type of preservice mentoring generally shows positive mentoring effects. For instance, Stamm and Beddeberg-Fischer (2011) find that medical residents who receive mentoring during medical residency, either in the form of a mentoring relationship with a single physician or through participation in a mentoring support network, have higher measures of both objective (e.g., salary) and subjective (e.g., self-reported satisfaction) success in their future careers. However, this study is representative of the broader mentorship literature

[6] Rockoff does find, consistent with Smith and Ingersoll (2004), that some measures of mentor quality (e.g., prior mentor experience in the same school as a mentee) do predict mentee retention.

discussed above and summarized in Ely et al. (2007), as it focuses on the *presence* or *style* of mentoring, as opposed to investigating any characteristics of the mentor.

Mentoring is also an important ingredient in teacher preparation. Indeed, apprenticeships with mentored clinical experiences for teacher candidates are characterized as "a key component—even 'the most important' component of—pre-service teacher preparation" (Anderson & Stillman, 2013, p. 3) and they are required for traditional teacher licensure (Goldhaber et al., 2014).[7] There is a widespread belief that mentors "influence the career trajectory of beginning teachers for years to come" (Ganser, 2002, p. 380). The mentor teacher (also often referred to as the "cooperating teacher" in Washington State, the setting for this study) is a K–12 teacher who hosts a mentee (or "teacher candidate") as they take on some or all of lead teaching responsibilities.

There is a large theoretical and case study literature describing the role of mentor teachers in the development of teacher candidates. This suggests that mentors serve as models of instructional effectiveness, providing feedback and support to teacher candidates who are just learning to practice their craft (e.g., Ambrosetti & Dekkers, 2010; Grossman et al., 2014; Schwille, 2008; Yendol-Hoppey, 2007; Zeichner & Gore, 1990). Some also argue that mentors help to prepare teacher candidates for the realities of K–12 classrooms, which may be different from the expectations set up in their teacher education programs (Hargreaves & Jacka, 1995).

There is relatively little quantitative evidence about the relationship between mentor teachers and later teacher candidate performance. Matsko et al. (forthcoming) find positive

[7] Student teaching generally occurs in the last year of a teacher candidate's teacher education experience. States sometimes require mentors to have a minimum level of teaching experience and, occasionally, a minimum performance evaluation; generally, however, states provide little specific guidance about who should serve as a mentor (Greenberg et al., 2011, 2013). States also have other preservice requirements associated with licensure, such as passing various licensure tests (Goldhaber, 2007).

correlations between teacher candidates' feelings of preparedness and both their reports of the instructional quality of their mentor teachers as well as the performance evaluations (observational ratings) their mentor teachers receive. Similarly, Ronfeldt et al. (2018a, 2018b) also find positive correlations between the observational ratings of mentor teachers and the teacher candidates they mentor who eventually become teachers. These studies certainly support the notion that the quality of a mentor affects the later performance of their mentees, but they are also limited by the subjective measure of observational ratings. Observational ratings have been shown to vary considerably from one district to another in ways that do not reflect differences in teacher quality across districts (Cowan et al., 2018), and since teacher candidates tend to find jobs in the school districts in which they completed their student teaching (Krieg et al., 2016, 2018), the positive correlations between mentor and mentee observation ratings could simply be an artifact of the rigor of school district ratings. Moreover, observational ratings tend to be only weakly related to student achievement (Blazar, 2015; Cowan et al., 2018; Kane et al., 2013).[8]

We are only aware of one published study relating the productivity of mentors and mentees using an objective measure of productivity. As in this study, Ronfeldt et al. (2018a) assess whether having a more effective mentor teacher (i.e., having higher value added) is associated with the later effectiveness of those mentees. They find that a one standard deviation increase in the effectiveness of the mentor is associated with about a 5% of a standard deviation increase in the value added of mentees who enter the profession in value-added grades and subjects.[9]

[8] Kane et al. (2013), for instance, use data from the Measures of Effective Teaching study (in which teachers were randomly assigned to classrooms within schools and grades) and find that a 1-point increase in a teacher's classroom observation score is correlated with about a 0.10 standard deviation increase in student performance.
[9] The authors report that this is about a third of the estimated return to the first year of teaching experience.

The analysis in Ronfeldt et al. (2018a) provides important direct evidence connecting mentors to the students of their mentees but is also somewhat hampered by data limitations. In particular, the relatively short time panel of student teaching apprenticeships and mentee outcome years necessitates that Ronfeldt et al. consider value-added measures *from the year the teacher hosted the mentee* as predictors of the mentee's future value added. As we discuss in more detail in Section 4, this raises questions about whether the apprenticeship or the specific mentee are contributing to these measures of mentor value added. The short panel also means that Ronfeldt et al. are unable to explore the persistence of these relationships as mentees remain in the teacher workforce.[10]

Thus beyond the utility of investigating this same question in a different context, our analysis—based on nine years of student teaching data and student-level achievement data—is able to build substantially on this prior analysis by (a) considering a measure of mentor quality that is calculated from student-level data *entirely from years prior to the apprenticeship*, (b) estimating the persistence of the relationships between mentor and mentee effectiveness as mentees gain experience in the teaching workforce, and (c) evaluating the sensitivity of our estimates of these relationships under different assumptions about the nonrandom sorting of mentees to mentors and K–12 students to different classrooms. In the next section, we describe the unique dataset of student teaching placements that allows us to build on this prior research base.

[10] It is also worth noting that the measures of both mentor and mentee value added come from the Tennessee Value-Added System (TVAAS), and methodological issues have been raised by researchers (e.g., Ballou & Springer, 2015; Vosters et al., 2018) about various aspects of the way TVAAS works.

3. Data and Setting

For this research we combine data from Washington State's Office of the Superintendent of Public Instruction (OSPI) on in-service public school teachers and students with longitudinal data on student teaching apprenticeships provided by a group of 15 teacher education programs in Washington State that are participating in the Teacher Education Learning Collaborative (TELC).[11] The OSPI data include annual student test scores (for Grades 3–8) in reading and math as well as student demographic and program participation data for all K–12 students in the state. From 2006–07 through 2008–09, students in Grades 3–5 can be linked to their classroom teacher by their proctor on the state exam.[12] From 2009–10 through the most recent year of available data, 2016–17, the state's CEDARS data system allows students to be linked to their classroom teachers through unique course IDs.[13] Because we are estimating value-added models (described in more detail below) that require student-teacher links and both current and prior-year test scores, we limit the sample of in-service teachers (both mentor and mentees) to those who teach self-contained classes in Grades 4–5 between 2006–07 and 2016–17 and in math or reading in Grades 6–8 between 2009–10 and 2016–17.

The OSPI data can be linked to the TELC dataset through unique teacher IDs for both the mentor teacher and mentee of each student teaching placement. Specifically, the TELC data

[11] 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 six 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.

[12] 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.

[13] CEDARS data include fields designed to link students to their individual teachers, based on reported schedules. However, limitations of reporting standards and practices across the state may result in ambiguities or inaccuracies around these links.

include information on the mentor teacher who supervises the apprenticeship of teacher candidates for the 15 TELC programs, and data on teacher candidates can be linked with the state's teaching credential database that permits further connection to data on in-service teachers in the OSPI data.[14] The most recent year of TELC data is 2015–16 but the earliest years of data from each program in the TELC dataset vary, with some programs providing data on apprenticeships that date back to the late 1990s. We focus on nine years of student teaching data (2007–08 through 2015–16) because candidates in these years can be assigned to a mentor teacher with a prior measure of value added (i.e., 2006–07 is the first year in which value added can be calculated in Washington, so mentor teachers in 2006–07 and earlier cannot have a measure of prior value added).

The OSPI data also include other measures of the background and credentials of both mentors and mentees, including information on teachers' years of teaching experience; degree level (e.g., bachelor's or master's); teacher preparation institution; teaching endorsement areas; licensure test performance on the Washington Educator Skills Tests – Basic (WEST-B) in math, reading, and writing; and the institution from which they graduated. Because the state accepts a number of alternative tests that meet the WEST-B testing requirement for receiving a teaching credential, only 82% of mentees in the data have valid WEST-B scores.[15] Moreover, since the WEST-B has only been a licensure requirement since 2002, scores are missing for most of the

[14] 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. We also match 72% of these teacher candidates to observations on their in-service teaching positions in the OSPI data; the 28% of candidates who do not enter the workforce include candidates who teach in private schools or out-of-state, never become a teacher, or who are not successfully matched with the OSPI data (e.g., because of a name change between student teaching and the first teaching position).
[15] Passing scores for Praxis I, California Basic Educational Skills Test (CBEST), or the Pearson NES Essential Academic Skills test, as well as scores on the SAT and ACT above certain cutoffs (e.g., 515 on the math SAT) can be submitted as alternatives to the WEST-B exam (RCW 28A.410.220 & WAC 181-01-002).

