Labour Economics 63 (2020) 101792

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

Effective like me? Does having a more productive mentor improve the productivity of mentees?

Dan Goldhaber[a,b], John Krieg[c], Roddy Theobald[a,*]

[a] American Institutes for Research, USA
[b] University of Washington, USA
[c] Western Washington University, USA

ARTICLE INFO

Keywords:
Education
Mentorship
Teacher preparation
Value added

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, and would be positive and statistically significant even in the presence of non-random sorting on unobservables of the same magnitude as the sorting on observables. 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 many 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 that includes nearly a decade of data on the preservice apprenticeships ("student teaching placements") of teacher candidates in Washington State—linked to data on the workforce outcomes of these candidates once they become K-12 teachers—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. This work builds on prior work on the mentorship of *in-service* teachers (e.g., Papay et al., 2016; Rockoff, 2008), and follows a similar study that used more limited data (see Section 2) to investigate the same question about *pre-service* teachers in Tennessee (Ronfeldt et al., 2018a). Importantly, this is the first study to consider mentor productivity as measured *before* the mentor-mentee relationship to investigate this key question.

There are a number of reasons to focus on the connection between mentor and mentee productivity in the case of teaching. First, teachers are 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. Moreover, the vast majority of these three million public school teachers served as student teachers in the past as will the majority of teacher candidates today. Teachers have also been shown to play a critical role in the creation of future human capital.[2] Finally, and importantly for the purposes

* Corresponding author.
E-mail address: rtheobald@air.org (R. Theobald).

[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).

[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).

https://doi.org/10.1016/j.labeco.2019.101792
Received 7 May 2019; Received in revised form 13 December 2019; Accepted 16 December 2019
Available online 18 December 2019

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] It is important to note that value added is *not* direct measure of the quality of mentorship provided by mentors, but to the extent that teaching quality and mentorship quality are correlated, this measure helps us understand how much mentees benefit from a high-quality mentor.

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 novice and second-year teacher; in other words, the expected gain in teacher effectiveness from assignment to a more effective mentor is equivalent to the well-documented returns to the first year of teaching experience (e.g., Ladd and 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, 2019) and qualitative (St. John et al., 2018) evidence from Washington State (the setting of this study) documents considerable non-random 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, 2019)—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 by following the approach of Altonji et al. (2005, 2008) and Oster (2017) who provide methods for estimating the bias produced by different magnitudes of non-random sorting. We show that under plausible scenarios—including a scenario in which the amount of sorting on unobservable dimensions is the same as the amount of sorting on observables—the relationship between mentor effectiveness in math and the future effectiveness of their mentees in math is still positive and statistically significant. We interpret this as evidence that at least some of this relationship reflects a causal relationship between mentor effectiveness and the future effectiveness of their mentees in math.

[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.

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 and Dekkers, 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 a randomized set of schools in the treatment group were provided with a list of "suggested" mentor-mentee pairings of high-performing and low-performing teachers based on prior evaluation scores (and schools in the control group did "business as usual"). 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 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 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

[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).

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

candidates are characterized as "a key component—even 'the most important' component of—pre-service teacher preparation" (Anderson and 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 fairoeedback and support to teacher candidates who are just learning to practice their craft (e.g., Ambrosetti and Dekkers, 2010; Grossman et al., 2014; Schwille, 2008; Yendol-Hoppey, 2007; Zeichner and 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 and Jacka, 1995). We also note that, while all teacher candidates during our years of data were required to complete a student teaching placement, student teaching is *not* required in many alternative routes to the teaching profession (including some in Washington that have been established since these data were collected). Thus this study also provides complementary evidence to earlier studies about these alternative routes to teacher certification (e.g., Glazerman et al., 2006; Machin and McNally, 2008) that do not require student teaching.

Importantly for this study, there is both quantitative (Krieg et al., 2016, 2019) and qualitative (Meyer, 2016; St. John et al., 2018) evidence about the factors that influence that matching of mentees to mentors in student teaching placements, much of it from Washington State (the setting of this study). For example, a qualitative study in Washington (St. John et al., 2018) summarizes interviews with the individuals responsible for student teaching placements for teacher education programs and for partnering schools and districts in the state and finds tremendous variation in placement processes across programs and districts. Two quantitative studies (Krieg et al., 2016, 2019) investigate the observable characteristics of potential mentors that predict hosting a student teacher and find that (at least among observable variables) both time-invariant (e.g., licensure test scores) and time-variant (e.g., experience) characteristics are associated with serving as a mentor.[8]

There is comparatively little quantitative evidence about the relationship between mentor teachers and later teacher candidate performance. Matsko et al. (2019 forthcoming) find positive 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).[9]

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.[10]

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.[11]

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 non-random 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.

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).[12] The OSPI data include annual

[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., 2013, 2011). States also have other preservice requirements associated with licensure, such as passing various licensure tests (Goldhaber, 2007).

[8] There is little additional data about student teaching placement processes that is systematically collected in Washington, though w Meyer (2016) quantifies some student teaching policies in Missouri.

[9] 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.

[10] The authors report that this is about a third of the estimated return to the first year of teaching experience.

[11] 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.

[12] 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

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.[13] 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.[14] Because we estimate 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 and 07 and 2016–17 and in math or reading in Grades 6–8 between 2009 and 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 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.[15] 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 some candidates in these years are 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).

It is important to note that this timing of data collection results in three different types of teachers in the final analytic data set. First, there are teachers who begin teaching in a tested grade immediately after their student teaching experience. For these teachers, we observe their performance immediately after their mentorship experience—exactly at the time one would think mentoring would have its greatest effect. The second type of teacher takes some time after their student teaching placement to entering the workforce in a tested grade, while the final type of teacher teaches for a few years before teaching in a tested grade In Section 4, we explain how we estimate models that ultimately consider all three groups of teachers in the analysis.

