NATIONAL
CENTER for ANALYSIS of LONGITUDINAL DATA in EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

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

AIR

Identifying Promising Clinical Placements Using Administrative Data: Preliminary Results From ISTI Placement Initiative Pilot

Matthew Ronfeldt
Dan Goldhaber
James Cowan
Emanuele Bardelli
Joy Johnson
Christopher Daniel Tien

WORKING PAPER 189 • April 2018

Identifying Promising Clinical Placements Using Administrative Data: Preliminary Results From ISTI Placement Initiative Pilot

Matthew Ronfeldt
University of Michigan

Dan Goldhaber
American Institutes for Research/CALDER
University of Washington

James Cowan
American Institutes for Research/CALDER

Emanuele Bardelli
University of Michigan

Joy Johnson
University of Michigan

Christopher Daniel Tien
University of Washington

Contents

Acknowledgements

We appreciate the generous financial support that was provided for this research by the Bill and Melinda Gates Foundation, IES grant R305C120008, and 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 the foundations supporting CALDER, see http://caldercenter.org/about-calder. Emanuele Bardelli received pre-doctoral support from the Institute of Education Sciences (IES), U.S. Department of Education (PR/Award R305B150012). We also appreciate comments from Susanna Loeb on an earlier draft of this paper. This project would not have been possible without the participation of Tennessee Tech University (TTU) and data provided by the Tennessee Department of Education. We are particularly indebted to Julie Baker from TTU for all the help on the ground in carrying out the experiment. Please note that the views expressed are those of the authors and do not necessarily reflect those of this study's sponsors, TTU, the Tennessee Department of Education, or the institutions to which the authors are affiliated.

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

Identifying Promising Clinical Placements Using Administrative Data: Preliminary Results From ISTI Placement Initiative Pilot

Matthew Ronfeldt, Dan Goldhaber, James Cowan, Emanuele Bardelli, Joy Johnson, Christopher Daniel Tien

CALDER Working Paper No. 189
April 2018

Abstract

Improving the skill set of new teachers is a potentially important policy lever to increase student achievement. There are, however, doubts about the extent to which teacher education programs contribute to the development of teachers' skills. In this paper, we describe findings from an experiment designed to test whether one aspect of teacher education – the assignment of teacher candidates to student teaching internships – has a causal impact on the type of coaching they receive and their feelings of preparedness to teach. Specifically, we use administrative data on potential student teaching placements to predict those placements more likely to be promising, and then we randomly assign teacher candidates from one large program from Tennessee into either more promising (high index) or less promising (low index) internships. Based on data from a post-student teaching survey, we find consistently strong evidence of large effects of being assigned to high versus low index placements, particularly in terms of teacher candidates' perceptions of the quality of instruction of their cooperating teachers and the quantity and quality of the coaching that they received. And while not as large or consistently significant, we also found that teacher candidates in high versus low placements are more likely to report better working conditions in their placement schools, higher quality collaboration among teachers, more opportunities to learn to teach, and feeling better prepared to teach. Our findings provide evidence that teacher education can have a causal effect on the development of teacher candidates and it offers practical implications for programs and districts about how to use administrative data to inform internship placement decisions.

INTRODUCTION

Preservice teacher preparation has recently received increased attention and scrutiny. A large body of quasi-experimental evidence finds that the effects of in-service teachers on student achievement are both dramatic (relative to other schooling variables) and heterogeneous.[1] Thus, improving the skill set of new teachers is a potentially important policy lever to increase student achievement. There are, however, doubts about the extent to which teacher education programs (TEPs) contribute to the development of teachers' skills, or whether such programs can improve teacher education.[2]

Progress is being made. In the past decade, a growing number of correlational studies have begun to identify features of teacher preparation, and especially clinical preparation, that are positively associated with workforce outcomes, including measures of teaching effectiveness and retention (Boyd et al., 2009; Goldhaber et al., 2017; Krieg et al., 2016; Ronfeldt, 2012, 2015). These studies found, for example, that new graduates are more instructionally effective when they learn to teach in schools with less teacher turnover and with more instructionally effective cooperating teachers (Ronfeldt, 2012; Ronfeldt, Brockman, & Campbell, under review). While these positive associations suggest that certain features of preparation improve workforce outcomes, we cannot draw causal conclusions from these studies based on observational data, since the findings could reflect several forms of selection (Goldhaber and Ronfeldt, forthcoming).[3]

This paper describes what we believe is the largest randomized controlled trial focused on features of teacher preparation. The Improving Student Teaching Initiative (ISTI) is an experiment designed to supply causal evidence on the value of placements in higher quality student teaching internships. The findings are based on the pilot year in which ISTI was implemented in one large TEP at Tennessee Tech University (TTU).[4] As described in more detail below, we used administrative data on potential student teaching

[1] For more on the variation of teacher effectiveness or effects of teachers on student outcomes, see (Aaronson et al., 2007; Chetty et al., 2014; Goldhaber et al., 1999, 2018; Jackson, 2018; Rivken et al., 2005).
[2] See Goldhaber (forthcoming) and Peck and McDonald (2013) for more background on some of the critiques leveled at TEPs and on the information (and culture) that might be necessary for TEP improvement. Also see von Hippel and Bellows (2018), who conclude that there is little variation in the effectiveness of teachers across TEPs.
[3] For instance, teacher candidates with better skills might be more likely to be matched to types of internship schools or cooperating teachers during student teaching.
[4] ISTI now includes eight TEPs across three states.

placements to predict those placements more likely to be promising. The features of schools and teachers that we used in the algorithm were chosen based on evidence from prior nonexperimental literature on the associations between aspects of clinical placements and in-service teacher and student outcomes. After splitting potential placements into two groups—those we expected to be more promising (high index) and less promising (low index)—we randomly assigned teacher candidates to one group of placements or the other.

Drawing on data from a post-student teaching survey, we tested whether candidates assigned to high-index placements reported better quality clinical placements and more opportunities to learn to teach, and felt more instructionally prepared than candidates assigned to low-index placements. We found some consistently strong evidence of large effects of being assigned to high versus low placements, particularly in terms of teacher candidates' perceptions of the quality of instruction of their cooperating teachers and the quantity and quality of the coaching that they received. And while not as large or consistently significant, we also found that teacher candidates in high versus low placements are more likely to report better working conditions in their placement schools, higher quality collaboration among teachers, more opportunities to learn to teach, and feeling better prepared to teach.

Our findings offer practical implications for TEPs and districts that work with them to assign clinical placements: they can use existing information collected regularly to guide placement decisions. The index that we constructed based on administrative data appears to be a relatively low-cost, evidence-based method for improving clinical preparation. It will be important to demonstrate that this method predicts better workforce outcomes in the first year of teaching. Based on the work we now have in progress, we believe it is nontrivial that those assigned to high-index placements report substantially better opportunities for learning to teach. First, it provides proof of concept that the placement index indeed identifies more promising clinical placements. Second, because the index is based on literature about features of preparation associated with better workforce outcomes, ISTI is a promising example of how researchers, practitioners, and policy makers can use existing correlational evidence to design randomized controlled trials to study the causal effects of different kinds of preparation.

LITERATURE REVIEW

At the turn of the present century, few studies had linked preservice teacher preparation to the workforce outcomes of graduates. However, many studies have since compared workforce outcomes of teachers who enter the profession through different pathways (e.g., alternative versus traditional)[5] or TEPs.[6] Some of these studies provide evidence that certain programs or pathways may be related to better outcomes on average for graduates, but they fail to identify the specific causal mechanisms. More recently, researchers have focused on identifying specific features of preparation, cutting across TEPs, that predict better workforce outcomes for graduates. Because the emphasis of ISTI is to design clinical preparation initiatives that are likely to improve workforce outcomes, our review focuses primarily on this more recent body of literature, and particularly on studies that consider specific features of clinical preparation.

One of the first large-scale studies linking many features of clinical preparation to workforce outcomes was the New York City (NYC) Teacher Pathways project. Specifically, Boyd et al. (2009) collected survey data on recent graduates from all major providers in NYC, as well as from a review of program documentation and interviews with program leaders. They then linked this rich set of information on preparation to graduates' effectiveness at raising their students' achievement during their early years of teaching in NYC. Among other results, Boyd et al. (2009) found that graduates were more effective at raising student achievement when they graduated from programs that exert more oversight over student teaching, including more supervision and more alignment with methods coursework; and when programs required a capstone project that relates experiences of student teaching to other aspects of the preparation that candidates receive. The authors also found that graduates had better achievement gains when they reported more preservice opportunities to practice aspects of teaching. Although not explicitly about clinical

[5] See, for instance, Boyd et al. (2006), Constantine et al. (2009), Goldhaber and Brewer (2000), Kane et al. (2008), and Glazerman et al. (2006).

[6] For results on the variation in the effectiveness of EPP completers from individual states, see Boyd et al. (2009) for New York City, Goldhaber et al. (2013) for Washington, Koedel et al. (2015) for Missouri, Mihaly et al. (2013) for Florida, Ronfeldt and Campbell (2016) for Tennessee, and von Hippel et al. (2016) for Texas. For a recent re-analysis of the empirical findings across sites, see von Hippel and Bellows (2018). The literature on differences in workforce outcomes is significant and is summarized by Goldhaber (2011).

experiences, the majority of opportunities to practice teaching likely occurs during preservice student teaching and other kinds of clinical experiences.

Because these results identified clinical experiences as a promising aspect of preparation, Ronfeldt (2012) drew upon the same NYC Teacher Pathways database to test whether the characteristics of the schools in which teacher candidates completed their preservice student teaching experiences were related to better achievement gains or job retention among recent graduates. The author found that candidates had better achievement gains and job retention when they completed student teaching in schools with better average rates of teacher retention (higher stay ratio). Based on survey responses, graduates in these kinds of schools reported better conditions for teaching and learning to teach, including better school leadership, more support, and better collaboration among faculty.