(relatively more experienced) mentor teachers in the sample, though 15% of mentor teachers can be linked to these licensure test scores.

The merged dataset includes 1,044 mentee observations in math (with 924 unique mentors; mentors supervise apprenticeships an average of 1.12 times in our data) and 944 mentee observations in ELA, all of whom are linked both with a prior measure of mentor value added and with student test scores in an in-service teaching position. In all, we have 2,534 mentee-year observations linked to 78,458 student observations in math and 2,423 mentee-year observations linked to 65,632 student observations in ELA.

Table 1 provides selected summary statistics for mentors and mentees in this dataset. We provide overall summary statistics in columns 1 and 5 (for math and ELA, respectively), and for the top, middle two, and bottom quartiles. Testing the means in the top and bottom quartile against the middle two shows no more statistically-significant differences than we would expect by chance, providing cursory evidence that there are not strong mentor and mentee matching patterns, at least based on observable characteristics.

Table 2 repeats this exercise for the student characteristics that will be the control variables in the models described in the next section. Here we see more nonrandom sorting of students to classrooms by the value added of the teacher's mentor. For example, student teachers whose mentor is in the top quartile of ELA value added tend to have considerably higher performing students once they enter the workforce than student teachers whose mentor is in the bottom quartile of mentor value added. These differences may be driven by two factors documented in prior work in Washington State: Both student teaching placements and teacher hiring tend to be very localized (Goldhaber et al., 2014, 2017b; Krieg et al., 2016, 2018), and there are significant differences in both student performance and average teacher value added across districts in the state (Goldhaber et al., 2015, 2018b). Put together, this suggests that teachers in some parts of the state are both more likely to be assigned to an effective mentor teacher and more likely to enter a classroom with high-achieving students than teachers in other parts of the state. The analytic models described in the next section account for this nonrandom sorting along observable dimensions, and we also describe in the next section a robustness check that uses the nonrandom sorting by observable variables documented here as a proxy for the amount on nonrandom sorting we might expect on unobservable dimensions.

4. Empirical Strategy

Central to our study is the need to obtain unbiased measures of the productivity of both mentor teachers and their mentees. A significant literature investigating teachers is devoted to assessing the impacts of individual teachers on students (e.g., Aaronson et al., 2007; Chetty et al., 2014a; Rivken et al., 2005) as well as the extent to which value-added models (VAMs) can be used to obtain unbiased estimates of the contribution of individual teachers to student test score gains (Bacher-Hicks et al., 2014; Chetty et al., 2014b; Goldhaber & Chaplin, 2014; Kane & Staiger, 2008; Kane et al., 2013; Rothstein, 2009, 2014). While this issue is not settled,[16] we argue that appropriately specified VAMs show minimal bias (Koedel et al., 2015), especially in estimating teacher effectiveness in math.[17]

[16] See, for instance, the debate between Chetty et al. (2014a, 2016) and Rothstein (2014).
[17] Kane et al. (2013) show that value-added estimates produce nearly unbiased predictions of student achievement differences when classrooms are randomly assigned to teachers within schools, and Chetty et al. (2014a) show that the changes in out-of-sample value added at the grade-school level associated with teachers switching grades and schools is an unbiased predictor of changes in student achievement in those grades and schools.

To make our calculations concrete, define t_{jk}^{M} as the year that teacher *j* serves as a mentor to a mentee *k*. The measure of mentor value added that we use in our subsequent models is calculated from the following VAM specification (we also test variants of this):

$$Y_{ijst'} = \alpha_0 + \alpha_1 Y_{i(t'-1)} + \alpha_2 X_{it'} + \sum_k \alpha_{k+3} I(Exp_{jt'} = k) + \tau^{M}_{jks(t'<t^{M}_{jk})} + \varepsilon_{ijst'} \quad (1)$$

In (1), $Y_{ijst'}$ is the state test score for each student *i* with teacher *j* in subject *s* (math or reading) and year *t'*, normalized within grade and year; $Y_{i(t'-1)}$ is a vector of student *i*'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'* (gender, race, FRL status, English language learner status, gifted status, special education status, learning disability status); and $Exp_{jt'}$; is the experience of teacher *j* in year *t'* (included as indicators for different years of teaching experience).[18] The estimate of mentor value added that we use for mentee *k*, $\tau^{M}_{jks(t'<t^{M}_{jk})}$, represents the contribution of teacher *j* to student test scores in subject *s for all years prior to the student teaching placement*. We shrink these estimates using Empirical Bayes methods that remove some measurement error in these estimates.[19]

The use of a prior measure of mentor value added is motivated by potential endogeneity concerns. As mentioned in the context of Ronfeldt et al. (2018a), there is a possibility that hosting a mentee impacts teacher value added in the year of their apprenticeship. Goldhaber et al. (2018a) show that, while there is no detectable *average* effect of hosting a student teacher on

[18] Note that our inclusion of teacher experience indicators means that our measures of mentor value added control for the experience of the mentor teacher. We include mentor teacher experience as an additional control in the second-stage models described below.
[19] 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.

student achievement in the student teaching ("host") classroom, there are large negative effects for host classrooms in which the mentor is in the lowest quartile of value added, which would suggest a downward bias in the Ronfeldt et al. estimates of the relationship between mentor and mentee effectiveness. Finally, Goldhaber et al. (2018a) show that hosting a student teacher has a positive impact on the *mentor's* teaching effectiveness in later years, so we also do not use data from years following the placement as these estimates appear to be impacted by the apprenticeship itself.

Ultimately, we are interested in understanding the impact that effective mentor teachers have on their mentees after these mentees enter classrooms of their own. To investigate this, we estimate models predicting student achievement *in the classrooms of mentee k once they enter the workforce*. We therefore estimate variants of the following model predicting student performance in the classroom of mentee *k* as a function of the estimated value added of mentor *j* (calculated from equation 1) and the same set of controls:

$$Y_{ikst} = \beta_0 + \beta_1 Y_{i(t-1)} + \beta_2 X_{it} + \sum_k^4 \beta_{k+3} I(Exp_{kt} = k) + \beta_8 \hat{\tau}^M_{jks(t'<t^M_{jk}\}} + \varepsilon_{ijst} \qquad (2)$$

The variables in equation 2 are defined the same as above, and the coefficient of interest, β_8, represents the relationship between mentor value added and the performance of students in the mentee's classroom.

There are at least four arguments for using caution when interpreting β_8 as the causal impact of a cooperating teacher on the effectiveness of mentees. First, for some student teachers, there is a significant lag between their student teaching experience and our observations of them in the classroom. We hypothesize that time that has elapsed since student teaching can dilute the impact of a mentor. In order to account for this, we explicitly measure the impact of time between the student teaching year t_{jk}^M and the current year *t* and interact the log of this term with mentor value added:[20]

$$Y_{ikst} = \gamma_0 + \gamma_1 Y_{i(t-1)} + \gamma_2 X_{it} + \sum_k^4 \gamma_{k+3} I(Exp_{kt} = k) + \gamma_8 \hat{\tau}_{jks(t'<t_{jk}^M\}}^M + \gamma_9 \log(t - t_{jk}^M) + \gamma_{10} \log(t - t_{jk}^M) * \hat{\tau}_{jks(t'<t_{jk}^M\}}^M + \varepsilon_{ijst} \quad (3)$$

In the specification in equation 3, γ_8 represents the relationship between mentor value added and student achievement the year immediately following student teaching (i.e., when $t - t_{jk}^M = 1 \Rightarrow \log(t - t_{jk}^M) = 0$). The parameter γ_9 represents the relationship between the time since student teaching and student achievement (conditional on return to teaching experience; i.e., this term is identified exclusively by teachers with a delay between student teaching and the first time they are observed in classrooms), while γ_{10} captures the rate at which the relationship between mentor value added and student achievement decays as the time since student teaching increases.