The OSPI data also include other measures of the background and credentials of both mentors and mentees, including information on years of teaching experience; degree level (e.g., bachelor's or master's); 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.[16] Moreover, since the WEST-B has only been a licensure requirement since 2002, scores are missing for most of the (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 1044 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 2534 mentee-year observations linked to 78,458 student observations in math and 2423 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 of mentor value added. 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 serve as the control variables in the models described in the next section. Here we see some non-random sorting of students to classrooms related to 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, 2019), 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 non-random sorting along observable dimensions. This section also describes a robustness check that uses the non-random sorting by observable variables as a proxy for the amount on non-random 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; Rivkin 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, 2015; Kane and Staiger, 2008; Kane et al., 2013; Rothstein, 2009, 2014). While this issue is not settled,[17] we argue that appropriately specified

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.

[13] 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.

[14] 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.

[15] 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).

[16] 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).

[17] See, for instance, the debate between Chetty et al. (2014a, 2016) and Rothstein (2014).

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	14.416	14.288	13.648	14.650	15.403	14.414	14.371
	(8.119)	(8.726)	(7.854)	(7.989)	(8.233)	(8.827)	(8.090)	(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	0.241	0.070	0.398+	0.117	0.272	−0.020	0.222
	(0.702)	(0.711)	(0.710)	(0.630)	(0.771)	(0.696)	(0.810)	(0.718)
Mentor WEST-B Reading	0.175	0.340	0.069	0.209	0.256	0.295	0.275	0.171
	(0.781)	(0.728)	(0.655)	(0.963)	(0.729)	(0.642)	(0.709)	(0.847)
Mentor WEST-B Writing	0.148	0.214	0.022	0.293	0.195	0.289	0.217	0.043
	(0.661)	(0.708)	(0.687)	(0.524)	(0.712)	(0.667)	(0.719)	(0.722)
Panel B: Mentee Characteristics								
Mentee Experience	2.225	2.123	2.209	2.358	2.190	2.292	2.028	2.413+
	(1.918)	(1.978)	(1.860)	(1.965)	(1.902)	(2.067)	(1.746)	(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	0.218+	0.382	0.257	0.135	0.187	0.176	0.003+
	(0.718)	(0.784)	(0.635)	(0.789)	(0.765)	(0.677)	(0.688)	(0.951)
Mentee WEST-B Reading	0.110	0.101	0.112	0.115	0.112	0.178	0.161	−0.049
	(0.775)	(0.720)	(0.776)	(0.823)	(0.897)	(0.797)	(0.732)	(1.212)
Mentee WEST-B Writing	0.159	0.149	0.160	0.167	0.245	0.297	0.296	0.096*
	(0.752)	(0.723)	(0.789)	(0.703)	(0.715)	(0.652)	(0.680)	(0.815)
Unique Mentees	1044	243	536	265	994	220	497	277
Unique Mentors	924	220	472	232	895	198	447	250
Mentee Years	2534	599	1276	659	2423	548	1221	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	0.022	0.014	−0.050	0.000	0.110**	0.001	−0.112**
	(0.959)	(0.969)	(0.956)	(0.952)	(0.966)	(0.955)	(0.965)	(0.965)
Prior Score in ELA (Standardized)	0.000	0.027	0.017	−0.061+	0.000	0.121**	0.003	−0.127**
	(0.976)	(0.990)	(0.967)	(0.977)	(0.965)	(0.942)	(0.965)	(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: $+p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

VAMs show minimal bias (Koedel et al., 2015), especially in estimating teacher effectiveness in math.[18]

To make our calculations concrete, define t^{M}_{jk} 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):

$$Y_{ijst\prime} = \alpha_0 + \alpha_1 Y_{i(t\prime-1)} + \alpha_2 X_{it\prime} + \sum_k \alpha_{k+3} I\left(Exp_{jt\prime} = k\right) + \tau^{M}_{jks(t\prime<t^{M}_{jk}\}} + \varepsilon_{ijst\prime} \quad (1)$$

In (1), $Y_{ijst\prime}$ is the state standardized test score for each student i with teacher j in subject s (math or reading) and year $t\prime$, normalized within grade and year[19]; $Y_{i(t\prime-1)}$ is a vector of student i's scores the previous year in both math and reading, also normalized within grade and year; $X_{it\prime}$ is a vector of student attributes in year $t\prime$ (gender, race, FRL status, English language learner status, gifted status, special education status, learning disability status); and $Exp_{jt\prime}$ is the experience of teacher j in year $t\prime$ (included as indicators for different years of teaching experience). The estimate of mentor value added for mentee k, $\tau^{M}_{jks(t\prime<t^{M}_{jk}\}}$, 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

[18] 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.

[19] We use the notation t' to represent the years in which we're estimating mentor teacher value added, which are distinct from the years t in which we estimate future mentee effectiveness (see equation 2).

estimates using empirical Bayes methods; as described in Jacob and Lefgren (2008), shrinking value-added estimates before including them as predictors in Eq. (2) accounts for some bias in the estimated coefficients due to measurement error and uncertainty in these estimates.[20]

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 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.

The experience controls in Eq. (1) are also important because, as described above, mentor value added is being estimated from different years of data than the year in which the mentor hosts a student teacher. This means that there are two measures of mentor experience that are important; their experience in the years we observe student test scores, and their experience in the year they host a student teacher. The first measure is a confounder—given well-documented returns to teaching experience, we would not want mentors to be considered as more effective just because we observe student test scores in years in which they have more experience—while the second measure is a variable of interest (i.e., are there benefits to assignment to a more experienced mentor?). This motivates the inclusion of experience controls in Eq. (1) (i.e., to create an experience-adjusted measure of mentor value added), and then our consideration of mentor experience in the student teaching year as one of the mentor characteristics considered in some specifications of the second-stage models in equation 2.[21]

A final concern is about the specification of the value-added model in Eq. (1). In particular, the measure of value-added produced by the model in equation compares all teachers across the state, as opposed to within a given district (i.e., with a district fixed effects model) or a given school (i.e., with a school fixed effects model). Our preferred specifications of mentor teacher value added do not include any other fixed effects because TEPs typically place student teachers across a large number of districts—the median TEP in our sample places student teachers in 39 different districts over the years of data we consider, and every TEP places student teachers in at least 6 districts—so a measure of teacher quality that's comparable across the whole state seems preferable for policy purposes. We do experiment with different fixed effects in our second-stage models (see Eq. (2)) that make comparisons within specific internship districts and schools.