Building on these studies, Goldhaber et al. (2017) and Ronfeldt (2015) tested whether the stay ratio or other characteristics of placement schools were related to better student achievement gains or teacher retention. These studies generally supported the finding that teacher candidates have higher future value added (i.e., regression adjusted student achievement) and retention when they complete their student teaching in schools with higher stay ratios and measures of collaboration. Additionally, Ronfeldt (2015) found that teacher candidates have higher future value added when they intern in schools that have a history of strong achievement gains, and Goldhaber et al. (2017) found that future value added is higher if there is a closer demographic match of students between the school in which an internship is completed and a teacher's first teaching assignment.[7]

One of the most central figures during preservice student teaching is the cooperating or mentor teacher (CT). CTs are the teachers of record who host and mentor the teacher candidates as they take on some or all of lead teaching responsibilities. While much has been written about CTs, especially in the 1970s

[7] Specifically, they measured the difference in the proportion of students qualifying for free- or reduced-price lunches or underrepresented minority students between internship schools and hiring schools. Such a relationship was also hypothesized and tested in Ronfeldt (2015), who found a positive but not statistically significant relationship. Goldhaber et al. (2017) extended the Ronfeldt analysis to consider nonlinearities in the relationship, finding that the effects are disproportionately significant for teacher candidates in schools with compositions that differ significantly from their internship placements.

and 1980s, most accounts have described their beliefs, roles, and actions. In an oft-cited chapter reviewing earlier literature on teacher socialization, Zeichner and Gore (1990) concluded that cooperating teachers, as well as the schools in which they work, are powerful socializing agents but not always in the best direction. Most literature suggests that cooperating teachers tend to shape the beliefs and actions of teacher candidates to fall in line with their more custodial and managerial perspectives rather than the more progressive perspectives promoted by preparation programs. Although teacher education scholars often dismiss CTs as working against the kinds of preparation that they promote, they have continued to outsource a critical part of training for CTs. While most of the teacher education literature characterizes cooperating teachers and field placement schools as countering or obstructing the progressive preparation received by candidates early in their programs, Hargreaves and Jacka (1995) took another angle, characterizing progressive programs as "Disneyworlds" that prepare, or in their words "seduce," candidates for schools that do not actually exist. Either way, the point is that the literature suggests schools of education and P–12 schools often promote different forms of teaching and learning, and cooperating teachers (as well as teacher candidates) are often caught in the middle.

Another tension in the literature about cooperating teachers is whether they function primarily as models of instructional effectiveness or as coaches or mentors who provide feedback and support to teacher candidates. Most state policies focus on the former, often establishing minimum requirements for number of years of teaching experience and, in some cases, instructional effectiveness. For example, to serve as a cooperating teacher in Tennessee, a teacher must score at least a 3 out of 5 on the state evaluation rubric. We are not aware of any state policies for minimum requirements *for being an effective mentor* of preservice teachers. By contrast, most research on teacher education suggests that the mentoring role of cooperating teachers—including the kinds of feedback and support they provide—is most critical. For example, in a cross-national analysis of 26 mentor–novice pairs, Schwille (2008) identified 10 forms of mentoring provided by mentors with strong reputations. Likewise, Yendol-Hoppey (2007) argued for an elite and rare class of cooperating teachers as mentors who "teach interns much like they teach children" (p. 690) by encouraging constructivist practices, drawing on concepts of developmental appropriateness, encouraging risk-taking, and emphasizing problem-

solving. However, neither of these studies linked mentoring activities to the instructional performance of teacher candidates, so we cannot be certain that these forms of mentoring actually promote better instructional performance. The majority of the literature on cooperating teachers consists of small-scale case studies of individual programs focused on teacher candidates' beliefs and self-efficacy as outcomes, which are typically measured through interviews and surveys during or immediately following student teaching (Grossman, Ronfeldt, & Cohen, 2014; Zeichner & Gore, 1990).

Matsko et al. (forthcoming) recently examined more than 1,000 teacher candidates from programs across Chicago who registered to student teach in Chicago Public Schools. In addition to surveying all candidates before and after their student teaching experiences, the authors surveyed the cooperating teachers, and then linked the survey data to administrative and evaluation data on the cooperating teachers and their schools. Focusing on teacher candidates' survey-based perceptions of readiness to teach in different instructional areas at the end of preparation, Matsko et al. (forthcoming) found evidence that teacher candidates felt better prepared when they rated the instruction of their cooperating teachers as stronger and when their cooperating teachers had received better observation ratings on the district evaluation rubric. Teacher candidates also felt better prepared when their CTs offered more instructional support, frequent and adequate feedback, collaborative mentoring activities (e.g., co-teaching, co-planning), job search support, and a balance of autonomy and encouragement.

A limitation of this study is that it focused on teachers' self-reports of preparedness. Thus, while candidates might have felt better prepared, we cannot tell if their instruction was actually more effective as a result of their teacher preparation experiences. Attending to this limitation in a follow-up study, Ronfeldt et al. (2018) linked characteristics of cooperating teachers and clinical experiences more generally to district observation ratings of graduates in their first year of teaching. The authors found that recent graduates received stronger first-year observation ratings when their cooperating teachers (1) also received stronger observation ratings and (2) reported providing stronger coaching around the instructional domains evaluated on the district rubric.

Using statewide data in Tennessee, Ronfeldt, Brockman, and Campbell (under review) linked evaluation data on cooperating teachers to evaluation data on program completers who worked with them to examine whether recent program completers were more instructionally effective when they worked with CTs who were also more instructionally effective. Indeed, the authors found that program completers received better observation ratings during their first few years of teaching when their CTs also received better observation ratings. Compared with program completers whose CTs had average ratings of 3.0, program completers whose CTs had average ratings of 5.0 performed as though they had been teaching for an additional year. Likewise, program completers received stronger value-added to student achievement measures (VAMs) when their cooperating teachers also had stronger VAMs.[8]

Recent studies, as described in this section, provide promising evidence that teacher candidates are more instructionally effective when their cooperating teachers are more instructionally effective and when they are placed in schools with low teacher turnover and a history of better achievement gains. However, these findings are based on correlational evidence that cannot establish definitively causal conclusions. As Goldhaber and Ronfeldt (forthcoming) argue, many forms of selection among candidates—to programs, to field placements, and to first-year teaching positions—could account for the relationships observed in these correlational studies. These authors argue for experimental studies and especially for studies that randomly assign teacher candidates in preparation programs to different forms of preparation, which is consistent with the ISTI experiment.

OVERVIEW & INITIATIVE DESIGN

The ISTI placement initiative[9] is designed to assess the value of higher or lower quality internships for teacher candidates. Thus, it was necessary for each candidate in the study to have the possibility of at least

[8] Goldhaber et al. (2017) also link CTs to the later outcomes of those teacher candidates who become public school teachers. The authors found few statistically significant effects (other than a negative association between CTs with master's degrees and later value added of the teacher candidates they supervised), but the sample of CTs in the study was considerably smaller than Ronfeldt et al. (under review).

[9] Beyond the placement initiative, ISTI also includes a feedback initiative which focuses on improving the feedback that teacher candidates during their clinical experiences. For more information about the ISTI feedback initiative, see Goldhaber and Ronfeldt (forthcoming).

two internship placements (note that this does not necessarily imply different schools as it could be different CTs in the same school). Therefore, we asked program leaders at TTU to over-recruit CTs with whom they might want their teacher candidates to be placed. Drawing on evidence (described above) on the features of preparation likely to support better workforce outcomes, we then used state-level administrative data on the recruited CTs and their schools to create an index designed to identify placements that are more (high index) and less (low index) likely to promote better workforce outcomes. We then randomly assigned teacher candidates in this program to either low-index or high-index placements. We surveyed teacher candidates before and after their residency experiences. On the latter survey, we asked teacher candidates about their residency experiences and how well prepared to teach they felt in various instructional areas.[10] Based on these surveys, we compared teacher candidates from low- and high-index placements on their clinical experiences, on their opportunities to learn to teach, and on how well prepared to teach they felt after completing their clinical training. Specifically, we asked the following research questions:

1. Compared with teacher candidates assigned to low-index placements, do candidates assigned to high-index placements report:
    (a) that their cooperating teachers are more instructionally effective?
    (b) better working conditions in their field placement schools?
2. Compared with teacher candidates assigned to low-index placements, do candidates assigned to high-index placements report:
    (a) that their cooperating teachers provide more and higher quality coaching?
    (b) more opportunities to learn to teach?
3. Compared with teacher candidates assigned to low-index placements, do candidates assigned to high-index placements report feeling better prepared to teach?

Setting and Participants

[10] Our ultimate interest is to link graduates to workforce outcomes—including employment, retention, and measures of instructional effectiveness (including observation ratings and TVAAS scores).

The pilot for the ISTI placement initiative took place in collaboration with TTU, a large provider of teachers in the state of Tennessee.[11] TTU uses a residency model for their culminating clinical experience, placing their teacher candidates in the same residency classrooms and with the same cooperating teachers across an entire academic year. The present study focuses on 183 teacher candidates who completed their residency experiences during the 2016–2017 academic year.

Given the large number of teacher candidates that TTU prepares each year, program leaders collaborate with many district central offices across the state to help identify potential cooperating teachers and classrooms placements. In the year in which the pilot occurred, teacher candidates were placed in 32 districts across the state. While the placement process varies somewhat across teacher candidates and districts, early in the process TTU typically asks all teacher candidates to request specific districts or counties in which they would like to complete their residency experiences according to their geographic preferences and other needs. Once TTU knows how many teacher candidates are seeking placements in each district, it contacts district leaders to request that they recruit the same number of teachers to serve as cooperating teachers for the teacher candidates, who also match the grade level (e.g., early elementary, secondary) and subject area (e.g., special education, mathematics, foreign language) in which the teacher candidates are being endorsed. (See Appendix A1 for breakdowns by grade level and subject.) To serve as a cooperating teacher, Tennessee State law requires CTs to receive a summative rating of at least 3 (out of 5)[12] on the district evaluation. Therefore, district leaders must also consider information from teacher evaluations when recruiting teachers to serve as cooperating teachers.