Another advantage of this decay specification is that it serves as one check of a second threat to the validity of our results: that more effective mentees systematically sort to more effective mentors. Specifically, suppose there is a time-invariant, unobserved variable for mentee *k*, θ_k, that is correlated both with mentor value added (so $Cov\left(\theta_k, \hat{\tau}_{jks(t'<t_{jk}^M\}}^M\right) > 0$) and future teacher effectiveness (so $Cov\left(\theta_k, \varepsilon_{ijst}\right) > 0$). This omitted variable basis clearly results in a positive bias in the estimated coefficients $\hat{\beta}_8$ and $\hat{\gamma}_8$ in equations 2 and 3, respectively. However, if θ_k (or any other time-invariant confounders) is the *only* source of confounding in equations 2 and 3 *and* if the true value of the parameter γ_8 in equation 3 is zero, we would expect to see that

[20] We selected the log specification through a model selection procedure in which we compared the BIC between models with linear and polynomial terms of $(t - t_{jk}^M)$, as well as a formal exponential decay model used to model decay in teacher preparation program effects in prior work (Goldhaber et al., 2013).

$\hat{\gamma}_{10} = 0$, as there would be no reason for the relationship between mentor value added and student achievement to change over time.

The specific confounding story described above (in which a time-invariant variable is the *only* source of confounding) is probably implausible, so we pursue a number of additional extensions to minimize this source of bias. First, while the models in equations 2 and 3 include a rich set of variables intended to control for potential of bias, it is possible that nonrandom sorting of mentees remain a threat to causal interpretations. For instance, if high-ability student teachers (as proxied by their licensure test scores) are supervised by mentors with higher value added, a finding corroborated by Krieg et al. (2016, 2018), then $\hat{\beta}_8$ and $\hat{\gamma}_8$ in equations 2 and 3 will be biased upward.[21] We attempt to minimize this bias by adding a number of controls that come in three types: mentor controls, mentee controls, and district fixed effects. The mentor and mentee controls include: licensure test scores in math, reading, and writing; indicators for the teacher education program attended; an indicator for whether the teacher has a master's degree; and indicators for subject endorsement areas. In addition, the mentor controls include the years of teaching experience at the time the mentor hosted the student teacher. If mentees are sorted to mentors based upon these characteristics, then including them eliminates any bias caused by that selection.

In addition, in some specifications we include two types of district fixed effects: fixed effects for the district in which the student teaching took place and fixed effects for the district where the mentee is observed earning their value added. The student teaching fixed effects control for the sorting of mentees to student teaching placements and teaching districts, which is important because of prior evidence linking the location of teacher education programs to job

[21] This holds because prior work (e.g. Goldhaber et al., 2017a) finds modest relationships between the performance of teachers on licensure tests and their value added.

placement (Krieg et al., 2016, 2018) and to districts where teacher trainees grew up (Boyd et al., 2005). The district fixed effects for the mentees' current teaching positions control for the large differences in teacher value added across different districts in Washington State documented in prior work on "teacher quality gaps" (Goldhaber et al., 2015, 2018b).

Another reason to be wary about interpreting $\hat{\beta}_8$ and $\hat{\gamma}_8$ as causal effects has to do with the possibility that mentors influence the workforce participation of mentees. For instance, a more effective mentor may increase the likelihood that teacher candidates with different unobserved teaching capacities pursue a teaching career. Since we only observe outcome measures for student teachers who enter teaching, if this selection issue exists, then we are more likely to observe student teachers with more effective mentors. This is mitigated by controlling for the observables described above, and we investigate this issue further by estimating a logit model that predicts workforce entry based upon mentor's value added and find no significant relationships. However, it is still possible that effective mentors have *differential* impacts on the workforce entry by mentee experience—that is, more effective mentors may make their effective mentees more likely to enter the workforce and their less effective mentees less likely—and given that we do not observe a direct measure of mentee productivity prior to student teaching, we cannot test this possibility directly. Thus, the estimated relationship between mentor value added and future mentee effectiveness likely includes both within-mentee effects (i.e., changes of productivity due to working with a more effective mentor) and cross-mentee effects due to any differential impacts of mentors on the workforce entry of their mentees. Given that 72% of student teachers entered the labor force in our sample, we expect that the impact of this type of bias is likely small.

A related concern has to do with attrition of mentees from the sample. If more effective teachers who were supervised by more effective mentees are differentially likely to leave the workforce, this would also bias our estimates $\hat{\beta}_8$ and $\hat{\gamma}_8$. We are able to test for this possibility directly by estimating models predicting attrition of mentee *k* from the sample in subject *s* year *t*, A_{kst}, as a function of their effectiveness, their mentor's effectiveness, and the interaction:

$$logit(A_{kst}) = \omega_0 + \omega_1 \hat{\tau}^M_{jks\left(t'<t^M_{jk}\right\}} + \omega_2 \hat{\tau}_{ks(t'\leq t\}} + \omega_3 \hat{\tau}^M_{jks\left(t'<t^M_{jk}\right\}} * \hat{\tau}_{ks(t'\leq t\}}$$

$$+ \sum_k^4 \omega_{k+3} I(Exp_{kt} = k) \qquad (4)$$

If mentees of different effectiveness leave teaching and this decision is connected to their mentor's effectiveness, then we would observe $\omega_3 \neq 0$, something we test for in the next section. However, as a simpler way to remove this potential source of bias we also estimate variants of the models in equations 2 and 3 only for first-year teachers (i.e., before they are able to leave the workforce).

Of the four potential sources of bias described above, despite the extensive controls and number robustness checks described above, we are primarily concerned about the potential nonrandom sorting of more effective mentors to more effective mentees and, by extension, the students in their mentees' future classrooms along unobserved dimensions. We therefore pursue one additional extension to quantify the potential implications of this source of bias. Specifically, we follow Oster (2017), who extends the work of Altonji et al. (2005, 2008) on identifying the extent of bias caused by selection on unobservables.

Under this methodology, let W_{ijkst} capture *all unobserved variables* that are correlated with the value added of mentor *j,* $\hat{\tau}^M_{jks(t'<t^M_{jk}\}}$ and student performance in the classroom of mentee *k,* Y_{ikst}. Further define δ as the magnitude of sorting on W_{ijkst} relative to sorting on *all* the

observable variables V_{ijkst} in equation 2 (formally, $\delta \frac{\sigma_{V\tau}}{\sigma_V^2} = \frac{\sigma_{W\tau}}{\sigma_W^2}$, where $\sigma_{V\tau} = Cov\left(V_{ijkst}, \hat{\tau}^M_{jks(t'<t^M_{jk}\}}\right)$, $\sigma_{W\tau} = Cov\left(W_{ijkst}, \hat{\tau}^M_{jks(t'<t^M_{jk}\}}\right)$). Oster (2017) derives that, under some restrictive assumptions, the adjusted value of $\hat{\beta}_8^*$ in equation 2—that is, the value of $\hat{\beta}_8$ we would have estimated *if we had been able to control for* W_{ijkst}—can be calculated as a function of the estimate $\hat{\beta}^0$ and the R-squared of a null model regressing Y_{ikst} against only $\hat{\tau}^M_{jks(t'<t^M_{jk}\}}$, R^0, the observed estimate $\hat{\beta}_8$ and the R-squared of the model in equation 2, $\tilde{R}$, and the maximum possible R-squared from a model predicting Y_{ikst}, R_{max}:[22]

$$\hat{\beta}_8^* \approx \hat{\beta}_8 - \delta\left(\frac{R_{max}-\tilde{R}}{\tilde{R}-R^0}\right)\left(\hat{\beta}^0 - \hat{\beta}_8\right) \qquad (5)$$

We choose the values $R_{max} = \delta = 1$—that is, that the outcome could (theoretically) be perfectly predicted and the amount of sorting on unobservables is equivalent to the amount of sorting on observables—and use equation 5 to derive the adjusted estimate $\hat{\beta}_8^*$. We then bootstrap standard errors for $\hat{\beta}_8^*$ to test whether the estimated relationship between mentor value added and student performance would still be statistically significant if we had been able to control for W_{ijkst} under this scenario. More intuitively, this approach tests whether the estimated relationship between mentor value added and student performance could be "explained away" by this amount of sorting on unobservables.[23]

[22] The most restrictive of these assumptions is that the relative contribution of each variable to Y_{ikst} must be the same as their contribution to $\hat{\tau}^M_{jks(t'<t^M_{jk}\}}$.