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 Eq. (1)) and the same set of controls:

$$Y_{ikst} = \beta_0 + \beta_1 Y_{i(t-1)} + \beta_2 X_{it} + \sum_{k}^{4} \beta_{k+3} I\left(Exp_{kt} = k\right) + \beta_8 \hat{\tau}^{M}_{jks(t\prime<t^{M}_{jk}\}} + \varepsilon_{ijst} \tag{2}$$

The variables in Eq. (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 after then mentee begins their teaching career. We cluster standard errors at the teacher level to account for correlated errors for different students with the same teacher.[22]

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. We account for this in two separate ways. In our preferred models, we restrict estimation of (2) to mentee teachers in their first full year of teaching. While this restriction reduces the sample size, it also focuses exclusively on teachers at the point nearest to their mentorship experience which is where one would expect to see the greatest impact of mentorship. By the same logic, we further restrict the observations in some additional specifications to mentee teachers who were student teachers *in the prior year*, thereby eliminating teachers who took time to find their first job.

In a second approach where we include all observations, we explicitly measure the impact of time between the student teaching year t^{M}_{jk} and the current year t and interact the log of this term with mentor value added :[23]

$$Y_{ikst} = \gamma_0 + \gamma_1 Y_{i(t-1)} + \gamma_2 X_{it} + \sum_{k}^{4} \gamma_{k+3} I\left(Exp_{kt} = k\right) + \gamma_8 \hat{\tau}^{M}_{jks(t\prime<t^{M}_{jk}\}} + \gamma_9 log\left(t - t^{M}_{jk}\right) + \gamma_{10} log\left(t - t^{M}_{jk}\right) * \hat{\tau}^{M}_{jks(t\prime<t^{M}_{jk}\}} + \varepsilon_{ijst} \tag{3}$$

In the specification in Eq. (3), γ_8 represents the relationship between mentor value added and student achievement the year immediately following student teaching (i.e., when $t - t^{M}_{jk} = 1 \Rightarrow log(t - t^{M}_{jk}) = 0$). 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. The benefit of this approach is it can be applied to all observations, including those teachers who did

[20] 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.

[21] A third measure of experience that is potentially important is a mentor's experience in a mentorship role, but the limited years of data preclude us from creating such a measure (only 10% of mentors are observed hosting a student teacher more than once).

[22] One concern with equation (2) is the inclusion of an estimated variable (mentor value added) on the right-hand side. It is well known that ignoring the fact that an explanatory variable is estimated can lead to standard errors that are biased downwards. However, there is also the possibility that the Bayesian shrinkage "over shrinks" the estimates and causes the standard error to be biased upwards (McCaffrey et al., 2009). As a check on these possibilities, we calculate cluster bootstrapped standard errors for a subset of model specifications by first sampling with replacement teachers in the data used to estimate mentor value added, estimating mentor value added (equation 1) and then estimating the second stage model only for these mentor teachers (equation 2), and then repeating this procedure 500 times. We find the bootstrapped standard errors are slightly less conservative than the conventionally-estimated clustered standard errors, so we report the conventional standard errors so as to not overstate the statistical significance of our findings.

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

not immediately move from student teaching into a tested grade. The potential downside is that these specifications are potentially subject to non-random attrition bias, which we discuss below.

A second concern with both Eqs. (2) and (3) are non-random sorting of mentees into mentors. While the models in Eqs. (2) and 3 include a rich set of variables controlling for potential bias, it is possible that non-random sorting of mentees remain a threat to causal interpretations. For instance, if high-ability student teachers are supervised by mentors with higher value added, a finding corroborated by Krieg et al. (2016, 2019) using licensure test scores, then $\hat{\beta}_8$ and $\hat{\gamma}_8$ in Eqs. (2) and (3) will be biased upward.[24] 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 experiment with fixed effects for the school or district in which the student teaching took place, as well as fixed effects for the school or district where the mentee teaches. These fixed effects control for the sorting of mentees to student teaching schools and districts and their later hiring schools and districts, which is important because of prior evidence linking the location of teacher education programs to job placement (Krieg et al., 2016, 2019) and to districts where teacher trainees grew up (Boyd et al., 2005). On the other hand, the hiring of mentees into specific schools and districts may be endogenous to mentor effectiveness, so we primarily view these fixed effects models as robustness checks for our primary results.

A third 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 summary statistics of mentees that do and do not appear in the final analytic sample show very minimal differences in mentor value added (see Table A1 in the appendix).

That said, we investigate this issue further by estimating a logit model that predicts appearance in our analytic sample based upon mentor's value added pursuing a bounding exercise adapted from Lee (2009) by Carrell et al. (2018).[25] Specifically, suppose the relationship between mentor value added and sample entry is positive. We randomly and incrementally drop teachers in the sample whose mentor value added is above the mean until the point estimate from the sample entry logits goes to zero (i.e., is within 0.002 of zero), estimate the relationship between mentor value added and student achievement from Eq. (2), and repeat this procedure 500 times to approximate the range of estimates that could be observed under differential sample entry of the magnitude estimated from the logit models.

However, even after performing these robustness checks, 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.

A fourth concern has to do with attrition of mentees from the sample. If more effective teachers who were supervised by more effective mentors are differentially likely to leave the workforce, this would also bias our estimates $\hat{\beta}_8$ and $\hat{\gamma}_8$. For models where we restrict the sample to only first year teachers, this concern is moot; we observe these teacher's value added prior to any possibility of attrition. For models without this restriction, we test for non-random attrition directly by estimating models predicting the probability of 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 of these two variables:

$$log\left(\frac{A_{kst}}{1-A_{kst}}\right) = \omega_0 + \omega_1 \hat{\tau}^{M}_{jks(t\prime<t^{M}_{jk})} + \omega_2 \hat{\tau}_{ks(t\prime\leq t)} + \omega_3 \hat{\tau}^{M}_{jks(t\prime<t^{M}_{jk})} * \hat{\tau}_{ks(t\prime\leq t)} + \sum_{k}^{4} \omega_{k+3} I\left(Exp_{kt} = k\right) \tag{4}$$

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 in the next section.