Experimental Design of the Initiative and Experiment

A goal for ISTI was to design the intervention to align, as much as possible, with existing program practices and structures. In the pilot year, the placement process began in a similar way to that of TTU—by

[11] According to Title II data, TTU is the largest TEP in Tennessee, graduating about 300 teacher candidates in 2014–15 (the last year for which Title II data are available). Note that this is a larger number of teacher candidates than is included in the experiment we describe because some teacher candidates could not be randomized to CTs based on the way placements were done in particular programs in TTU and the availability of potential CTs in certain teaching specializations. The participation of teacher candidates in the experiment is described more extensively below.
[12] For school year 2015–2016, 95.42% of the teachers in Tennessee received a total observation score of 3 or better, and 51.35% received a score of 4 or better.

identifying the number of teacher candidates who needed placements in each district, and, more specifically, the number in each grade level and subject combination. Going forward, we will refer to these district-by-grade-by-subject combinations as "placement blocks" or "blocks," since these are the blocks in which randomization eventually occurred. After the number of teacher candidates in each placement block was identified, the next step of the placement process differed somewhat from typical procedures. Rather than have districts recruit the same number of cooperating teachers as teacher candidates in each placement block, to ensure we would have enough cooperating teachers for randomization to be successful and to offset any cooperating teacher attrition, TTU asked districts to over-recruit cooperating teachers. We suggested that districts aim to recruit twice as many cooperating teachers as teacher candidates per placement block, but this was often not possible, especially in smaller districts and high-needs endorsement areas where the pools of potential cooperating teachers were often limited. Several placement blocks contained only one teacher candidate (referred to as "singletons"—see Appendix A1), particularly in small districts and in less common endorsement areas (e.g., art, foreign language, science). To be included in the experiment, at least two cooperating teachers had to be recruited for these singleton blocks (one for each condition). Thus, we encouraged TTU—in their negotiations with district leaders—to prioritize over-recruitment, particularly in these singleton blocks. In the end, 412 cooperating teachers were recruited for 183 teacher candidates.

We then linked all recruited cooperating teachers to administrative and evaluation data on them and to the schools in which they worked. In determining what constituted high or low placements, we focused on five characteristics of clinical placements that are typically available in state administrative data systems and that research suggests are likely positively related to stronger workforce outcomes. Three of these variables are teacher characteristics: their observation ratings based on the district rubric, their scores on the Tennessee Value-Added Assessment System (TVAAS), and their number of years of teaching experience.[13] The other

[13] While the literature suggests that instructional effectiveness of program completers is related to cooperating teachers' instructional effectiveness (see Literature Review), evidence for the positive effects of the number of years of experience of cooperating teachers is less well established. We include this variable in our placement algorithm for two reasons. First, there is substantial evidence that number of years of teaching experience is positively related to instructional effectiveness, as measured by value added and observation ratings (e.g., Jiang & Sporte, 2016; Rivken et al., 2005; Rockoff, 2004). Second, most states set minimum requirements for number of years of teaching experience.

two variables are placement school characteristics: average teacher retention rates at the school level and average TVAAS scores at the school level. More details about each variable are provided below.

Cooperating Teacher Characteristics. As a part of Tennessee's First to the Top Act, the state established and implemented a teacher evaluation system during the 2011–2012 academic year. Under this system, teachers are evaluated on (a) growth in students' test scores, as measured by TVAAS, (b) student achievement on another selected measure,[14] and (c) classroom observation rubrics. With respect to observation ratings, the Tennessee State Board of Education adopted the Tennessee Educator Acceleration Model (TEAM) as the statewide observational rubric.[15] The TEAM rubric includes four domains—instruction, environment, planning, and professionalism—and several indicators are associated with each domain. Administrators are required to observe multiple domains during a classroom visit, except for the professionalism domain, which is evaluated only at the end of the year. Teachers are rated on a scale of 1 to 5 on each indicator within a domain, with 1 being "significantly below expectations," 3 "at expectations," and 5 "significantly above expectations." For our analysis, overall observation ratings are an average of a teacher's ratings across the four domains and observed lessons.

In addition to observation ratings, we also used a measure of a teacher's contributions to student learning, as measured by TVAAS scores (also referred to as value added in our placement index).[16] The Tennessee Department of Education provided us with TVAAS estimates by subject, grade, and year for each teacher assigned to a tested subject. The teacher-level TVAAS score is available for each teacher who has at least six valid student scores in the same grade and subject and is calculated as a composite weighed index of the average students' deviation from the statewide average score gain in each grade level and subject divided by its standard error. This procedure rescales the raw teacher-level TVAAS to a unit-less Normal Curve

[14] These measures can include ACT/SAT scores, early postsecondary assessments (e.g., Advanced Placement, International Baccalaureate, and dual credit exams), industry certifications, graduation rates, or other off-the-shelf assessments.

[15] The state evaluation policy allows districts to develop their own observation rubrics and apply for approval by the department of education for an alternate observation model. Nineteen districts (13.5%) received approval for these alternate models in 2014–2015.

[16] For more information, see page 38 of https://tvaas.sas.com/support/TVAAS-TechnicalDocumentation-2017.pdf.

Equivalent (NCE) measure.[17] Because teachers received different TVAAS scores for different tests, [18] grades, and subjects, we first standardized within test, grade, and subject. We then created a standardized composite TVAAS score for each teacher by averaging all of his/her TVAAS scores across tests, grades, and subjects. A potential limitation of this methodology is that it assumes comparability across subject areas, grade levels, and standardized tests; however, this methodology is similar to how Tennessee evaluates teachers who teach multiple subjects and grades. In addition to observation ratings and TVAAS scores, the state also provided information about the number of years of teaching experience for each teacher.

Placement School Characteristics. To calculate teacher retention at the school level, we began with the teacher evaluation data file; because it is state law that all teachers in the state be evaluated, evaluation files provide the most accurate way to identify all teachers in the state and where they teach. In each school in each year, we then calculated the proportion of teachers in that school in year (t) that remained in the subsequent year (t+1). For the present study, we calculated this stay ratio for all schools from 2012 to 2013, 2013 to 2014, and 2014 to 2015. We then averaged these to get 3-year stay ratios for each school.

In addition to teacher-level TVAAS scores, the state of Tennessee also calculates school-level TVAAS scores for each school in each year. These estimates report a single TVAAS score for each school site, which is the composite for teacher-level estimates weighted by the number of students in each grade, subject, and school. As with the school-level turnover measure, we averaged school-level TVAAS scores from 2012–2013 to 2014–2015.

Placement Procedures. We then proceeded by standardizing all five variables (described above) using the entire population of teachers and schools in Tennessee. To create the overall placement index, we calculated separately a teacher index and a school index, using the following equations:

$$\text{TchIndex} = \text{TchObsRating}^{\text{STD}} + \text{TchTVAAS}^{\text{STD}} + \text{Experience}^{\text{STD}} \quad (1)$$

$$\text{SchIndex} = \text{SchStayRatio}^{\text{STD}} + \text{SchTVAAS}^{\text{STD}} \quad (2)$$

[17] For more information about how the state calculates teacher- and school-level TVAAS scores, see https://tvaas.sas.com/support/TVAAS-TechnicalDocumentation-2017.pdf.
[18] TVAAS estimates were taken from three different standardized tests: SAT10, TCAP, and end-of-course exams.

In cases where any of the five variables were missing for teacher candidates, we replaced missing information with the mean score on that variable for the district-by-grade-by-subject block. Finally, we created the overall index by weighting the teacher index at 75% and the school index at 25%[19]:

$$\text{PlacementIndex} = 0.75 \cdot \text{TchIndex} + 0.25 \cdot \text{SchIndex} \quad (3)$$

We then split the recruited cooperating teachers into two lists—low index and high index—using the district-grade-subject block median as the cutoff. Cooperating teachers who had index scores at the block median were assigned to the high index.[20] In the end, there were 219 cooperating teachers in the high index and 193 cooperating teachers in the low index (the differential results from assigning placements at the median to the high index). Figure 1 shows the distribution of overall index scores among placements assigned to the low-index (red) and high-index (blue) lists.

After the cooperating teachers and school placement lists were finalized, we randomized teacher candidates within district-grade-subject blocks to either the high-index group or the low-index group. Specifically, after candidates were assigned a random number, we used these random numbers to establish ranks within district-subject-grade blocks. For blocks with even numbers of teacher candidates, the half of the block with the highest ranks were assigned to the high-index placement list, and the half with the lowest ranks were assigned to the low-index placement list. For blocks with odd numbers of teacher candidates greater than one, we repeated the same procedures except we assigned teacher candidates at the median rank to the high-index list. Our reason for doing this was that some partnering programs had concerns that their teacher candidates might end up with worse average placements than usual; this practice ensured more teacher candidates ended up in the high-index group than would be expected by chance. Because this changed the odds of placement somewhat, we included probability weights in our main model specifications (elaborated below). For blocks with only one teacher candidate, however, we randomly assigned them to either the low- or high-index condition.

[19] We decided to weight the teacher index more heavily because teacher candidates spend most of their residency time in their placement classrooms and with their cooperating teachers, so we assumed the characteristics of cooperating teachers would be more influential than school characteristics.

[20] This provided more cooperating teacher options in the high-index condition to accommodate our randomization procedures, which also assigned somewhat more teacher candidates to the high-index condition, as elaborated below.

At this juncture, it is important to underscore that we were not assigning teacher candidates to specific cooperating teachers and placement schools but instead to one list or the other, each of which is based on both cooperating teacher and school factors. Thus, we created two placement lists based on index scores (high and low index) and two lists of teacher candidates (those randomly assigned to the high or low index placement lists). We then returned the lists to TTU and asked program leaders to make final placements by placing teacher candidates who had been randomly assigned to the high-index group only to placements on the high-index list, and by placing teacher candidates who had been randomly assigned to the low-index group only to placements on the low-index list. During this final step, TTU leaders were blind to the conditions (rather than label lists as "high" or "low" index, we used "list A" and "list B"). More specifically, program leaders at TTU received two Excel spreadsheets (A and B). Spreadsheet A included two tabs – one for "Cooperating Teacher List A" and one for "Teacher Candidate List A." We asked program leaders at TTU to place each teacher candidate on List A to one of the cooperating teachers on List A. We asked them to repeat these steps separately on Spreadsheet B.

Sample

A total of 183 teacher candidates were randomized within blocks to either the high-index (n = 97) or low-index (n = 86) placement condition making up our Intent to Treat (ITT) sample. Of these individuals, 123 responded to our survey, equating to a 67% (123/183) response rate (discussed below). Because the outcome measures for this paper are based on survey responses, this group makes up our ITT *analytic* sample. Given the attrition of teacher candidates from the program (e.g., Praxis scores did not permit completing residency), cooperating teacher attrition, and other factors (see prior section), 175 of the 183 initially randomized teacher candidates were assigned to a cooperating teacher who was originally selected to participate in the placement initiative for the duration of the residency (our Local Average Treatment Effect [LATE] group). Of these, 113 responded to the survey, making up the LATE analytic sample for this paper.