[23] Our implementation of this procedure is very conservative both in our choice of R_{max} and δ and in our use of *all* covariates (student, mentor, and mentee) in calculating the amount of sorting on observables. An alternative is to consider just mentee characteristics in the vector of observable variables V_{ijkst}—that is, to account directly for our concern that more effective mentees sort to more effective mentors—but this procedure actually results in a larger adjusted estimate due to the somewhat negative sorting of mentees to mentors along observable dimensions (as can be seen in Table 1).

5. Results

In this section we describe the relationships between mentor and mentee effectiveness, but prior to focusing on the main findings of interest, a few peripheral findings warrant brief notice. In both the math and ELA samples we observe that black students, participants in the free and reduced-price lunch program, and/or those in special education score lower than their reference groups (the coefficients for these student level control variables are reported in Table A1 in the appendix for the model specifications that are reported in Table 2). All of these findings are quite consistent with the broader literature.[24] We also see evidence of returns to teaching experience; for example, students with a teacher who has 3 or more years of experience outperform students with novice teachers by about 5% of a standard deviation in both math and ELA (the estimates for all the mentor and mentee characteristics can be found in Table A2 in the appendix).

5.1 Primary Findings

Table 3 shows the primary relationships of interest between mentor effectiveness and the later effectiveness of their mentees. We begin in column 1 (for math) and column 6 (for ELA) with a sparse model that omits controls for mentors other than their value added and also does not include measures of mentee quality (i.e., equation 2). In math we see strong evidence that value-added measures of mentor effectiveness are related to mentees' value-added effectiveness; a one standard deviation increase in mentor effectiveness is associated with a 12% of a standard deviation increase of the effectiveness of their mentees; this is roughly half of the difference between a novice teacher and one with at least three years of experience (see Appendix Table

[24] For instance, see Aaronson et al. (2007) and Rivkin et al. (2005).

A2) and about twice as large as the comparable estimate in Ronfeldt et al. (2018a).[25] The estimated relationship from the specification in ELA is not statistically significant (though only slightly smaller than the estimated magnitude of the comparable relationship in Ronfeldt et al., 2018a).

In columns 2 and 7 we add controls and interactions for the (logged) amount of time elapsed since student teaching (i.e., the specification in equation 3 of the previous section) to account for the possibility that the effects of working with more effective mentors decay over time. The coefficient on the interaction between time and mentor value added is marginally significant and negative, suggesting that the magnitude of the relationship between mentor and mentee value added does decrease over time. The magnitude of the interaction effect suggests that the relationship between mentor and mentee value added in the first year after student teaching, 0.190, disappears entirely by a teacher's 10th year, a period beyond the range of our observed data, so we simply conclude this relationship persists but decays significantly.[26]

This conclusion can be seen visually in Panel A of Figure 1, which plots predicted student achievement from the specification in column 2 of Table 3 for mentees assigned to mentors of different levels of value added and as a function of time since student teaching (and also incorporates expected returns to teaching experience). The differences between mentee effectiveness are considerable the first year after student teaching, and while the lines get closer over time, mentees with more effective mentors are still more effective (all else equal) many years after they enter the workforce. Unfortunately, we cannot determine whether this decay is related to the decay of mentor effects or a mentor's impact on workforce attrition (this is

[25] A 0.12 standard deviation increase in teacher effectiveness is equivalent to approximately a 0.024 standard deviation increase in student performance, while the returns to the first two years of teaching experience is approximately 0.05 standard deviations of student performance in math (see Appendix Table A2).
[26] $0.190 - \log(10) * .082 \approx 0.$

discussed in more detail in Section 5.2). The analogous decay term in ELA is not statistically significant, but accounting for the possibility of decay does produce a marginally statistically-significant relationship between mentor and mentee value for the year immediately following student teaching in ELA. As can be seen in Panel B of Figure 1, though, the magnitudes of these relationships are considerably more modest in ELA than in math.

To explore the potential that the findings on mentor effectiveness are related to other observable mentor characteristics, we add a number of mentor controls in columns 3 and 8. Though these mentor controls explain a significant amount of variation in both math and ELA, there is little change in the estimated coefficients on mentor value added associated with these additions to the model, which is not surprising given that (as can be seen in Appendix Table A2) these mentor characteristics are generally weak predictors of mentee value added.[27] Similarly, in columns 4 and 9, we show the findings when we add analogous controls for preservice mentee quality (including indicators for the institution from which each mentee graduated). There is little change in the coefficients on mentor effectiveness, which again is not surprising given that (as shown in Appendix Table A2) these mentee characteristics are only weakly predictive of mentee value added.[28] This provides some cursory evidence that the results are not related to the nonrandom matching of mentor and mentee quality (at least based on observables).

Finally, there is ample evidence (Boyd et al., 2005; Krieg et al., 2016, 2018; Mihaly et al., 2013) of strong geographic links between teacher education programs, student teaching

[27] An F-test on the mentor controls in math results in an F-statistic of 124.94, while the F-statistic is 15.09 in ELA, both highly statistically significant. Mentor experience is a negative predictor of student performance in math and a positive predictor of student performance in ELA, but the magnitudes of the coefficients are very small (implying in each case that a 10-year increase in mentor experience is correlated with only a 0.02 standard deviation change in student performance). These weak relationships are consistent with Ronfeldt et al. (2018a).

[28] As in prior work in Washington (Goldhaber et al., 2013), some of the institution indicators are statistically significantly different from each other, but these institution indicators explain only about 1% of the variation in mentee value added.

placements, and the likelihood of mentees being employed in particular school systems. As we described in Section 3, this could be another source of nonrandom sorting of more effective mentees to more effective mentors. To account for this possibility, we include (in columns 5 and 10) fixed effects for the school districts in which the apprenticeships and teaching experience took place. In these models the coefficient on mentor effectiveness are being identified based on the within-district variation in both mentor and mentee value added. The estimates from these specifications are slightly more modest, but we still see statistically significant main effects of similar magnitude in both math and ELA.

5.2 Robustness Checks

We now describe the various robustness checks described in Section 4 that explore the implications of the various potential sources of bias in the estimates presented above. In Table 4, we present estimates from the attrition model described in equation 4 that is intended to explore differential sample *attrition* by mentor and mentee value added. While we find some evidence that more effective mentees are more likely to leave the analytic sample, we do not see systematic heterogeneity in this relationship by mentor value added, which suggests that differential attrition from the sample is unlikely to be a major source of bias.

With that said, we still estimate relationships between mentor and mentee value added only for first-year teachers (i.e., observations not impacted by nonrandom attrition) and report the results in Table 5. As expected from the main effects in Table 3, the relationships are all positive and statistically significant in math, whether or not the sample is further restricted to the first year after student teaching. The estimates in ELA are positive but inconsistently significant across specifications. We use the specifications in columns 1 and 7 of Table 5 to further investigate nonlinearities in these relationships by swapping in quartiles of mentor value added for the continuous measure discussed to this point and plot to estimated effects (relative to the lowest quartile of mentor value added) in Figure 2.[29] In both subjects, the positive relationships appear to be driven by mentor teachers in the top quartile of the distribution.

Finally, as described in Section 4, we follow Oster (2017) and test whether the estimated relationship between mentor value added and student performance would still be statistically significant under a hypothetical scenario where the magnitude of nonrandom sorting to more effective mentors along unobserved dimensions is just as large as the observed nonrandom sorting on observed variables. To include the most observable variables possible (and to simplify the interpretation of our estimates), we use the specification in equation 2 (i.e., without decay) but include the full array of mentor and mentee controls in columns 4 and 9 of Table 3. The estimated relationship between mentor value added and mentee value added in math in this model is 0.109, meaning that a standard deviation in mentor value added is correlated with a 0.109 standard deviation increase in mentee value added. Under unobserved sorting of the same magnitude as sorting on observables, the adjusted relationship is 0.031 with a bootstrapped 95% confidence interval of (0.006, 0.055). In other words, even under this extreme example in which there is considerable bias due to unobservable sorting to more effective mentors, the relationship between mentor value added and mentee value added in math is still positive and statistically significant. On the other hand, the estimated relationship in ELA (0.052) is easily "explained away" under this same scenario, with an adjusted estimate of -0.191. Our conclusion from this is that, at least in math, some of the relationship between mentor and mentee quality likely reflects a causal relationship between mentor effectiveness and the future effectiveness of their mentees in math.