Of these potential sources of bias, we are primarily concerned about the potential non-random 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 an 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 that could be caused by selection on unobservables. Under this methodology, let W_{ijkst} represent *all unobserved variables* that are jointly correlated with the value added of mentor j, $\hat{\tau}^{M}_{jks(t\prime<t^{M}_{jk})}$ 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 Eq. (2) (formally, $\delta\frac{\sigma_{V\tau}}{\sigma^2_V} = \frac{\sigma_{W\tau}}{\sigma^2_W}$, where $\sigma_{V\tau} = Cov(V_{ijkst}, \hat{\tau}^{M}_{jks(t\prime<t^{M}_{jk})})$ and $\sigma_{W\tau} = Cov(W_{ijkst}, \hat{\tau}^{M}_{jks(t\prime<t^{M}_{jk})})$). 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 $\hat{\tau}^{M}_{jks(t\prime<t^{M}_{jk})}$, R^0, the observed estimate $\hat{\beta}_8$ and the R-squared of the model in Eq. (2), $\tilde{R}$, and the maximum possible R-squared from a model predicting Y_{ikst}, R_{max}:[26]

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

In Eq. (5) δ represents the magnitude of non-random sorting on unobservables. A value of $\delta = 1$ represents the case where sorting on unobservables is of the same magnitude as the sorting on our extensive set of observed covariates. We experiment with different values of δ to explore the sensitivity of our results to different amounts of sorting on unobservables (as a proportion of the sorting on observables). We 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

[24] 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.

[25] The dependent variable in these regressions is a binary indicator for appearing in the analytic sample in at least one year.

[26] 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 $\hat{\tau}^{M}_{jks(t\prime<t^{M}_{jk})}$.

Table 3
Relationships Between Mentor Math Value Added and Mentee's Students' Math Achievement .

	1	2	3	4	5	6	7	8	9	10	11	12
Mentor VA	0.188**	0.217**	0.160**	0.173**	0.126*	0.146**	0.116**	0.190**	0.163**	0.169**	0.141**	0.167**
	(0.059)	(0.066)	(0.055)	(0.059)	(0.050)	(0.053)	(0.038)	(0.061)	(0.059)	(0.055)	(0.054)	(0.062)
Log Time Since ST (Time)								0.007	0.009	0.007	0.039*	0.031
								(0.019)	(0.019)	(0.019)	(0.019)	(0.020)
Mentor VA * Time								−0.082+	−0.055	−0.067	−0.063	−0.095+
								(0.048)	(0.048)	(0.048)	(0.046)	(0.055)
Teachers	474	376	474	376	474	376	1044	1044	1044	1044	1044	1044
Students	15,266	12,253	15,266	12,253	15,266	12,253	78,458	78,458	78,458	78,458	78,458	78,458
R^2	0.698	0.695	0.704	0.702	0.710	0.710	0.720	0.720	0.722	0.724	0.731	0.747
First-Year Only	X	X	X	X	X	X						
Year After ST Only		X		X		X						
Mentor Controls			X	X	X	X			X	X	X	X
Mentee Controls					X	X				X	X	X
Current district FEs											X	
Current school FEs												X

Note: FE = fixed effect; ST = student teaching; 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: $+p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

between mentor value added and student performance could be "explained away" by different amounts of sorting on unobservables.[27]

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 A2 in the appendix for the model specifications that are reported in Table 2). All of these findings are quite consistent with the broader literature.[28] 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 A3 in the appendix).

5.1. Primary findings

Tables 3 and 4 show the primary coefficients of interest between mentor effectiveness and the later effectiveness of their mentees. We begin in column 1, Table 3 (for math) and Table 4 (for ELA) with a sparse model that omits controls for mentors other than their value added, does not include measures of mentee quality, and restricts the sample to first-year teachers (i.e., Eq. (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 18% of a standard deviation increase of the effectiveness of their mentees; this is roughly the half of the difference between a novice teacher and one with one to two years of experience (see Appendix Table A3) and about three times as large as the comparable estimate in Ronfeldt et al. (2018a).[29] The estimated relationship from the specification in ELA is only marginally statistically significant (and only slightly smaller than the comparable estimated relationship in Ronfeldt et al., 2018a).

In column 2 of both tables we restrict the sample to only those teachers who served as a mentee in the immediate prior year. The purpose of this restriction is to examine teachers as close as possible to the time served with their mentor. This additional restriction causes the magnitude of the math coefficient to increase, perhaps as a result of less decay happening between internship and first job.

To explore the potential that the findings on mentor effectiveness are related to other observable mentor characteristics, we add the mentor characteristics listed in Table 2 to the regressions in columns three and four (of both Tables) with the difference in the pair of columns being the first-year teacher restriction. Though these mentor controls explain a statistically significant (though modest) amount of variation in both math and ELA scores, 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 A3) these mentor characteristics are generally weak predictors of mentee value added.[30] Similarly, in columns five and six, we show the findings when we add analogous controls for preservice mentee quality (including indicators for the institution from which each mentee graduated). The fact that the math results remain statistically significant after including this set of controls provides cursory evidence that the results are not driven by the non-random matching of mentor and mentee quality (at least based on observables). However, relative to the initial mod-

[27] Our implementation of this procedure is very conservative both in our choice of $R_{max} = 1$ 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).

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

[29] 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 A3).

[30] 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).

Table 4
Relationships Between Mentor ELA Value Added and Mentee's Students' ELA Achievement .