Table 1 summarizes the characteristics of all cooperating teachers and their schools who were assigned to teacher candidates (Column 1), of those assigned to the high-index group (Column 2), and of those assigned to the low-index group (Column 3). This sample of cooperating teachers had average

observation ratings of 4.12, TVAAS scores of 0.07, and almost 13 years of teaching experience. About 87% of teachers in cooperating teachers' schools were retained from one year to the next, and these schools had 0.02 school-level TVAAS scores. Over the same period, the statewide average observation rating was 3.86 (standard deviation [s.d.] = .58), average TVAAS score was -0.05 (s.d. = 0.82), and average number of years of teaching experience was 12.62 (s.d. = 10.11). The average statewide retention rate was 82.6% (s.d. = 12.3), and the average school-level TVAAS score was -.03 (s.d. = .59). Across variables, the high-index means were higher than the low-index means.

Table 1 (Column 4) reports the standardized treatment contrast between high- and low-index placements. Teacher candidates assigned to high-index placements had between 0.7 and 0.9 standard deviation units higher mean scores on cooperating teacher characteristics than the teacher candidates assigned to low-index placements; all differences were statistically significant. Differences on the school variables were nonsignificant and much smaller in magnitude; these results are not all that surprising given that, in the construction of the placement index, school characteristics were weighted at 25% and cooperating teacher characteristics at 75%.

Attrition Analysis

Table 2 reports the block-adjusted attrition rates. Baseline attrition for actual placement with a cooperating teacher in the initiative was 17.6 percentage points and differential attrition for teacher candidates in the high-index group was -6.32 percentage points. The nonresponse rate on the survey was 27.6 percentage points and the differential nonresponse rate for high-index placement was 6.78 percentage points. These attrition rates are within the "tolerable threat of bias under optimistic assumptions" and low attrition rate according to What Works Clearinghouse's guidelines (What Works Clearinghouse, 2017, p. 11).

Balance Checks

Table 3 (Panel A) summarizes the sample of all teacher candidates (Column 1) who were initially assigned to either condition, those assigned to low-index placements (Column 2), and those assigned to high-index placements (Column 3). The information comes from program administrative data collected on all teacher candidates. The sample of teacher candidates is mostly White (96%) and female (90%), has a mean

cumulative grade point average (GPA) of 3.6, and has a mean inclusive GPA of 3.4. Across all variables, the low-index and high-index groups were statistically similar prior to being their teacher residency; an omnibus test confirmed the null hypothesis that no significant differences existed between conditions and, thus, that randomization was successful (see Table 3, Panel B).[21]

Survey and Response Rates

We administered online surveys to all teacher candidates after their residency experiences. The post-placement survey was administered in the final weeks of residency in April 2017 and was open for 5 weeks, extending past graduation. A total of 126 teacher candidates in our ITT sample completed the post-placement survey, for a 67% response rate. To incentivize participation in the post-placement survey, teacher candidates who responded to the survey received electronically a $10 gift card for Amazon.

Using administrative data provided by TTU, we tested whether the ITT sample who took the survey differed from those in the ITT sample who did not take the survey. Using tests of proportions and *t*-tests, we found no differences between these groups in terms percent White, percent female, cumulative GPA, mean Praxis score, and mean number of times each Praxis test was taken. See Appendix Table A2 for results from these tests.

The post-placement survey posed questions related to teacher candidates' placement experiences, including their feelings about preparation to teach; perceptions of their cooperating teacher, field supervisor, and placement school; opportunities to practice particular tasks of teaching; and types of feedback received. Finally, we asked about teacher candidates' post-graduation plans.

Measures

We analyzed survey items using confirmatory factor analyses (CFA). This technique allowed us to estimate a single scale score from multiple survey items. We prefer this method over a more customary principal component factor analysis for two reasons. First, CFA separately estimates measurement error from the idiosyncratic error term (Kline, 2015). This allowed us to account in the models for potential sources of

[21] As a robustness check, we also tested for teacher candidate characteristics that they reported on the pre-placement survey. The high-index and low-index groups are balanced on all pre-placement survey items.

measurement bias (i.e., omitted variables). For example, teacher candidates were asked to rate the instructional effectiveness of their cooperating teachers on differentiating instruction and using assessment data to inform instruction. We found evidence that the error terms for these two items are correlated, indicating that a competing construct alongside cooperating teacher's instructional quality might be at play when teacher candidates respond to the survey. Second, CFA provides fit indices that we used to assess whether our model was a good fit for the data. These statistics measure the deviation of the model covariance matrix from the observed covariance matrix (Kline, 2015).

We modeled each hypothesized latent construct in separate CFAs. To account for possible non-normality in the survey response items, we used the Satorra-Bentler adjustment (Satorra & Bentler, 2001) to calculate robust standard errors, fit statistics, and *p* values.[22] After fitting each hypothesized model, we used modification indices to guide us in selecting items that might share a common source of error covariance. Table A3 reports the fit indices for these final models. Overall, the models have good fit to the data following customary guidelines (Hu & Bentler, 1999). We then used the predicted factor scores from these models for all subsequent analyses. Appendix Table A4 provides more details about survey items included in each factor and their factor loadings.

Analysis Methods

To compare the candidates assigned to low-index and high-index placements, we utilized the following base model specification:

$$Y_i = \beta_0 + \beta_1 \cdot \text{Placement}_i + \varepsilon_i \quad (4)$$

where Y_i is the predicted factor score of interest for candidate i, Placement_i is a dummy variable indicating whether candidate *i* was placed in a high-index placement. β_1 is the main coefficient of interest, and it captures the average difference in outcome variable Y_i between low-index and high-index placement groups.

We estimated several variations of this base model specification. In Model 1, we used the ITT sample to construct the Placement_i indicator. That is, teacher candidates were assigned a 1 if they were originally

[22] As a robustness check, we also fit these models using polytomous Rasch models that allow the measurement model to use ordinal items. Item loadings and *p* values are qualitatively equal to our preferred measurement model specification.

randomized in a high-index placement and a 0 if they were placed in a low-index placement regardless of their actual placement. In Model 2 (Local Average Treatment Effect [LATE]), we estimated actual field placement on intended placement, specifically:

$$Y_i = \beta_0 + \beta_1 \cdot \widehat{\text{Placement}}_i + \varepsilon_i \quad (5)$$

$$\text{Placement}_i = \alpha_0 + \alpha_1 \, \text{RandomizedList}_i + \eta_I \quad (6)$$

where RandomizedList_i is a dummy variable indicating the placement to which the teacher candidate was originally randomized, and Placement_i is the endogenous observed placement. Therefore, the coefficient β_1 is the Wald estimator of the treatment effects and can be read as the LATE on the teacher candidates who were induced to receive treatment by randomization.

As described above, we grouped candidates in blocks defined by licensure field, level, and geographical preferences to facilitate their random assignment to placement lists. This design necessarily resulted in several blocks containing a single candidate. To account for differences in the likelihood of treatment assignment across randomization blocks while retaining useful variation from candidates in singleton blocks, we weighted each observation by the inverse probability of being randomized to the high-index placement. This approach accounts for the different selection probabilities of candidates in blocks of different sizes, which may introduce an association between unobservable characteristics of candidates and treatment assignment. In particular, each candidate had the following placement probabilities:

$$p(\text{High Index}_i \mid \text{Block}_i) = \frac{1}{2} \qquad \text{when } N_{block} = 1 \text{ or } N_{block} \text{ is even} \quad (7)$$

$$p(\text{High Index}_i \mid \text{Block}_i) = \frac{1}{2} \cdot \frac{N_{Block}+1}{N_{Block}} \qquad \text{when } N_{block} \text{ is odd} \quad (8)$$

where High Index_i is the indicator for high-index placement for candidate i; Block_i is the randomization block for individual i; and N_{block} is the total number of candidates in randomization block i. We then calculated analytical weights as

$$w_i = \frac{1}{p(\text{High Index}_i \mid \text{Block}_i)} \qquad \text{when Placement}_i = 1 \quad (9)$$

$$w_i = 1 - \frac{1}{p(\text{High Index}_i \mid \text{Block}_i)} \qquad \text{when Placement}_i = 0 \quad (10)$$

where $p(\text{High Index}_i \mid \text{Block}_i)$ is the placement probability calculated in equations (7) and (8), and Placement_i is the intended placement for candidate i. We checked the robustness of this adjustment in Models 4 and 5 in Appendix Tables A5–A12. Model 4 reports the results of the naïve reduced-form, ordinary least squares regression without block adjustments. Model 5 reports the results of reduced-form models that include block-level fixed effects.

RESULTS

Table 4 summarizes main results from comparisons between teacher candidates assigned to low-index and high-index placements about the kinds and quality of clinical training they reported experiencing during residency. Models 1 and 2 summarize estimates from our ITT and LATE analyses respectively. Each residency factor measure listed in Table 4 is based on several survey questions; Appendix Tables A5–A11 summarize the results for each of the survey questions included in the factor measures.

Overall, results suggest that teacher candidates who were assigned to high-index placements reported better quality and more frequent clinical training than those assigned to low-index placements. Across all models and residency experience factors, estimates trended positively toward the high-index placement group, with differences ranging in magnitude from between 8% and 51% of a standard deviation unit. On all residency experience factor measures there were significant differences between groups on at least one model specification, except for the *Placement School Working Conditions* factor. Not surprisingly, estimates for the LATE analyses tended to be larger in magnitude and more often statistically significant.

RQ1: Compared with teacher candidates assigned to low-index placements, do candidates assigned to high-index placements report (a) that their cooperating teachers are more instructionally effective? (b) better working conditions in their field placement schools?

Teacher candidates in high-index placements reported that their cooperating teachers were more instructionally effective on average than teacher candidates in low-index placements, but differences were significant at the $p < 0.05$ level only for the LATE estimates. These results are consistent with what we would expect, given that three measures of instructional quality—observation ratings of cooperating teachers, number of years of experience, and TVAAS scores—were included in the placement index. As indicated in

Appendix Table A5, differences tended to be greatest in terms of teacher candidates' perceptions about how effectively their cooperating teachers assessed students and handled classroom management and discipline issues.

In contrast to factor scores related to cooperating teachers, differences between conditions on the one factor related to the placement schools (*Placement School Working Conditions*) were smaller in magnitude and not statistically significant, but generally still positive. This is not entirely surprising since school characteristics were weighted less than cooperating teacher characteristics in the construction of the placement index. That said, teacher candidates in the high-index placements were significantly more likely to agree with the following statement: "Teachers were willing to collaborate with me about instruction" (see Appendix Table A6). While estimates on most other survey items trended positive for those assigned to the high-placement condition, results were more mixed.

RQ2: Compared with teacher candidates assigned to low-index placements, do candidates assigned to high-index placements report (a) their cooperating teachers provide more and higher quality coaching and (b) more opportunities to learn to teach?