[29] See Table A3 in the appendix for the point estimates from these models.

6. Conclusions

First and foremost, this study has clear and direct implications for K–12 education. Despite decades of research and billions of dollars of investment in efforts to enhance the teacher workforce, interventions that improve the productivity of individual teachers are somewhat elusive. Yet one of the most widely acknowledged empirical findings is that individual teachers do improve, as there are well-documented returns to early-career teaching experience. Several states and policymakers have therefore sensibly turned to preservice teacher preparation as one potential way of moving some of these early-career returns to the years before teachers have a classroom of their own.[30]

This study suggests one specific mechanism through which this can occur. In fact, Figure 1 illustrates that first-year teachers who student taught with a highly-effective mentor teacher in math (i.e., 2 standard deviations above the mean) are predicted to be just as effective as third-year teachers who worked with an average mentor. While it is certainly possible that some of these differences reflect the nonrandom sorting of mentees to mentors (and thus reflect cross-mentee differences in effectiveness), the decay in these relationships over time and the robustness of these relationships under extreme sorting on unobservables (in which the relationship is still significant and positive) both suggest that assignment to higher quality mentors induces a causal and within-mentee improvement in quality. Thus, the assignment of

[30] As one specific example, the Massachusetts Department of Elementary and Secondary Education states as a policy goal that "*... by 2022, candidates prepared by Massachusetts' providers will enter classrooms and demonstrate results on par with peers in their third year of teaching.*" http://www.doe.mass.edu/edprep/EPIC/

student teachers to more effective mentor teachers appears to be a sensible low-cost approach to inducing marginal improvements in beginning teacher quality.[31]

These findings also speak to the more general issue of the heterogeneity in teacher effectiveness; that is, consistent with the well-known evidence that teachers differ significantly from one another in their impacts on student achievement, we find evidence that *the same teachers who have positive impacts on their own students' learning* also appear to be more effective mentors to beginning teachers. This broad conclusion clearly has implications for any field with a significant preservice mentoring component (e.g., nursing, medicine, etc.). As discussed in Section 2, while the vast majority of the broader mentorship literature to date has focused on the presence or type of mentoring, this study points to a promising future direction of research: investigating the productivity of the *specific mentors assigned to each mentee* as predictors of outcomes for those mentees. This approach could greatly improve our understanding of what constitutes an effective mentorship in a variety of contexts and potentially lead to more systematic and effective apprenticeships in many fields.

[31] Goldhaber et al. (2018a) show that only about 3% of in-service teachers host a student teacher in any given year, which means that even under the extreme example of placing student teachers only with teachers in the top quartile of teaching effectiveness, less than 1 in 8 of these teachers would need to host a student teacher in a given year. Moreover, Fives et al. (2016), for instance, note that the average compensation that mentor teachers received in 2012–13 was $232, far lower than the nearly $1,600 (adjusted for inflation) that was typical back in 1959, which suggests that there is the potential for substantially more investment in this area as well.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics, 25*(1), 95–135. doi:10.1086/508733

Ambrosetti, A., & Dekkers, J. (2010). The interconnectedness of the roles of mentors and mentees in pre-service teacher education mentoring relationships. *Australian Journal of Teacher Education, 35*(6).

American Association of Medical Colleges. (2017). *Table B-2.2: Total Graduates by U.S. Medical School and Sex, 2012–2013 through 2016–2017* [Table]. Retrieved from https://www.aamc.org/download/321532/data/factstableb2-2.pdf

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, *113*(1), 151–184.

Altonji, J. G., Elder, T. E., & Taber, C. R. (2008). Using selection on observed variables to assess bias from unobservables when evaluating Swan-Ganz catheterization. *American Economic Review*, *98*(2), 345–350.

Anderson, L. M., & Stillman, J. A. (2013). Student teaching's contribution to preservice teacher development. *Review of Educational Research, 83*(1), 3–69. doi:10.3102/0034654312468619

Aryee, S., Wyatt, T., & Stone, R. (1996). Early career outcomes of graduate employees: The effect of mentoring and ingratiation. *Journal of Management Studies*, *33*, 95–118

Bacher-Hicks, A., Kane, T., & Staiger, D. (2014). *Validating teacher effect estimates using changes in teacher assignments in Los Angeles.* (Working Paper, No. 20657). The National Bureau of Economic Research (NBER). doi:10.3386/w20657

Ballou, D., & Springer, M. G. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher*, 44(2), 77–86.

Blazar, D. (2015). Effective teaching in elementary mathematics: Identifying classroom practices that support student achievement. *Economics of Education Review, 48*, 16–29. doi: 10.1016/j.econedurev.2015.05.005

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.

Bozeman, B., & Feeney, M. K. (2007). Toward a useful theory of mentoring. *Administration & Society, 39*(6), 719–739. doi:10.1177/0095399707304119

Bureau of Labor Statistics, U.S. Department of Labor. (2018). Occupational outlook handbook. Retrieved from https://www.bls.gov/ooh/

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593–2632. doi: 10.1257/aer.104.9.2593

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–2679.

Chetty, R., Friedman, J. N., & Rockoff, J. (2016). Using lagged outcomes to evaluate bias in value-added models. *American Economic Review*, *106*(5), 393–99.

Cowan, J., Goldhaber, D., Hayes, K., & Theobald, R. (2016). Missing elements in the discussion of teacher shortages. *Educational Researcher,* 45(8), 460–462. doi:10.3102/0013189x16679145

Cowan, J., Goldhaber, D., & Theobald, R. (2018). An exploration of sources of variation in teacher evaluation ratings across classrooms, schools, and districts. CALDER Working Paper 197-0618-1.

Council on Social Work Education. (2015). Annual statistics on social work education in the United States. Retrieved from https://www.cswe.org/getattachment/992f629c-57cf-4a74-8201-1db7a6fa4667/2015-Statistics-on-Social-Work-Education.aspx

Crosby, O. (2002). Career training, credentials—and a paycheck in your pocket. *Occupational Outlook Quarterly*. Bureau of Labor Statistics, U.S. Department of Labor.

Eby, L. T., Rhodes, J. E., & Allen, T. D. (2007). Definition and evolution of mentoring. In T. D. Allen & L. T. Eby (Eds.), *The Blackwell handbook of mentoring: A multiple perspectives approach* (pp. 7–20). Oxford, England: Blackwell. doi:10.1111/b.9781405133739.2007.00002

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). Building the capacity of school districts to design, implement, and evaluate effective new teacher mentor programs: Action points for colleges and universities. *Mentoring & Tutoring: Partnership in Learning, 10*(1), 47–55. doi:10.1080/13611260220133144

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. D., Brewer, D. J., & Anderson, D. J. (1999). A three-way error components analysis of educational productivity. *Education Economics, 7*(3), 199–208. doi:10.1080/09645299900000018

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

Goldhaber, D., Gratz, T., and Theobald, R. (2017a). What's in a teacher test? Assessing the relationship between teacher licensure test scores and student secondary STEM achievement and course taking. *Economics of Education Review, 61*, 112–129.

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*, *42*, 106–124.

Goldhaber, D., Krieg, J. M., & Theobald, R. (2017b). Does the match matter? Exploring whether student teaching experiences affect teacher effectiveness. *American Educational Research Journal, 54*(2), 325–359. doi:10.3102/0002831217690516

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

Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher, 44*(5), 293–307. doi:10.3102/0013189x15592622

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

Goldhaber, D., Quince, V., & Theobald, R. (2018b). Has it always been this way? Tracing the evolution of teacher quality gaps in U.S. public schools. *American Educational Research Journal*, *55*(1), 171–201.