	1	2	3	4	5	6	7	8	9	10	11	12
Mentor VA	0.114+	0.113	0.107	0.132+	0.099	0.083	0.050	0.103+	0.110+	0.116+	0.073	0.065
	(0.066)	(0.073)	(0.067)	(0.072)	(0.069)	(0.071)	(0.035)	(0.062)	(0.059)	(0.059)	(0.065)	(0.068)
Log Time Since ST (Time)								−0.020	−0.017	−0.009	0.010	0.020
								(0.016)	(0.016)	(0.016)	(0.016)	(0.018)
Mentor VA * Time								−0.052	−0.057	−0.045	−0.061	−0.017
								(0.047)	(0.047)	(0.047)	(0.050)	(0.054)
Teachers	452	347	452	347	452	347	994	994	994	994	994	994
Students	12,523	9570	12,523	9570	12,523	9570	65,632	65,632	65,632	65,632	65,632	65,632
R^2	0.646	0.641	0.648	0.644	0.652	0.648	0.683	0.683	0.684	0.685	0.691	0.703
First-Year Only	X	X	X	X	X	X						
Year After ST Only		X		X		X						
Mentor Controls			X	X	X	X			X	X	X	X
Mentee Controls					X	X				X	X	X
Current district FEs											X	
Current school FEs												X

Note: ELA = English Language Arts; FE = fixed effect; ST = student teaching; 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: $+p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

els with no controls, the math results are about forty percent smaller after adding these of controls, suggesting that further investigation into the role of non-random sorting of mentees to mentors is warranted.

We now turn to the full sample results which includes all mentees listed including those who are observed after several years in the workforce. Because these specifications include some experienced teachers observed only after entering tested grades (as well as additional observations for teachers observed earlier in their careers as well), we include binary variables accounting for different teaching experience levels as controls as well as interactions for the (logged) amount of time elapsed since student teaching (i.e., the specification in Eq. (3) of the previous section) to account for the possibility that the effects of working with more effective mentors decay over time. Before discussing these results, we note that our attrition models (reported in Appendix Table A4 and estimated from the model Eq. (4)) do not suggest that non-random attrition (at least on observed dimensions) is a significant source of bias in these models. In particular, none of these marginal effects are statistically significant, and most importantly, we do not see systematic heterogeneity in this relationship by mentor value added.

The full-sample results are presented in columns 7 through 12 of Tables 3 and 4. For reference, the seventh column of these tables report the full sample results without controlling for the decay of mentor effects over time. Both the math and ELA in these columns are positive, though only the math results are statistically significant. We add the time variables in the eighth column and find that the coefficient on the interaction between time and mentor value added is marginally significant and negative (in math), 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.[31]

This conclusion can be seen visually in Panel A of Fig. 1, which plots predicted student achievement from the specification in column 8 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, the increasing importance of unobserved in-service influences (e.g., in-service mentors), or even a mentor's impact on workforce attrition (this is 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 Fig. 1, though, the magnitudes of these relationships are considerably more modest in ELA than in math.

Columns nine and ten of both tables successively add mentor-level controls and mentee-level controls. For both the math and ELA results, the additional control variables cause the estimates of β_8 to shrink some, though the math results remain statistically significant in all specifications. 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 placements, and the likelihood of mentees being employed in particular school systems. As we described in Section 3, this could be another source of non-random sorting of more effective mentees to more effective mentors. To account for this possibility, we successively include (columns 11 and 12) fixed effects for the school districts and the school buildings in which the mentee taught. In these models the coefficient on mentor effectiveness are being identified based on the within-district (or school) variation in both mentor and mentee value added. The estimates are slightly more modest (especially in ELA), but we still see statistically significant main effects in math.

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 5, we summarize the results of the bounding exercise described in Section 4 and adapted from Carrell et al. (2018) and Lee (2009) that is intended to approximate the range of estimates that could be observed under the observed differential sample selection by mentor value added. Columns 1 and 4 of Table 5 repeat the estimated relationships between mentor value added and stu-

[31] $0.190 - log(10) * .082 \approx 0$.

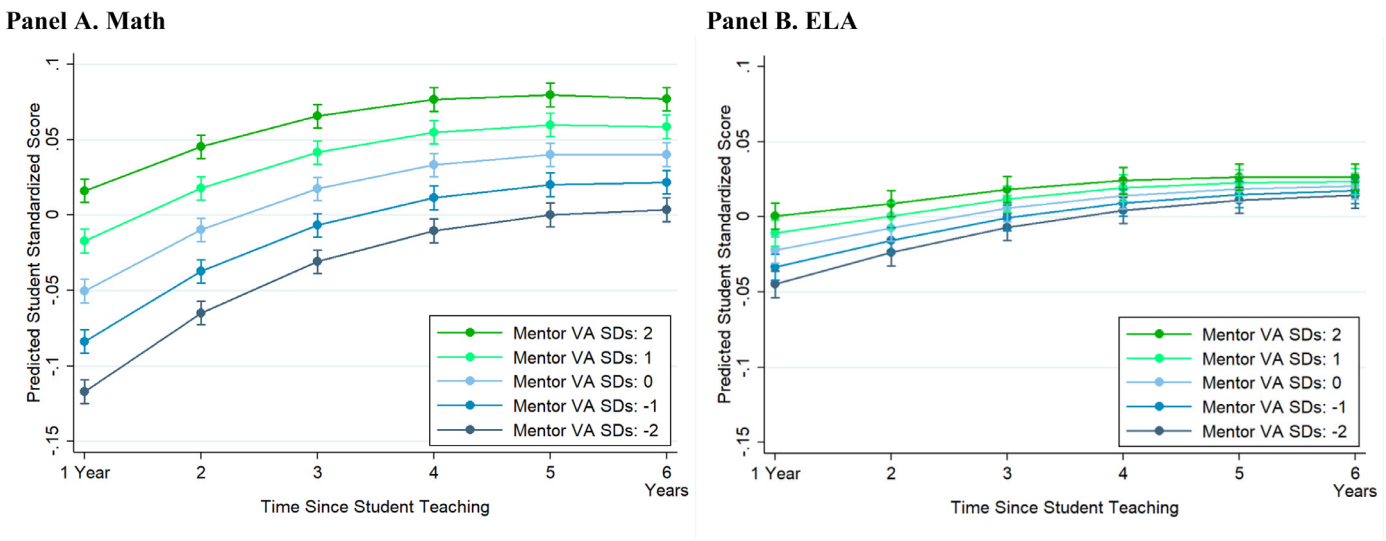

Fig. 1. Predicted Student Achievement by Time Since Student Teaching and Mentor Value Added.