Teacher candidates in high-index placements also reported that their cooperating teachers offered more frequent feedback, engaged more frequently in various coaching activities, and were more effective coaches generally. In terms of feedback frequency, differences between low-index and high-index groups were greatest in terms of frequency in asking reflective questions and sharing specific data when providing feedback (see Appendix Table A7). Regarding frequency of coaching activities, the most pronounced difference between conditions was that teacher candidates in high-index placements reported that their cooperating teachers observed them teach far more often—by between 61% and 78% of a standard deviation unit (see Appendix Table A8). Teacher candidates in high-index placements also reported that their cooperating teachers analyzed student work with them much more frequently— by between 33% and 46% of a standard deviation unit.[23]

[23] Both survey questions were rated on a 7-point scale, and the standard deviations of these items were 1.25 and 2.08, respectively.

Consistent with the finding that high-index teacher candidates reported more frequent observations from cooperating teachers, teacher candidates in high-index placements were much more satisfied with the frequency of observations conducted by their cooperating teachers (by about one-half of a standard deviation unit across specifications). In terms of survey items included in the *Perceptions of Coaching Quality* factor, differences were greatest on this item (see Appendix Table A9).

If cooperating teachers in the high-index condition provided more and better coaching, then we might also expect their teacher candidates to report more opportunities to practice different teaching skills. Indeed, we found this to be the case. Table 5 summarizes results from analyses comparing teacher candidates assigned to low-index and high-index placements in terms of how many opportunities they had to practice different aspects of teaching during their residency experiences. Across different domains of instruction, candidates in the high-index group reported more opportunities to practice; across all measures and models, estimates were positive and ranged between 17% and 39% of a standard deviation unit. Differences were largest in magnitude and significant across specifications only for *Opportunities to Establish a Learning Environment*. Appendix Table A10 summarizes differences on the survey items that contributed to these factor measures. Of the 18 individual survey items included in these factors, differences were rarely significant but were consistently positive across main model specifications (Models 1–2) for 14 items. These trends are unlikely to be due to chance and suggest that teacher candidates in high-index placements consistently experienced more opportunities to practice teaching across instructional areas.

RQ3: Compared with teacher candidates assigned to low-index placements, do candidates assigned to high-index placements report feeling better prepared to teach?

We also compared teacher candidates assigned to high-index and low-index placements in terms of how well prepared to teach they felt at the end of their residency experiences. Table 6 summarizes differences we observed on four different factor measures. Although not statistically significant, estimates across all measures and models trended positive. *General Preparedness* is a factor measure based on a set of standard questions that we asked across all TEPs participating in ISTI. Teacher candidates in high-index placements tended to feel better prepared on this factor measure—by about one-third of a standard deviation unit—but differences were significant only at the $p < 0.1$ level. In terms of differences on individual survey items

(Appendix Table A11), differences were greatest in magnitude and significance in relation to handling classroom management and discipline issues. This finding is consistent with two earlier results: teacher candidates in high-index placements reported (1) observing their cooperating teachers manage classrooms substantially better and (2) substantially more opportunities to practice managing classrooms.

The other three factor measures—*Preparedness to Plan*, *Preparedness to Establish a Learning Environment*, and *Preparedness to Deliver Instruction*—are based on the rubric that TTU uses to assess its teacher candidates during their residency experiences. In the case of TTU, the clinical assessment rubric is the same rubric used by the state to evaluate in-service teachers (the TEAM rubric).[24] Consistent with prior results, estimates on all factor measures across all models trended positive but were not statistically significant. Also consistent with the trends related to classroom management, only one individual survey item had significant differences across model specifications: "Establish a classroom environment" (see Appendix Table A12). Although estimates tended not to be significant on individual survey items, out of the 24 items included in these four Preparedness factors, estimates trended positive across focal models (Models 1–2) on 22 of the items.

DISCUSSION AND CONCLUSIONS

ISTI is an example of how evidence from well-designed, earlier correlational studies can be used to design randomized controlled trials in teacher education that provide more definitively causal evidence on the effects on features of teacher education on the outcomes of teacher candidates. The findings should be interpreted with caution for two reasons: First, they are based on assessing the intervention from a single TEP, which appears to have implemented the intervention quite well; moreover this is a TEP that was able to over recruit CTs for the study. These are important caveats in thinking about the generalizability of the findings. Second, they are based entirely on perceptions garnered through surveys of teacher candidates. Ultimately, we would hope to see that ISTI affects the outcomes of teacher candidates who end up in the workforce (and their students), which is the subject of our ongoing research.

[24] The first three factors match three of the instructional domains that the TEAM rubric is designed to capture: *Planning, Establishing a Learning Environment,* and *Delivering Instruction*. The TEAM rubric also includes a *Professionalism* domain. We included survey items based on the indicators used for this domain but could not construct a reliable factor based on them.

The above cautions notwithstanding, we view the initial findings as quite promising. Relative to being assigned to the lower index group, teacher candidates assigned to the higher index group of CTs and schools rated their CTs as more instructionally effective on all eight of the items in the survey; suggested that their CTs provided more frequent coaching, feedback, and practice on 16 of the 18 items in the survey; and rated the quality of the coaching they received as higher on all nine of the items. Collectively, these were all highly significant and in some cases the effects are quite large. For instance, 65.6% of the teacher candidates in the high-index list reported that their cooperating teachers observed them daily, but only 50.9% of the teacher candidates in the low-index placement reported daily observations. Additionally, teacher candidates in high-index placements reported that when compared with their counterparts, their cooperating teachers offered more concrete suggestions (77.0% vs. 60.4%) and asked reflective questions (65.6% vs. 45.7%) often or all the time. Indeed, we believe this type of additional coaching is arguably beneficial regardless of whether it translates into statistically identifiable differences in the labor market outcomes of teacher candidates.

Teacher candidates perceived fewer differences in school working conditions, although findings were still generally positive for the high-index group. It is not clear whether this is because candidates have fewer opportunities to observe school culture more generally (as opposed to instruction of their specific CTs), or because the algorithm we used to determine assignment to the high- or low-index groups placed only a 25% weight on schoolwide measures. We also did not find consistently significant differences in feelings of preparedness between groups, but again, the teacher candidates in high-index placements tended to report (on 22 of the 24 items in the survey) feeling better prepared than those in low-index placements.

We plan to assess the effects of placements on candidate outcomes in future years, but the effects on candidate perceptions of coaching quality are also independently notable. Mentorship and coaching are important components of TEP missions. Using a small number of readily available indicators of school and teacher quality, our results suggest that programs can assign candidates to more effective coaches. Improved screening of mentors may help coordinators and policy makers overcome difficulties in scaling coaching programs (Kraft et al., 2018).

Our findings are very much in line with earlier nonexperimental evidence that placements with more instructionally effective mentors increased the instructional effectiveness of teacher candidates. This study is broadly supportive of the nonexperimental findings that teacher candidates assigned to higher rated CTs also rate them as more instructionally effective. The survey evidence in this paper also suggests mechanisms through which instructional effectiveness might be transmitted from CTs to teacher candidates (and then carried into the workforce). Prior studies hypothesized that teacher candidates who are assigned to more instructionally effective mentors improve by observing their CTs and mentors who model better teaching (Matsko et al., forthcoming). Our findings on candidates' perception of the instructional effectiveness of their CTs are consistent with this mechanism. However, we also observed a plausible second mechanism: teacher candidates assigned to more instructionally effective CTs also receive more frequent and better coaching (including feedback) and receive more opportunities to practice teaching.

That teachers who are more instructionally effective, based on both TVAAS scores and observational ratings, also appear to be more effective mentors in helping to supervise student teaching is not entirely surprising. One might expect that individuals who are better teachers of P–12 students will likely have some capacities that also translate to adult learners (teacher candidates); these individuals will also likely have better knowledge of quality teaching and as such, will be in a better position to teach this content to others. That said, in making cooperating teacher placements, many TEP leaders indicate that they prioritize recruiting individuals who they believe to be good coaches of teacher candidates over individuals who are effective teachers of P–12 students. Our results suggest that this distinction may be blurrier than is sometimes assumed. Importantly, if our findings hold up when we assess the relationship between index group assignment and teacher workforce outcomes, then this intervention and study would provide an evidence-based approach for selecting promising placements based on administrative data, and would make a strong case to school systems that the quality of placements is fundamental to the development of new teachers.

REFERENCES

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

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

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

Carinci, J., Meyer, S., & Jackson, C. (Eds.) (in press). *Linking Teacher Preparation Program Design and Implementation to Outcomes for Teachers and Students*. Charlotte, NC: Information Age Publishing.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633-79.

Constantine, J., Player, D., Silva, T., Hallgren, K., Grider, M., & Deke, J. (2009). An Evaluation of Teachers Trained through Different Routes to Certification. Final Report. NCEE 2009-4043. *National Center for Education Evaluation and Regional Assistance*.

Glazerman, S., Mayer, D., & Decker, P. (2006). Alternative routes to teaching: The impacts of Teach for America on student achievement and other outcomes. *Journal of Policy Analysis and Management*, *25*(1), 75-96.

Goldhaber, D. (2011). Licensure: Exploring the value of this gateway to the teacher workforce. In *Handbook of the Economics of Education* (Vol. 3, pp. 315-339). Elsevier.

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational evaluation and policy analysis*, *22*(2), 129-145.

Goldhaber, D. D., Brewer, D. J., & Anderson, D. J. (1999). A three-way error components analysis of educational productivity. *Education Economics*, *7*(3), 199-208.

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

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

Goldhaber, D. & Ronfeldt, M. (forthcoming). Toward causal evidence on effective teacher preparation. In J. S. Carinci, S. Meyer, & C. Jackson (Eds.). *Linking teacher preparation program design and implementation to outcomes for teachers and students.* Charlotte, NC: Information Age Publishing.

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

Hansen, B. B., & Bowers, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, *23*(2), 219–236.

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

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non-test score outcomes. *Journal of Political Economy*.

Jiang, J. & Sporte, S. (2016). *Teacher evaluation in Chicago: Differences in observation and value-added scores by teacher, student, and school characteristics*. Chicago, IL: University of Chicago Consortium on School Research.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education review*, *27*(6), 615-631.

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling* (4th ed.). Guilford Publications.

Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2015). Teacher preparation programs and teacher quality: Are there real differences across programs?. *Education Finance and Policy*, *10*(4), 508-534.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence. *Review of Educational Research*, Advance online publication.