Greenberg, J., Pomerance, L., & Walsh, K. (2013). *Student teaching in the United States.* National Council on Teacher Quality (NCTQ). Retrieved from https://www.nctq.org/dmsView/Student_Teaching_United_States_NCTQ_Report

Greenberg, J., Walsh, K., & Mckee, A. (2013). *Teacher prep review: A review of the nation's teacher preparation programs*. National Council on Teacher Quality (NCTQ). Retrieved from https://www.nctq.org/dmsView/Teacher_Prep_Review_2013_Report

Grossman, P., Cohen, J., Ronfeldt, M., & Brown, L. (2014). The test matters: The relationship between classroom observation scores and teacher value added on multiple types of assessment. *Educational Researcher, 43*(6), 293–303. doi:10.3102/0013189x14544542

Hargreaves, A., & Jacka, N. (1995). Induction or seduction? Postmodern patterns of preparing to teach. *Peabody Journal of Education, 70*(3), 41–63. doi:10.1080/01619569509538834

Ingersoll, R., & Smith, T. M. (2004). Do teacher induction and mentoring matter? Retrieved from http://repository.upenn.edu/gse_pubs/134

Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics, 1*(4), 85–108. doi:10.1257/app.1.4.85

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

Kane, T. J., & Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation (No. w14607). National Bureau of Economic Research.

Kleiner, M. M., & Krueger, A. B. (2013). Analyzing the extent and influence of occupational licensing on the labor market. *Journal of Labor Economics, 31*(2) 2nd Ser., S173–S202. doi:10.1086/669060

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review, 47*, 180–195. doi:10.1016/j.econedurev.2015.01.006.

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. doi:10.3102/0162373716630739

Ladd, H. F., & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school. *Education Finance and Policy*, *12*(2), 241–279.

Matsko, K. K., Ronfeldt, M., Greene, H., Reininger, M., & Brockman, S. (forthcoming). The role of cooperating teachers in preparing pre-service teachers: A district-wide portrait. *Journal of Teacher Education.*

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.

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

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237–257. doi:10.3102/01623737026003237

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 1–18.

Papay, J., Taylor, E., Tyler, J., & Laski, M. (2016*). Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data* (Working Paper, No. 21986). National Bureau of Economic Research. doi:10.3386/w21986

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

Ronfeldt, M., Brockman, S., & Campbell, S. (2018a). Does cooperating teachers' instructional effectiveness improve preservice teachers' future performance? *Educational Researcher*.

Ronfeldt, M., Matsko, K. K., Nolan, H. G., Reininger, M. (2018b). *Who knows if our teachers are prepared? Three different perspectives on graduates' instructional readiness and the features of preservice preparation that predict them* (Working Paper, No.18-01). Stanford Center for Education Policy Analysis (CEPA). Retrieved from https://cepa.stanford.edu/sites/default/files/wp18-01-v201801.pdf

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247–252.

Rockoff, J. (2008). *Does mentoring reduce turnover and improve skills of new employees? Evidence from teachers in New York City* (Working Paper, No. 13868). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w13868

Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy, 4*(4), 537–571.

Rothstein, J. (2014). Revisiting the impacts of teachers. UC-Berkeley Working Paper.

Schwille, S. A. (2008). The professional practice of mentoring. *American Journal of Education 2008, 115*(1), 139–167.

Stamm, M., & Buddeberg-Fischer, B. (2011). The impact of mentoring during postgraduate training on doctors' career success. *Medical Education, 45*(5), 488–496. doi:10.1111/j.1365-2923.2010. 03857

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.

U.S. Department of Health and Human Services. (2014). The future of the nursing workforce: National- and state-level projections, 2012–2025. Retrieved from https://bhw.hrsa.gov/sites/default/files/bhw/nchwa/projections/nursingprojections.pdf

Vosters, K., Guarino, C., & Wooldridge, J. (2018). Understanding and evaluating the SAS EVAAS Univariate Model (URM) for measuring teacher effectiveness. *Economics of Education Review, 66*, 191–205.

Yendol-Hoppey, D. (2007). Mentor teachers' work with prospective teachers in a newly formed professional development school: Two illustrations. *Teachers College Record, 109*(3), 669–698.

Zeichner, K. M., & Gore, J. M. (1990). Teacher socialisation. In W. R. Houston (Ed.) *Handbook of research on teacher education.* New York: Macmillan.

Table 1. Mentor and Mentee Summary Statistics

Subject:	Math				ELA			
Column:	1	2	3	4	5	6	7	8
Sample:	All	Q4 Mentor VA	Q2-3 Mentor VA	Q1 Mentor VA	All	Q4 Mentor VA	Q2-3 Mentor VA	Q1 Mentor VA
Panel A: Mentor Characteristics								
Mentor Experience	14.160 (8.119)	14.416 (8.726)	14.288 (7.854)	13.648 (7.989)	14.650 (8.233)	15.403 (8.827)	14.414 (8.090)	14.371 (7.847)
Mentor Adv. Degree	0.741	0.738	0.743	0.742	0.780	0.745	0.797	0.780
Mentor WEST-B Math	0.207 (0.702)	0.241 (0.711)	0.070 (0.710)	0.398+ (0.630)	0.117 (0.771)	0.272 (0.696)	-0.020 (0.810)	0.222 (0.718)
Mentor WEST-B Reading	0.175 (0.781)	0.340 (0.728)	0.069 (0.655)	0.209 (0.963)	0.256 (0.729)	0.295 (0.642)	0.275 (0.709)	0.171 (0.847)
Mentor WEST-B Writing	0.148 (0.661)	0.214 (0.708)	0.022 (0.687)	0.293 (0.524)	0.195 (0.712)	0.289 (0.667)	0.217 (0.719)	0.043 (0.722)
Panel B: Mentee Characteristics								
Mentee Experience	2.225 (1.918)	2.123 (1.978)	2.209 (1.860)	2.358 (1.965)	2.190 (1.902)	2.292 (2.067)	2.028 (1.746)	2.413+ (1.997)
Mentee Adv. Degree	0.309	0.354	0.29	0.302	0.41	0.451	0.379	0.432
Mentee WEST-B Math	0.310 (0.718)	0.218+ (0.784)	0.382 (0.635)	0.257 (0.789)	0.135 (0.765)	0.187 (0.677)	0.176 (0.688)	0.003+ (0.951)
Mentee WEST-B Reading	0.110 (0.775)	0.101 (0.720)	0.112 (0.776)	0.115 (0.823)	0.112 (0.897)	0.178 (0.797)	0.161 (0.732)	-0.049 (1.212)
Mentee WEST-B Writing	0.159 (0.752)	0.149 (0.723)	0.160 (0.789)	0.167 (0.703)	0.245 (0.715)	0.297 (0.652)	0.296 (0.680)	0.096* (0.815)
Unique Mentees	1,044	243	536	265	944	220	497	277
Unique Mentors	924	220	472	232	895	198	447	250
Mentee Years	2,534	599	1,276	659	2,423	548	1,221	654

Note. Adv. = advanced; ELA = English Language Arts; Q1 = bottom quartile; Q2-3 = middle quartiles; Q4 = upper quartile; VA = value added. *P*-values from two-sided *t*-tests in columns 2 and 4 relative to column 3 and in columns 6 and 8 relative to column 7: $+p < 0.1$; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

Table 2. Student Summary Statistics

Subject:	Math				ELA			
Column:	1	2	3	4	5	6	7	8
Sample:	All	Q4 Mentor VA	Q2-3 Mentor VA	Q1 Mentor VA	All	Q4 Mentor VA	Q2-3 Mentor VA	Q1 Mentor VA
Prior Score in Math (Standardized)	0.000 (0.959)	0.022 (0.969)	0.014 (0.956)	-0.050 (0.952)	0.000 (0.966)	0.110** (0.955)	0.001 (0.965)	-0.112** (0.965)
Prior Score in ELA (Standardized)	0.000 (0.976)	0.027 (0.990)	0.017 (0.967)	-0.061+ (0.977)	0.000 (0.965)	0.121** (0.942)	0.003 (0.965)	-0.127** (0.973)
Female	0.492	0.495	0.491	0.490	0.490	0.489	0.491	0.489
American Indian	0.013	0.011	0.015	0.011	0.014	0.014	0.015	0.011
Asian/Pacific Islander	0.099	0.110	0.102	0.084+	0.110	0.111	0.109	0.112
Black	0.053	0.052	0.050	0.060	0.053	0.048	0.051	0.061
Hispanic	0.258	0.270	0.228	0.304**	0.231	0.194	0.227	0.276+
White	0.503	0.485	0.526	0.473+	0.515	0.552	0.522	0.463*
Learning Disability	0.061	0.061	0.057	0.068	0.060	0.047*	0.058	0.078*
Special Education	0.119	0.116	0.114	0.130	0.120	0.102+	0.117	0.141+
Gifted	0.053	0.052	0.055	0.050	0.051	0.073	0.047	0.037
Limited English	0.101	0.106	0.088	0.122*	0.093	0.065**	0.093	0.119+
Free/Reduced Lunch	0.518	0.528	0.491	0.562*	0.489	0.432+	0.484	0.558**
Number of Students	78,458	19,606	39,205	19,647	65,632	16,399	32,803	16,430