Table 5
Lee/Carrell et al. Bounds for Sample Selection Bias.

	Math			ELA		
Column	1	2	3	4	5	6
Model	**Main**	**Entrance**	**Lee Bounds**	**Main**	**Entrance**	**Lee Bounds**
Mentor VA	0.169***	−0.041	0.169	0.116+	0.064	0.118
	(0.055)	(0.046)	[0.146,0.189]	(0.059)	(0.056)	[0.093,0.143]
Teachers	1044	2663	1044	994	2704	994
Students	78,458		78,458	65,632		65,632
Mentor Controls	X	X	X	X	X	X
Mentee Controls	X	X	X	X	X	X

Note: ELA = English Language Arts; VA = value added. Math maim models come from column 10 of Table 3, while ELA models come from column 10 of Table 4. Entrance models represent average marginal effects from logistic regressions predicting appearance in the analytic samples. Lee Bounds represent the average and 95% confidence intervals of estimates of the main model from 500 iterations of the Lee Bounds procedure described in Carrell et al. (2018), Section III, Part E. *P*-values from two-sided *t*-test: $+p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

dent performance from the specifications of Eq. (2) that include both mentee and mentor controls (i.e., column 10 of Tables 3 and 4) in math and ELA. Columns 2 and 5 then provide the marginal effect of mentor value added on sample selection; this relationship is negative in math, positive in ELA, and not statistically significant in either subject. Not surprisingly, then, the average estimates from the Carrell et al. (2018) and Lee (2009) bounding exercise in columns 3 and 6 are quite similar to the uncorrected estimates, with a range of estimates well above zero in both subjects. This suggests that sample selection bias is not a serious threat to validity in this analysis.

In Table 6, we present Oster's (2017) tests on whether the estimated relationship between mentor value added and student performance would still be statistically significant under different hypothetical scenarios where the non-random sorting of mentees to more effective mentors along unobserved dimensions changes magnitude. To include the most observable variables possible (and to simplify the interpretation of our estimates), we use the specification in Eq. (3) (i.e., with decay) and include the full array of mentor and mentee controls in column 10 of Tables 3 and 4. The Oster results are presented with different levels of hypothetical sorting captured by δ, which can be interpreted as the proportional amount of nonrandom sorting relative to the sorting on observed variables. For instance, a value of $\delta = 0.5$ means that there is half as much non-random sorting of mentees to mentors along unobserved dimensions as there is on observed dimensions.

For ease of comparison we begin the math and ELA portions of Table 5 with a value of $\delta = 0$ which corresponds to cases where we assume there is no effects of sorting between mentees and mentors; this is identical to the results presented in column 10 of Tables 3 and 4. We then increase δ until the coefficient on a mentor's value added falls below conventional levels of statistical significance. For the math results, this occurs for values of δ somewhere between 1.0 and 1.25, suggesting that the significant relationship between mentor and mentee quality can only be "explained away" if the amount the non-random sorting of mentees to mentors along unobserved dimensions is greater than the sorting on observables. Given that a simulation in Oster (2017) that "supports the idea of 1 as an upper bound on δ" (Oster, 2017, p. 11) and the fact that our models control for a number of important and potentially confounding variables, we view this amount of sorting on unobservables as unlikely. On the other hand, the relationship in ELA is much more sensitive to these assumptions; sorting on unobservables that is only 25% as great as the sorting on observables can explain away this relationship. 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.

Finally, we use the specifications in column 1 of Tables 3 and 4 to further investigate nonlinearities in these relationships by swapping in quartiles of mentor value added for the continuous measure discussed

Table 6
Oster Tests for Unobservable Sorting Bias .

	Math						ELA	
Column	1	2	3	4	5	6	7	8
Delta	**0**	**0.25**	**0.5**	**0.75**	**1.0**	**1.25**	**0**	**0.25**
Mentor VA	0.169***	0.156***	0.142**	0.128*	0.114*	0.100	0.116+	0.064
	(0.055)	(0.040)	(0.044)	(0.050)	(0.058)	(0.066)	(0.059)	(0.043)
Teachers	1044	1044	1044	1044	1044	1044	994	994
Students	78,458	78,458	78,458	78,458	78,458	78,458	65,632	65,632
Mentor Controls	X	X	X	X	X	X	X	X
Mentee Controls	X	X	X	X	X	X	X	X

Note: ELA = English Language Arts; VA = value added. Math models come from column 10 of Table 3, while ELA models come from column 10 of Table 4. Values of delta represent the amount of sorting on unobservables as a proportion of the observed sorting on observable variables, as described in Section 4 (and developed in Oster, 2018). Standard errors in columns 1 and 7 are clustered at the teacher level (as in Tables 3 and 4), while standard errors in remaining columns are calculated from 500 bootstrapped samples clustered at the teacher level. *P*-values from two-sided *t*-test: $+p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