Krieg, J. M., Theobald, R., & Goldhaber, D. (2016). A foot in the door: Exploring the role of student teaching assignments in teachers' initial job placements. *Educational Evaluation and Policy Analysis*, *38*(2), 364-388.

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

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

Peck, C. A., & McDonald, M. (2013). Creating "cultures of evidence" in teacher education: Context, policy, and practice in three high-data-use programs. *The New Educator*, *9*(1), 12-28.

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

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

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

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

Ronfeldt, M., Brockman, S. L., Campbell, S. L. (under review). Does cooperating teachers' instructional effectiveness improve preservice teachers' future performance? Revise and resubmit; *Educational Researcher*.

Ronfeldt, M., & Campbell, S. L. (2016). Evaluating teacher preparation using graduates' observational ratings. *Educational Evaluation and Policy Analysis*, *38*(4), 603-625.

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

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514.

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

von Hippel, P. T., & Bellows, L. (2018). How much does teacher quality vary across teacher preparation programs? Reanalyses from six states. *Economics of Education Review*.

von Hippel, P. T., Bellows, L., Osborne, C., Lincove, J. A., & Mills, N. (2016). Teacher quality differences between teacher preparation programs: How big? How reliable? Which programs are different?. *Economics of Education Review*, *53*, 31-45.

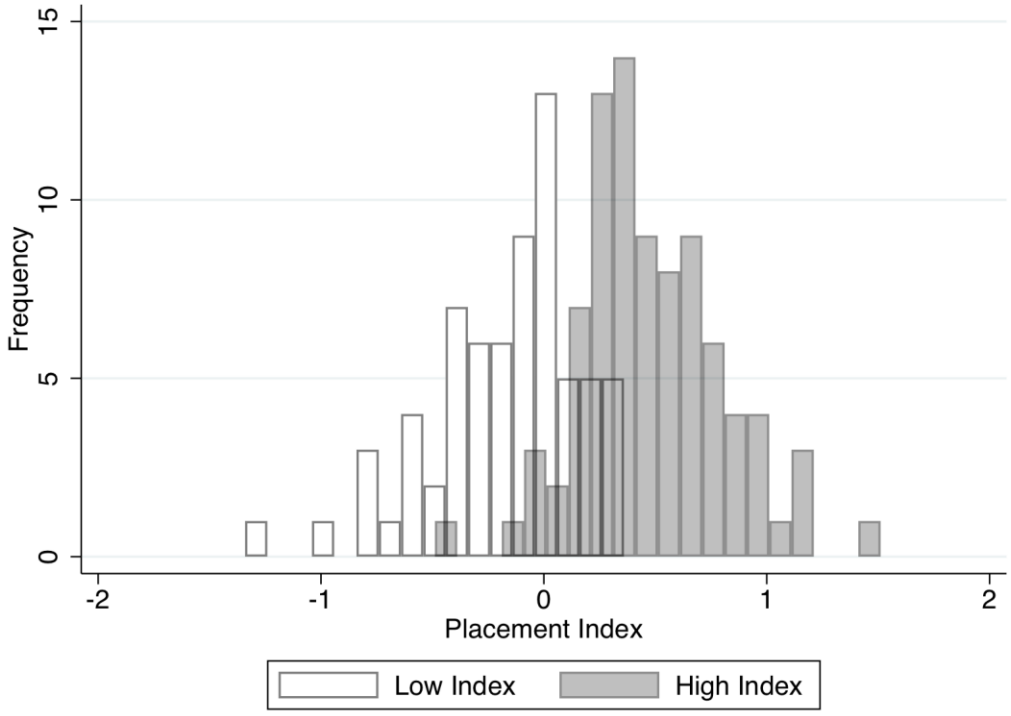
What Works Clearinghouse (2017). Standards handbook version 4.0. Institute of Education Sciences What Works Clearinghouse. Retrieved from: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf

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

Zeichner, K. G., & Gore, Y. (1990). Teacher socialization. *Handbook of Research on Teacher Education. New York: MacMillan*, 329-348.

FIGURES

Figure 1: Distribution of Placement Index Scores by Condition



Note: Placement index is constructed by weighing the teacher index at 75 percent and the school index at 25 percent. The teacher index is calculated by averaging the placement teacher's standardized observation ratings, standardized value-added measures, and standardized years of experience teaching. The school index is calculated by averaging the placement school's standardized teacher stay ratio and the standardized average teacher TVAAS scores.

TABLES

Table 1: Sample Characteristics

	Overall	High Index	Low Index	Standardized Contrast
	(1)	(2)	(3)	(4)
Panel A – Participant counts				
Intended placement	183	97	86	
Actual Placement	157	88	69	
Panel B – Cooperating teacher and school characteristics				
CT Observation Rating	4.12	4.31	3.87	0.705***
CT TVAAS	0.07	0.14	-0.02	0.733***
CT Experience	12.75	16.79	7.57	0.887***
School TVAAS	0.02	0.05	-0.03	0.162
School Stay-Ratio (retention)	0.87	0.88	0.86	0.005

Note. Cooperating Teacher Observation Ratings, Cooperating Teacher TVAAS, School TVAAS, and School Stay Ratio are averages over the three years prior to placement (2013 through 2015). Cooperating Teacher experience is the total number of years of teaching reported the year prior to placement (2015). Standardized contrast reports differences in standard deviation units between high index and low index placement. All variables are standardized at the state level.
**** p < 0.001 ** p < 0.01 * p < 0.05 + p < 0.10*

Table 2: Attrition Rates by Condition

	Intended to Treat	Survey No Response - Intended	Survey No Response - Treated
	(1)	(2)	(3)
Differential	-0.0632	0.0678	0.0586
	(0.0600)	(0.0846)	(0.0919)
Baseline	0.176***	0.276***	0.247***
	(0.0464)	(0.0566)	(0.0656)
N	183	183	157

Note. Columns (1) and (2) report the estimates for attrition on the intended placement sample. Column (3) reports the estimates of response rate attrition for the treated placement sample. Estimates are adjusted for placement block fixed effects.
*Robust standard error in parentheses. *** p < 0.001 ** p < 0.01 * p < 0.05 + p < 0.10*

Table 3. Balance Test on Teacher Candidate Administrative Data

	All (1)	Low Index (2)	High Index (3)	Diff	Effect Size
Female	0.895	0.898	0.893	-0.005	0.017
Male	0.105	0.102	0.107	0.005	0.017
White	0.957	0.956	0.957	0.001	0.007
Other	0.043	0.044	0.043	-0.001	0.007
Cumulative GPA	3.617	3.640	3.596	-0.044	0.138
Inclusive GPA	3.379	3.369	3.387	0.018	0.049
	Chi Square	df	p-value		
No Stratification	2.57	4	0.632		
Stratification	3.60	4	0.463		

Note. Administrative data is provided by TTU for all residents who completed their teacher preparation program (N = 176). Omnibus chi square test results are calculated using Hansen and Bowers (2008) xbalance program. As a sensitivity check, we also checked for balance on Praxis scores, number of teacher candidates took the Praxis test and pre-placement survey responses for a subsample of participants for which these data were available. All observed differences between conditions were not significant at the 95% level.

Table 4: Differences in Residency Experiences Between Low and High Index Placements

	Intent to Treat (1)	Local Average Treatment Effect (2)
Perceived cooperating teacher instructional effectiveness	0.287+	0.413*
	(0.173)	(0.190)
Placement school working conditions	0.0825	0.196
	(0.172)	(0.185)
Frequency of feedback	0.353*	0.499**
	(0.179)	(0.191)
Frequency of coaching activities	0.267	0.426*
	(0.179)	(0.193)
Perceived coaching quality	0.239	0.425*
	(0.174)	(0.182)
N	123	110

Note. This table displays treatment effects on field placement characteristics. Each row is the mean differences on a latent factor calculated using a Confirmatory Factor Analysis and that summarizes the student teacher's responses on one section of the field placement post-survey. All estimates are adjusted for placement stratification using an inverse probability weighing.
*Robust standard error in parentheses. *** p < 0.001 ** p < 0.01 * p < 0.05 + p < 0.10*

Table 5: Differences in Opportunities to Practice Teaching Between Low and High Index Placements

	Intent to Treat	Local Average Treatment Effect
	(1)	(2)
Opportunities to practice planning	0.261	0.341+
	(0.170)	(0.188)
Opportunities to establish a learning environment	0.357*	0.391*
	(0.168)	(0.174)
Opportunities to deliver instruction	0.170	0.267
	(0.175)	(0.192)
Opportunities to differentiate instruction	0.276	0.367*
	(0.172)	(0.179)
N	126	113

Note. This table displays treatment effects on field placement characteristics. Each row is the mean differences on a latent factor calculated using a Confirmatory Factor Analysis and that summarizes the student teacher's responses on one section of the field placement post-survey. All estimates are adjusted for placement stratification using an inverse probability weighing.
*Robust standard error in parentheses. *** p < 0.001 ** p < 0.01 * p < 0.05 + p < 0.10*

Table 6: Differences in Feelings of Preparedness Between Teacher candidates in Low and High Index Placements

	Intent to Treat	Local Average Treatment Effect
	(1)	(2)
General preparedness	0.310+	0.363+
	(0.179)	(0.198)
Preparedness to plan	0.136	0.158
	(0.182)	(0.203)
Preparedness to establish a learning environment	0.154	0.158
	(0.171)	(0.188)
Preparedness to deliver instruction	0.0873	0.0889
	(0.174)	(0.194)
N	122	109

Note. This table displays treatment effects on field placement characteristics. Each row is the mean differences on a latent factor calculated using a Confirmatory Factor Analysis and that summarizes the student teacher's responses on one section of the field placement post-survey. All estimates are adjusted for placement stratification using an inverse probability weighing.
*Robust standard error in parentheses. *** p < 0.001 ** p < 0.01 * p < 0.05 + p < 0.10*

APPENDIX TABLES

Table A1: Distribution of Initial Field Placement Grade Levels and Subject Areas

	N	Percent
Grades Pre-K to 3	19	9.9
Grades K to 5	115	59.9
Grades K to 6	2	1.04
Grades 4 to 8	11	5.73
Grades 7 to 12 - Biology	6	3.12
Grades 7 to 12 - English	12	6.25
Grades 7 to 12 - History	1	0.52
Grades 7 to 12 - Math	5	2.6
Grades 9 to 12 - Biology	1	0.52
Grades 9 to 12 - FCS	1	0.52
Grades 9 to 12 - Government	1	0.52
Grades 9 to 12 - History	1	0.52
Grades 9 to 12 - Math	3	1.56
Grades K to 12 - English as a Second Language	4	2.08
K to 12 - Physical Education	4	2.08
Special Education	6	3.12
Total	192	100

Note. Counts report teacher candidates' choice for field placement setting.