Note. ELA = English Language Arts; Q1 = bottom quartile; Q2-3 = middle quartiles; Q4 = upper quartile; VA = value added. *P*-values from two-sided *t*-tests in columns 2 and 4 relative to column 3 and in columns 6 and 8 relative to column 7: $+p < 0.1$; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

Table 3. Relationships Between Mentor Value Added and Student Achievement

Subject:	Math					ELA				
Column:	1	2	3	4	5	6	7	8	9	10
Mentor VA	0.116** (0.038)	0.190** (0.061)	0.163** (0.059)	0.169** (0.055)	0.112* (0.052)	0.050 (0.035)	0.103+ (0.062)	0.112+ (0.059)	0.117* (0.059)	0.113+ (0.067)
Log Time Since Student Teaching (Time)		0.007 (0.019)	0.009 (0.019)	0.007 (0.019)	0.041* (0.019)		-0.020 (0.016)	-0.017 (0.016)	-0.009 (0.016)	0.015 (0.017)
Mentor VA * Time		-0.082+ (0.048)	-0.055 (0.048)	-0.067 (0.048)	-0.065 (0.047)		-0.052 (0.047)	-0.058 (0.047)	-0.045 (0.047)	-0.042 (0.049)
Teachers	1044	1044	1044	1044	1044	994	994	994	994	994
Students	78458	78458	78458	78458	78458	65632	65632	65632	65632	65632
Mentor Controls			X	X	X			X	X	X
Mentee Controls				X	X				X	X
District Fixed Effects					X					X

Note: ELA = English Language Arts; VA = value added. Mentor value added calculated from all available years prior to student teaching placement. All models control for 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-price lunch, special education status and disability type, limited English proficiency indicator, migrant indicator, and homeless indicator. Mentor controls include WEST-B scores, institution attended, degree level, experience, and endorsement areas. Mentee controls include WEST-B scores, institution attended, degree level, and endorsement areas. District fixed effects include fixed effects both for the student teaching district and current school district. 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 4. Relationships Between CT Value Added, Teacher Value Added, and Attrition From Sample

	Math				ELA			
	1	2	3	4	5	6	7	8
Mentor VA	0.591 (0.559)		0.661 (0.612)	0.675 (0.609)	0.664 (0.651)		0.441 (0.719)	0.632 (0.724)
Mentee Prior VA		0.831+ (0.490)	0.789 (0.490)	0.719 (0.498)		-0.858 (0.743)	-0.866 (0.742)	-1.164 (0.773)
Mentor VA * Mentee Prior VA				1.877 (2.515)				6.784 (4.365)
Teachers	801	717	717	717	770	668	668	668

Note: ELA = English Language Arts; VA = value added. Mentor value added calculated from all available years prior to student teaching placement, and teacher value added calculated from all years up to the given school year. All coefficients are on the logit scale, and models control for indicators of annual teacher experience and the school year. Standard errors clustered at the teacher level are in parentheses. *P*-values from two-sided *t*-test: +$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

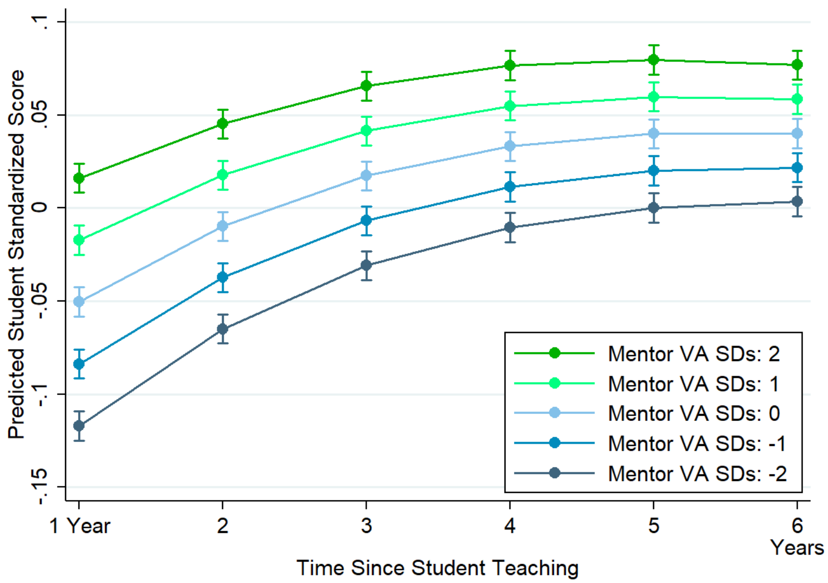
Table 5. Relationships Between Mentor Value Added and Student Achievement (First-Year Teachers Only)

	Math						ELA					
	1	2	3	4	5	6	7	8	9	10	11	12
Mentor VA	0.188** (0.059)	0.217** (0.066)	0.160** (0.055)	0.173** (0.059)	0.126* (0.050)	0.146** (0.053)	0.114+ (0.066)	0.113 (0.073)	0.107 (0.067)	0.132+ (0.072)	0.099 (0.069)	0.083 (0.071)
Teachers	474	376	474	376	474	376	452	347	452	347	452	347
Students	15266	12253	15266	12253	15266	12253	12523	9570	12523	9570	12523	9570
First-Year Teachers Only	X	X	X	X	X	X	X	X	X	X	X	X
Year After Student Teaching Only		X		X		X		X		X		X
Mentor Controls			X	X	X	X			X	X	X	X
Mentee Controls					X	X					X	X

Note: ELA = English Language Arts; VA = value added. Mentor value added calculated from all available years prior to student teaching placement. All models control for indicators 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-price lunch, special education status and disability type, limited English proficiency indicator, migrant indicator, and homeless indicator. Mentor controls include WEST-B scores, institution attended, degree level, experience, and endorsement areas. Mentee controls include WEST-B scores, institution attended, degree level, and endorsement areas. 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. Predicted Student Achievement by Time Since Student Teaching and Mentor Value Added

Panel A. Math



Panel B. ELA

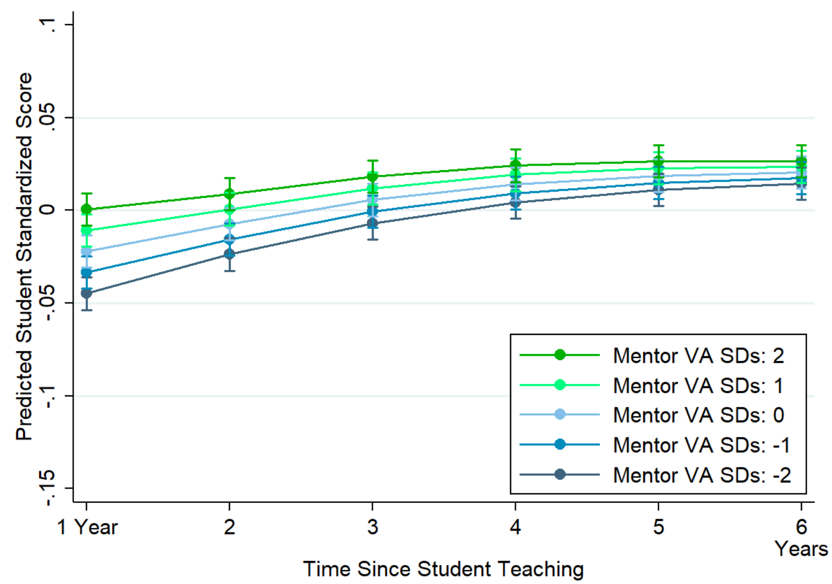


Figure 2. Marginal Effects on Student Achievement by Quartile of Mentor Value Added (First-Year Teachers Only)