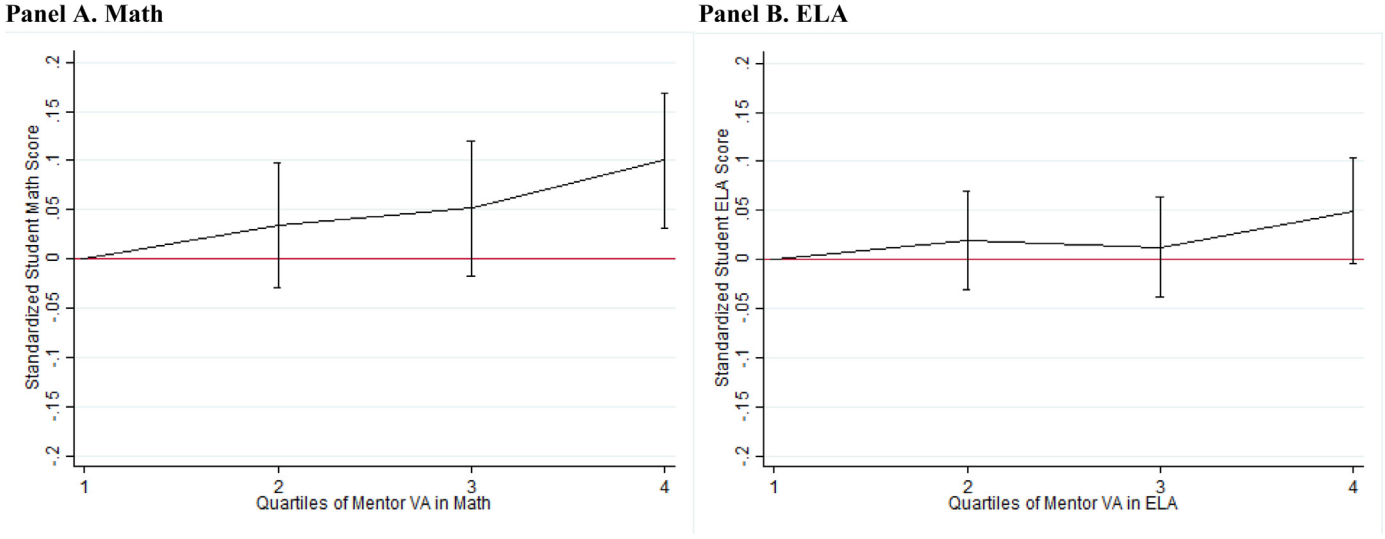

Fig. 2. Marginal Effects on Student Achievement by Quartile of Mentor Value Added (First-Year Teachers Only).

to this point and plot to estimated effects (relative to the lowest quartile of mentor value added) in Fig. 2. [32] These quartile indicators explain a significant portion of the variance in student test performance in math ($F = 2.83$) but not ELA ($F = 1.15$), and in both subjects, the positive relationships appear to be driven by mentor teachers in the top quartile of the distribution, and there are not significant differences between any of the top three quartiles.

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.[33]

This study suggests one specific mechanism through which this can occur. In fact, Fig. 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 non-random 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 student teachers to more effective mentor teachers appears to be a sensible low-cost approach to inducing marginal improvements in beginning teacher quality. States and districts are beginning to consider policies that take this approach; for instance, the Washington state legislature recently passed legislation (E2SHB 1139) stating that "Student teacher mentors should be highly effective as evidenced by… their most recent comprehensive performance evaluation", while Louisiana (Goldhaber and White, 2019) and Spokane (WA) Public Schools (Goldhaber et al., 2018) have also introduced poli-

[32] See Table A5 in the appendix for the point estimates from these models.

[33] 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/

cies that promote the selection of more effective mentors for student teacher placements.[34]

However, there are some clear challenges inherent in this policy recommendation. As discussed in Goldhaber et al. (2019), national data suggest that the average mentor teacher receives just over $200 in compensation per student teacher they host (Fives et al., 2016). In contrast, the average third-year teacher in Washington State is paid $3500 more than the average first-year teacher. Given the result discussed above—that assignment to a highly-effective mentor teacher, relative to an average mentor teacher, is the same magnitude as the average difference in effectiveness between 3rd-year teachers and 1st-year teachers—and the fact that only about 3% of teachers in Washington serve as a mentor teacher in a given year (Goldhaber et al., 2018), our conclusion is that there is significant scope for change in mentor teacher placements and that TEPs and districts are likely making a substantial underinvestment in mentor teachers.

Another concern with this policy recommendation is that there may not be enough highly-effective mentors to recruit to serve as mentor teachers. But as we demonstrate in Goldhaber et al. (2019), there are about the same number of highly-effective teachers (i.e., more than two standard deviations above average) who currently teach within 50 miles of a TEP and do not currently serve as mentor teachers as there are teachers who do serve as a mentor teacher in a given year, which suggests that this is not a major concern. That said, a lingering concern is that identifying high-quality mentors as well as restricting the ability to host student teachers to a subset of a school's teachers (i.e., those that are deemed to be "high-quality") could cause problems in a profession that has regularly been found to oppose differentiating or rewarding teachers by performance in any way (e.g. Goldhaber et al., 2011).

Our 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.

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 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. 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.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.labeco.2019.101792.

References

Aaronson, D., Barrow, L., Sander, W., 2007. Teachers and student achievement in the Chicago public high schools. J. Labor Econ. 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. Aust. J. Teach. Educ. 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. J. Polit. Econ. 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. Am. Econ. Rev. 98 (2), 345–350.

Anderson, L.M., Stillman, J.A., 2013. Student teaching's contribution to preservice teacher development. Rev. Educ. Res. 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. J. Manag. Stud. 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. Educ. Res. 44 (2), 77–86.

Blazar, D., 2015. Effective teaching in elementary mathematics: identifying classroom practices that support student achievement. Econ. Educ. Rev. 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. J. Policy Anal. Manage. 24 (1), 113–132.

Bozeman, B., Feeney, M.K., 2007. Toward a useful theory of mentoring. Adm. Soc. 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/ .

Carrell, S.E., Hoekstra, M., Kuka, E., 2018. The long-run effects of disruptive peers. Am. Econ. Rev. 108 (11), 3377–3415.

Chetty, R., Friedman, J.N., Rockoff, J.E., 2014a. Measuring the impacts of teachers I: evaluating bias in teacher value-added estimates. Am. Econ. Rev. 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. Am. Econ. Rev. 104 (9), 2633–2679.

Chetty, R., Friedman, J.N., Rockoff, J., 2016. Using lagged outcomes to evaluate bias in value-added models. Am. Econ. Rev. 106 (5), 393–399.

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

Cowan, J., Goldhaber, D., Hayes, K., Theobald, R., 2016. Missing elements in the discussion of teacher shortages. Educ. Res. 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 140618.

Crosby, O., 2002. Career training, credentials—and a paycheck in your pocket. Occup. Outlook Q. Bureau of Labor Statistics, U.S. Department of Labor.

[34] See http://lawfilesext.leg.wa.gov/biennium/2019-20/Pdf/Bills/Session%20Laws/House/1139-S2.SL.pdf, Section 202.