Table A2: T-Test of Background Characteristics between ITT Teacher Candidates Who Did and Did Not Respond to the Post-placement Survey

	All (1)	No Response (2)	Response (3)	Diff	Effect Size
Female	0.891	0.837	0.919	0.082	0.262+
Male	0.109	0.163	0.081	-0.082	0.262+
White	0.962	0.981	0.952	-0.029	0.150
Other	0.038	0.019	0.048	0.029	0.150
Cumulative GPA	3.619	3.564	3.645	0.082	0.255
Inclusive GPA	3.380	3.376	3.378	0.002	0.007
	Chi Square	df	p-value		
No Stratification	9.24	4	0.0554		
Stratification	8.39	4	0.0784		

Note. Administrative data is provided by TTU for all teacher candidates in the ITT sample (N = 183). Omnibus chi square test results are calculated using Hansen and Bowers (2008) xbalance program. As a sensitivity check, we also checked for balance on Praxis scores, number of teacher candidates took the Praxis test and pre-placement survey responses for a subsample of participants for which these data were available.

Table A3: Model Fit Indices for Confirmatory Factor Analyses

Factor	χ^2	d.f.	p-value	RMSEA	CFI	TLI	SRMR
Perceived cooperating teacher instructional effectiveness	326.97	28	0.232	0.041	0.969	0.952	0.043
Frequency of feedback	427.70	15	0.811	0.000	1.000	1.001	0.017
Frequency of coaching	848.27	45	0.085	0.050	0.977	0.969	0.047
Perceived coaching quality	511.06	28	0.319	0.031	0.982	0.968	0.028
School working conditions	381.69	28	0.360	0.025	0.980	0.962	0.042
Preparedness - General	363.58	28	0.947	0.000	1.000	1.010	0.025
Preparedness - Specific	1095.8	120	0.102	0.037	0.947	0.931	0.052
Practice - Specific	794.32	120	0.257	0.026	0.947	0.929	0.056

Note. RMSEA is the Root Mean Square Error of Approximation, CFI is the Comparative Fit Index, TLI is the Tucker-Lewis Index, SRMS is the Standardized Root Mean Residual. Good model fit requires a non-significant chi square value (i.e., $p < 0.05$), RMSEA < 0.06, CFI and TLI > 0.95, and SRMR < 0.05 (Hu & Bentler, 1999). Chi Square and RMSEA are the Satorra-Bentler adjusted values.

Table A4: Descriptive Statistics for Survey Items and Factor Loadings

	Mean	SD	Min	Max	Factor Loading
Factor: Perceived Cooperating Teacher Instructional Effectiveness					
Handle a range of classroom management or discipline situations	3.55	0.73	1	5	0.76
Use a variety of instructional methods	3.45	0.83	1	4	0.79
Teach your subject matter	3.71	0.59	2	5	0.75
Use computers in classroom instruction	3.15	1.01	1	5	0.41
Assess students	3.46	0.77	1	5	0.76
Differentiate instruction in the classroom	3.12	1.00	1	5	0.65
Use data from student assessments to inform instruction	3.32	0.88	1	5	0.70
Teach to state content standards	3.86	0.50	2	5	0.63
Factor: Placement school working conditions					
I regularly saw/communicated with administrators in the building	3.29	0.80	1	4	0.68
Teachers were willing to collaborate with me about instruction	3.37	0.82	1	4	0.76
School leadership supported me when I needed it	3.35	0.84	1	4	0.95
I felt welcome to attend faculty and departmental meetings at my student teaching placement	3.65	0.60	1	4	0.72
I felt included in the daily activities of the school	3.59	0.67	1	4	0.73
The school administration's behavior toward the staff was supportive and encouraging	3.60	0.69	1	4	0.78
The level of student misbehavior in this school did not interfere with teaching	3.32	0.73	1	4	0.55
Necessary materials, such as textbooks, supplies, and copy machines were available as needed by the staff	3.61	0.62	1	4	0.52
Factor: Frequency of Feedback					
Offer concrete suggestions	2.99	0.86	1	4	0.80
Ask reflective questions	2.67	1.02	1	4	0.88
Offer general observations	3.06	0.81	1	4	0.79
Refer to specific things you need to improve	2.65	0.93	1	4	0.78
Refer to specific things you did well	3.08	0.93	1	4	0.75
Share specific data when providing feedback	2.66	1.08	1	4	0.86
Factor: Frequency of coaching activities					
Observe you teach	6.16	1.25	2	7	0.58

Table A4 Continued

Ask you to observe an aspect of his/her teaching	4.22	2.33	1	7	0.62
Co-design lessons or parts of lessons with you	4.96	1.97	1	7	0.58
Co-teach lessons or parts of lessons with you	4.85	2.10	1	7	0.65
Analyze student work with you	5.26	2.08	1	7	0.66
Encourage you to practice specific aspects of your teaching	4.88	2.17	1	7	0.86
Share data or evidence about lessons s/he observed you teach	4.50	2.34	1	7	0.90
Offer you feedback on your teaching	5.16	1.93	1	7	0.87
Provide a formal evaluation of your teaching	3.11	2.05	1	7	0.70
Use evaluation data to provide recommendations for improvement	3.43	2.27	1	7	0.81
Factor: Perceived coaching quality					
My mentor observed me teach frequently enough	3.53	0.78	1	4	0.70
My mentor provided me with feedback frequently enough	3.17	0.97	1	4	0.82
The feedback that my mentor offered helped me learn to teach	3.39	0.86	1	4	0.88
The feedback that my mentor offered was consistent with the feedback I got from my university supervisor	3.30	0.89	1	4	0.84
When I struggled with my teaching, I felt comfortable going to my mentor for help	3.52	0.78	1	4	0.82
My mentor's expectations of me were appropriate to my experience	3.53	0.80	1	4	0.82
My mentor allowed me to make my own instructional decisions	3.64	0.66	1	4	0.76
I felt comfortable taking instructional risks in front of my mentor	3.50	0.82	1	4	0.81
Panel F: General preparedness					
Handle a range of classroom management or discipline situations	4.06	0.84	2	5	0.70
Use a variety of instructional methods	4.20	0.76	1	5	0.83
Teach your subject matter	4.31	0.66	2	5	0.68
Use computers in classroom instruction	4.17	0.86	1	5	0.63
Assess students	4.16	0.73	2	5	0.66
Differentiate instruction in the classroom	4.09	0.86	2	5	0.78
Use data from student assessments to inform instruction	4.09	0.75	1	5	0.57
Teach to state content standards	4.52	0.64	1	5	0.66
Factor: Preparedness to plan					
Design instructional plans	4.32	0.72	2	5	0.86

Table A4 Continued

Create student work assignments	4.33	0.68	2	5	0.94
Design assessment plans	4.20	0.72	2	5	0.80
Plan instructional activities and materials	4.33	0.67	2	5	0.89
Factor: Preparedness to set up learning environment					
Set academic expectations for all students	4.36	0.68	2	5	0.81
Manage student behavior	4.14	0.85	2	5	0.72
Establish a respectful culture in the classroom	4.53	0.64	2	5	0.80
Set learning objectives and standards	4.51	0.56	3	5	0.78
Motivate students	4.52	0.62	2	5	0.77
Establish a classroom environment	4.44	0.68	2	5	Dropped
Factor: Preparedness to deliver instruction					
Present instructional content	4.51	0.58	3	5	0.73
Structure and pace lessons	4.12	0.75	2	5	0.72
Provide academic feedback	4.25	0.71	3	5	0.71
Implement student grouping strategies in the classroom	4.36	0.68	2	5	0.70
Demonstrate content knowledge	4.43	0.60	3	5	0.84
Understand students' instructional needs	4.31	0.68	2	5	0.84
Provide opportunities for students to think about material on multiple levels	4.09	0.79	1	5	0.65
Question students about instructional content	4.29	0.65	3	5	Dropped
Implement activities that teach and reinforce multiple problem- solving strategies	4.07	0.84	1	5	Dropped
Professionalism					
Engage in professional growth and learning	4.33	0.64	2	5	-
Use data to support instruction	3.07	1.78	1	5	-
Engage positively with the school and community	2.62	1.92	1	5	-
Contribute to the school community by actively assisting, collaborating with, and/or mentoring other colleagues/staff	2.56	1.90	1	5	-
Factor: Practice planning					
Design instructional plans	5.61	1.44	2	7	0.84
Create student work assignments	5.62	1.47	2	7	0.96
Design assessment plans	4.99	1.70	1	7	0.81

Table A4 Continued

Plan instructional activities and materials	5.85	1.41	1	7	0.60
Factor: Practice setting up learning environment					
Establish a classroom environment	6.54	1.18	1	7	0.68
Set learning objectives and standards	6.80	0.72	1	7	0.90
Motivate students	6.92	0.36	4	7	0.78
Present instructional content	6.60	0.80	2	7	-0.39
Implement student grouping strategies in the classroom	5.81	1.53	1	7	-0.46
Set learning objectives and standards	5.82	1.56	1	7	Dropped
Factor: Practice instruction					
Plan instructional activities and materials	5.85	1.41	1	7	0.37
Present instructional content	6.60	0.80	2	7	0.83
Implement student grouping strategies in the classroom	5.81	1.53	1	7	0.93
Set academic expectations for all students	6.13	1.45	1	7	0.75
Manage student behavior	6.87	0.46	4	7	0.21
Structure and pace lessons	6.23	1.16	2	7	0.87
Provide academic feedback	6.58	0.85	2	7	0.54
Question students about instructional content	6.65	0.68	3	7	Dropped
Factor: Practice delivering tasks					
Demonstrate content knowledge	6.77	0.66	3	7	0.47
Understand students' instructional needs	6.75	0.74	1	7	0.87
Provide opportunities for students to think about material on multiple levels	6.37	0.90	3	7	0.91
Implement activities that teach and reinforce multiple problem-solving strategies	5.98	1.35	1	7	Dropped
Practice Professionalism					
Engage in professional growth and learning	6.18	1.28	2	7	-
Use data to support instruction	5.25	1.66	1	7	-
Engage positively with the school and community	6.43	1.32	1	7	-
Contribute to the school community by actively assisting, collaborating with, and/or mentoring other colleagues/staff	5.92	1.69	1	7	-

Note. Factor loadings are the standardized solution to confirmatory factor analyses. All loadings are significant at the customary level.