Panel A. Math **Panel B. ELA**

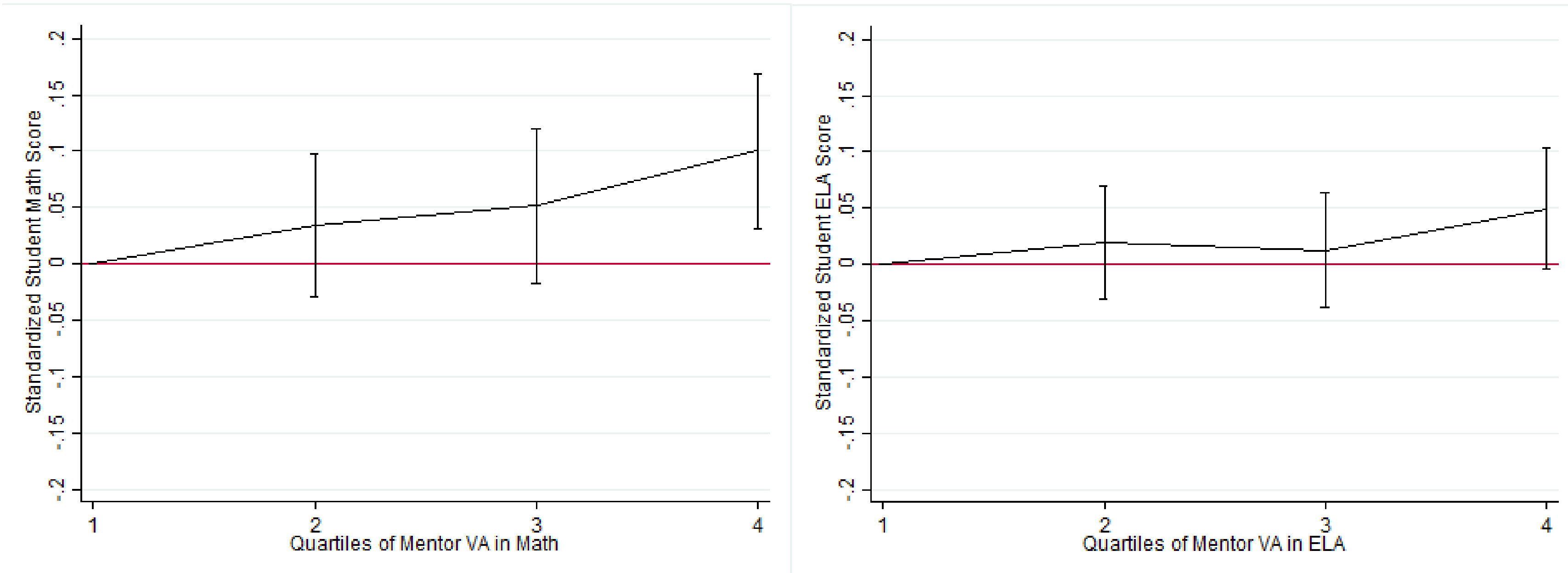


Table A1. Coefficients on Student Control Variables

Subject:	Math	ELA
Column:	1	2
Prior Score in Math (Standardized)	0.607*** (0.010)	0.261*** (0.007)
Prior Score in ELA (Standardized)	0.200*** (0.007)	0.513*** (0.008)
Female	-0.060*** (0.009)	0.133*** (0.009)
American Indian	-0.025 (0.034)	-0.141** (0.047)
Asian/Pacific Islander	0.150*** (0.020)	0.059*** (0.017)
Black	-0.090*** (0.027)	-0.086** (0.027)
Hispanic	-0.013 (0.015)	-0.028+ (0.015)
Learning Disability	-0.012 (0.025)	-0.025 (0.029)
Special Education	-0.154*** (0.019)	-0.163*** (0.020)
Gifted	0.231*** (0.029)	0.132*** (0.027)
Limited English	-0.039* (0.015)	-0.121*** (0.015)
Free/Reduced Lunch	-0.101*** (0.013)	-0.128*** (0.013)
Number of Students	78,458	65,632

Note. Column 1 reports estimated coefficients from specification reported in column 1 of Table 3, and column 2 reports estimated coefficients from specification reported in column 6 of Table 3. 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 A2. Coefficients on Mentor and Mentee Characteristics

Subject:	Math		ELA	
Column:	1	2	3	4
Mentor Experience	-0.002+ (0.001)	-0.002* (0.001)	0.002* (0.001)	0.002* (0.001)
Mentor Advanced Degree	0.002 (0.015)	0.002 (0.016)	0.005 (0.013)	0.012 (0.014)
Mentor WEST-B Math	-0.003 (0.025)	0.002 (0.023)	0.037 (0.024)	0.043+ (0.023)
Mentor WEST-B Reading	-0.025 (0.030)	-0.025 (0.027)	-0.022 (0.023)	-0.019 (0.022)
Mentor WEST-B Writing	0.024 (0.031)	0.010 (0.029)	-0.019 (0.026)	-0.025 (0.027)
Mentee 1–2 Years Experience	0.037* (0.017)	0.036* (0.018)	0.034* (0.015)	0.029+ (0.015)
Mentee 2–3 Years Experience	0.054* (0.023)	0.059* (0.023)	0.042* (0.020)	0.037+ (0.020)
Mentee 3–4 Years Experience	0.052+ (0.027)	0.055* (0.027)	0.049* (0.024)	0.042+ (0.023)
Mentee 4–5 Years Experience	0.034 (0.031)	0.034 (0.030)	0.065* (0.028)	0.054* (0.027)
Mentee 5+ Years Experience	0.057+ (0.034)	0.064+ (0.035)	0.097** (0.031)	0.089** (0.031)
Mentee Advanced Degree		0.024 (0.021)		-0.019 (0.014)
Mentee WEST-B Math		-0.008 (0.010)		-0.008 (0.008)
Mentee WEST-B Reading		-0.002 (0.011)		-0.003 (0.007)
Mentee WEST-B Writing		-0.012 (0.009)		0.004 (0.010)
Number of Students	78,458	78,458	65,632	65,632

Note. Column 1 reports estimated coefficients from specification reported in column 3 of Table 3, column 2 reports estimated coefficients from specification reported in column 4 of Table 3, column 3 reports estimated coefficients from specification reported in column 8 of Table 3, and column 4 reports estimated coefficients from specification reported in column 9 of Table

3. 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 A3. Relationships Between Quartiles of Mentor Value Added and Student Achievement (First-Year Teachers Only)

	Math						ELA					
	1	2	3	4	5	6	7	8	9	10	11	12
Mentor VA Q2 (Relative to Q1)	0.035 (0.032)	0.055 (0.038)	0.036 (0.032)	0.029 (0.036)	0.021 (0.031)	0.018 (0.035)	0.019 (0.026)	0.008 (0.029)	0.019 (0.026)	0.007 (0.029)	0.032 (0.026)	0.004 (0.030)
Mentor VA Q3 (Relative to Q1)	0.052 (0.035)	0.066 (0.041)	0.041 (0.036)	0.049 (0.042)	0.020 (0.036)	0.029 (0.041)	0.012 (0.026)	0.007 (0.030)	0.019 (0.025)	0.015 (0.029)	0.031 (0.026)	0.010 (0.029)
Mentor VA Q4 (Relative to Q1)	0.100** (0.035)	0.134*** (0.039)	0.091* (0.037)	0.109** (0.042)	0.056 (0.035)	0.082* (0.039)	0.049+ (0.028)	0.050 (0.031)	0.047 (0.029)	0.056+ (0.032)	0.054+ (0.030)	0.046 (0.031)
Teachers	474	376	474	376	474	376	452	347	452	347	452	347
Students	15266	12253	15266	12253	15266	12253	12523	9570	12523	9570	12523	9570
First-Year Teachers Only	X	X	X	X	X	X	X	X	X	X	X	X
Year After Student Teaching Only		X		X		X		X		X		X
Mentor Controls			X	X	X	X			X	X	X	X
Mentee Controls					X	X					X	X

Note: ELA = English Language Arts; Q1 = bottom quartile; Q2 = second quartile; Q3 = third quartile; Q4 = upper quartile; VA = value added. Mentor value added calculated from all available years prior to student teaching placement. All models control for indicators 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-price lunch, special education status and disability type, limited English proficiency indicator, migrant indicator, and homeless indicator. Mentor controls include WEST-B scores, institution attended, degree level, experience, and endorsement areas. Mentee controls include WEST-B scores, institution attended, degree level, and endorsement areas. Standard errors clustered at the teacher level are in parentheses. *P*-values from two-sided *t*-test: +$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.