Eby, L.T., Rhodes, J.E., Allen, T.D., 2007. Definition and evolution of mentoring. In: Allen, T.D., Eby, L.T. (Eds.), The Blackwell handbook of mentoring: A multiple Perspectives Approach. Blackwell, Oxford, England, pp. 7–20. 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. J. Teach. Educ. 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. Mentor. Tutor. Partn. Learn. 10 (1), 47–55. doi:10.1080/13611260220133144.

Glazerman, S., Mayer, D., Decker, P., 2006. Alternative routes to teaching: the impacts of teach for America on student achievement and other outcomes. J. Policy Anal. Manag. 25 (1), 75–96.

Goldhaber, D., 2007. Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? J. Hum. Resour. 42 (4), 765–794.

Goldhaber, D.D., Brewer, D.J., Anderson, D.J., 1999. A three-way error components analysis of educational productivity. Educ. Econ. 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. J. Res. Educ. Eff. 8 (1), 8–34.

Goldhaber, D., DeArmond, M., DeBurgomaster, S., 2011. Teacher attitudes about compensation reform: implications for reform implementation. ILR Rev. 64 (3), 441–463.

Goldhaber, D., Gratz, T., 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. Econ. Educ. Rev. 61, 112–129.

Goldhaber, D., Grout, C., Harmon, K., & Theobald, R. (2018). A practical guide to challenges and opportunities in student teaching: a school district's perspective. CALDER Working Paper No. 205-1018-1.

Goldhaber, D., Krieg, J., Naito, N., Theobald, R., 2019. Making the most of student teaching: the importance of mentors and scope of change. Educ. Finance Policy 1–21. Accepted Manuscriptaccessed from https://www.mitpressjournals.org/doi/pdf/10.1162/edfp_a_00305 .

Goldhaber, D., Krieg, J., Theobald, R., 2014. Knocking on the door to the teaching profession? Modeling the entry of prospective teachers into the workforce. Econ. Educ. Rev. 42, 106–124.

Goldhaber, D., Krieg, J.M., Theobald, R., 2017b. Does the match matter? Exploring whether student teaching experiences affect teacher effectiveness. Am. Educ. Res. J. 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 No. 207-1118-1.

Goldhaber, D., Lavery, L., Theobald, R., 2015. Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. Educ. Res. 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. Econ. Educ. Rev. 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. Am. Educ. Res. J. 55 (1), 171–201.

Goldhaber, D., White, J., 2019. Aspiring teachers deserve time with a mentor before going it along. The Hill. July 18, 2019. Accessed from https://thehill.com/opinion/education/453403-aspiring-teachers-deserve-time-with-a-mentor-before-going-it-alone.

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

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. Educ. Res. 43 (6), 293–303. doi:10.3102/0013189x14544542.

Hargreaves, A., Jacka, N., 1995. Induction or seduction? Postmodern patterns of preparing to teach. Peabody J. Educ. 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. Am. Econ. 1 (4), 85–108. doi:10.1257/app.1.4.85.

Jacob, B.A., Lefgren, L., 2008. Can principals identify effective teachers? Evidence on subjective performance evaluation in education. J. Labor Econ. 26 (1), 101–136.

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. J. Labor Econ. 31 (2), S173–S202. doi:10.1086/669060, 2nd Ser.

Koedel, C., Mihaly, K., Rockoff, J.E., 2015. Value-added modeling: a review. Econ. Educ. Rev. 47, 180–195. doi:10.1016/j.econedurev.2015.01.006.

Krieg, J., Goldhaber, D., Theobald, R., 2019. Teacher candidate apprenticeships: assessing the who and where of student teaching. J. Teach. Educ. doi:10.1177/0022487119858983.

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. Educ. Eval. Policy Anal. 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. Educ. Finance Policy 12 (2), 241–279.

Lee, D.S., 2009. Training, wages, and sample selection: estimating sharp bounds on treatment effects. Rev. Econ. Stud. 76 (3), 1071–1102.

Machin, S., McNally, S., 2008. The literacy hour. J. Public Econ. 92 (5–6), 1441–1462.

Matsko, K.K., Ronfeldt, M., Greene, H., Reininger, M., Brockman, S., 2019 forthcoming. The role of cooperating teachers in preparing pre-service teachers: a district-wide portrait. J. Teach. Educ..

McCaffrey, D.F., Sass, T.R., Lockwood, J.R., Mihaly, K., 2009. The intertemporal variability of teacher effect estimates. Educ. Finance Policy 4 (4), 572–606.

Meyer, S.J. (2016). Understanding field experiences in traditional teacher preparation programs in missouri. rel 2016-145. regional educational laboratory central.

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. Educ. Finance Policy 8 (4), 459–493.

Nye, B., Konstantopoulos, S., Hedges, L.V., 2004. How large are teacher effects. Educ. Eval. Policy Anal. 26 (3), 237–257. doi:10.3102/01623737026003237.

Oster, E., 2017. Unobservable selection and coefficient stability: theory and evidence. J. Bus. Econom. Statist. 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.G., Hanushek, E.A., Kain, J.F., 2005. Teachers, schools, and academic achievement. Econometrica 73 (2), 417–458.

Rockoff, J.E., 2004. The impact of individual teachers on student achievement: evidence from panel data. Am. Econ. Rev. 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

Ronfeldt, M., Brockman, S., Campbell, S., 2018a. Does cooperating teachers' instructional effectiveness improve preservice teachers' future performance. Educ. Res..

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.

Rothstein, J., 2009. Student sorting and bias in value-added estimation: selection on observables and unobservables. Educ. Finance 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. Am. J. Educ. 115 (1), 139–167 2008.

Stamm, M., Buddeberg-Fischer, B., 2011. The impact of mentoring during postgraduate training on doctors' career success. Med. Educ. 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 No. 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. Econ. Educ. Rev. 66, 191–205.

Yendol-Hoppey, D., 2007. Mentor teachers' work with prospective teachers in a newly formed professional development school: two illustrations. Teach. Coll. Rec. 109 (3), 669–698.

Zeichner, K.M., Gore, J.M., 1990. Teacher socialisation. In: Houston, W.R. (Ed.), Handbook of Research On Teacher Education. Macmillan, New York.