Table A5: Treatment Effects on Perceived Cooperating Teacher Instructional Effectiveness Factor

How effectively did your mentor teacher… ? [Not at all effectively; somewhat effectively; effectively; very effectively.]	Intent to Treat (IPW Adjusted) (1)	Local Average Treatment Effect (IPW Adjusted) (2)	Intent to Treat (OLS) (3)	Intent to Treat (Block Fixed Effects) (4)
Handle a range of classroom management or discipline situations	0.303+	0.430*	0.314+	0.406*
	(0.174)	(0.192)	(0.171)	(0.206)
Use a variety of instructional methods	0.125	0.199	0.148	0.276+
	(0.172)	(0.191)	(0.173)	(0.154)
Teach your subject matter	0.274	0.360+	0.280	0.349+
	(0.173)	(0.193)	(0.172)	(0.190)
Use computers in classroom instruction	0.0634	0.136	0.0633	0.139
	(0.173)	(0.185)	(0.174)	(0.188)
Assess students	0.458**	0.582**	0.457**	0.380*
	(0.164)	(0.182)	(0.162)	(0.183)
Differentiate instruction in the classroom	0.114	0.208	0.131	0.192
	(0.182)	(0.195)	(0.182)	(0.186)
Use data from student assessments to inform instruction	0.101	0.131	0.106	0.0476
	(0.182)	(0.196)	(0.181)	(0.201)
Teach to state content standards	0.189	0.289	0.177	0.224
	(0.167)	(0.184)	(0.166)	(0.180)
Factor: Perceived Mentor Instructional Effectiveness	0.287+	0.413*	0.296+	0.304+
	(0.173)	(0.190)	(0.172)	(0.179)
N	115	107	115	115

Note. The first row reports the treatment effect on the main factor and the other rows report effects on each item raw score. These survey items included the response option "Not enough information to evaluate". We mark these responses as missing and adjust the response counts accordingly.
**** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$ + $p < 0.10$*

Table A6: Treatment Effects on Working Conditions in Placement School

To what extent do you agree with the following statements about your school placement? [Strongly disagree; disagree; agree; strongly agree]	Intent to Treat (IPW Adjusted) (1)	Local Average Treatment Effect (IPW Adjusted) (2)	Intent to Treat (OLS) (3)	Intent to Treat (Block Fixed Effects) (4)
I regularly saw/communicated with administrators in the building	0.0930	0.165	0.0942	-0.0252
	(0.171)	(0.188)	(0.171)	(0.199)
Teachers were willing to collaborate with me about instruction	0.329*	0.446*	0.330*	0.301
	(0.166)	(0.183)	(0.165)	(0.194)
School leadership supported me when I needed it	0.101	0.232	0.109	0.0322
	(0.172)	(0.183)	(0.171)	(0.195)
I felt welcome to attend faculty and departmental meetings at my student teaching placement	-0.181	-0.110	-0.182	-0.160
	(0.165)	(0.179)	(0.165)	(0.169)
I felt included in the daily activities of the school	-0.138	-0.0459	-0.137	-0.139
	(0.165)	(0.170)	(0.166)	(0.188)
The school administration's behavior toward the staff was supportive and encouraging	-0.103	-0.0838	-0.109	-0.138
	(0.185)	(0.203)	(0.183)	(0.197)
The level of student misbehavior in this school did not interfere with teaching	0.108	0.193	0.104	0.0753
	(0.184)	(0.201)	(0.184)	(0.197)
Necessary materials, such as textbooks, supplies, and copy machines were available as needed by the staff	0.0582	0.105	0.0627	-0.0985
	(0.181)	(0.199)	(0.182)	(0.199)
Factor: Placement school working conditions	0.0825	0.196	0.0849	0.0242
	(0.172)	(0.185)	(0.171)	(0.191)
N	123	114	123	123

Note. The first row reports the treatment effect on the main factor and the other rows report effects on each item raw score.
*Robust standard error in parentheses. *** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$ + $p < 0.10$*

Table A7: Treatment Effects on Cooperating Teacher Frequency of Feedback Factor

How often did your mentor teacher … ? [Never; once in a while; often; all the time]	Intent to Treat (IPW Adjusted) (1)	Local Average Treatment Effect (IPW Adjusted) (2)	Intent to Treat (OLS) (3)	Intent to Treat (Block Fixed Effects) (4)
Offer concrete suggestions	0.275	0.388*	0.284	0.257
	(0.180)	(0.195)	(0.180)	(0.189)
Ask reflective questions	0.393*	0.547**	0.388*	0.452*
	(0.177)	(0.187)	(0.176)	(0.182)
Offer general observations	0.220	0.323+	0.238	0.156
	(0.180)	(0.192)	(0.180)	(0.219)
Refer to specific things you need to improve	0.216	0.299	0.250	0.252
	(0.183)	(0.200)	(0.182)	(0.183)
Refer to specific things you did well	0.215	0.331+	0.225	0.220
	(0.180)	(0.198)	(0.180)	(0.203)
Share specific data when providing feedback	0.362*	0.505**	0.366*	0.351+
	(0.178)	(0.191)	(0.177)	(0.182)
Factor: Frequency of feedback	0.353*	0.499**	0.364*	0.361*
	(0.179)	(0.191)	(0.179)	(0.184)
N	123	114	123	123

Note. The first row reports the treatment effect on the main factor and the other rows report effects on each item raw score.
**** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$ + $p < 0.10$*

Table A8: Treatment Effects on Cooperating Teacher Frequency of Coaching Activities Factor

On average, how often did your mentor teacher do the following…? [Never; less than once a month; once a month; 2-3 times a month; once a week; 2-4 times a week; daily]	Intent to Treat (IPW Adjusted) (1)	Local Average Treatment Effect (IPW Adjusted) (2)	Intent to Treat (OLS) (3)	Intent to Treat (Block Fixed Effects) (4)
Observe you teach	0.618***	0.705***	0.608***	0.608**
	(0.177)	(0.199)	(0.175)	(0.227)
Ask you to observe an aspect of his/her teaching	0.0552	0.133	0.0466	0.132
	(0.180)	(0.198)	(0.180)	(0.218)
Co-design lessons or parts of lessons with you	0.193	0.230	0.181	0.207
	(0.177)	(0.191)	(0.176)	(0.179)
Co-teach lessons or parts of lessons with you	0.147	0.294	0.147	-0.0305
	(0.175)	(0.189)	(0.175)	(0.202)
Analyze student work with you	0.360*	0.452*	0.379*	0.325
	(0.171)	(0.192)	(0.170)	(0.203)
Encourage you to practice specific aspects of your teaching	0.249	0.360+	0.278	0.177
	(0.175)	(0.189)	(0.174)	(0.198)
Share data or evidence about lessons s/he observed you teach	0.223	0.390*	0.232	0.167
	(0.181)	(0.194)	(0.181)	(0.201)
Offer you feedback on your teaching	0.275	0.431*	0.276	0.178
	(0.178)	(0.192)	(0.178)	(0.215)
Provide a formal evaluation of your teaching	0.0178	0.137	0.0227	0.105
	(0.179)	(0.196)	(0.180)	(0.212)
Use evaluation data to provide recommendations for improvement	0.00148	0.137	0.00361	0.0466
	(0.179)	(0.194)	(0.179)	(0.196)
Factor: Frequency of Coaching Activities	0.267	0.426*	0.277	0.210
	(0.179)	(0.193)	(0.178)	(0.204)
N	123	114	123	123

Note. The first row reports the treatment effect on the main factor and the other rows report effects on each item raw score.
**** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$ + $p < 0.10$*

Table A9: Treatment Effects on Perceived Coaching Quality

To what extent do you agree with the following statements about your mentor? [Strongly disagree; disagree; agree; strongly agree]	Intent to Treat (IPW Adjusted) (1)	Local Average Treatment Effect (IPW Adjusted) (2)	Intent to Treat (OLS) (3)	Intent to Treat (Block Fixed Effects) (4)
My mentor observed me teach frequently enough	0.491**	0.583**	0.466**	0.487*
	(0.176)	(0.195)	(0.173)	(0.214)
My mentor provided me with feedback frequently enough	0.283	0.440*	0.282	0.327+
	(0.175)	(0.188)	(0.174)	(0.193)
The feedback that my mentor offered helped me learn to teach	0.141	0.309	0.144	0.220
	(0.179)	(0.190)	(0.179)	(0.199)
The feedback that my mentor offered was consistent with the feedback I got from my university supervisor	0.0519	0.223	0.0487	0.140
	(0.172)	(0.178)	(0.174)	(0.203)
When I struggled with my teaching, I felt comfortable going to my mentor for help	0.195	0.383*	0.196	0.225
	(0.166)	(0.173)	(0.167)	(0.206)
My mentor's expectations of me were appropriate to my experience	0.260	0.410*	0.252	0.278
	(0.176)	(0.185)	(0.175)	(0.223)
My mentor allowed me to make my own instructional decisions	0.174	0.309	0.157	0.165
	(0.173)	(0.189)	(0.172)	(0.194)
I felt comfortable taking instructional risks in front of my mentor	0.125	0.310+	0.108	0.174
	(0.168)	(0.169)	(0.170)	(0.205)
Factor: Perceived coaching quality	0.239	0.425*	0.230	0.287
	(0.174)	(0.182)	(0.174)	(0.200)
N	123	114	123	123

Note. The first row reports the treatment effect on the main factor and the other rows report effects on each item raw score.
*Robust standard error in parentheses. *** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$ + $p < 0.10$*

Table A10: Treatment Effects on Opportunities to Practice

How much of an opportunity have you had to practice… ? [Never; less than once a month; once a month; 2-3 times a month; once a week; 2-4 times a week; daily]	Intent to Treat (IPW Adjusted) (1)	Local Average Treatment Effect (IPW Adjusted) (2)	Intent to Treat (OLS) (3)	Intent to Treat (Block Fixed Effects) (4)
Panel A. Practice Planning				
Design instructional plans	0.289+	0.370+	0.283	0.273
	(0.175)	(0.193)	(0.175)	(0.191)
Create student work assignments	0.265	0.340+	0.245	0.339+
	(0.169)	(0.187)	(0.169)	(0.201)
Design assessment plans	0.186	0.246	0.165	0.231
	(0.168)	(0.186)	(0.168)	(0.192)
Plan instructional activities and materials	0.141	0.213	0.115	0.226
	(0.170)	(0.188)	(0.169)	(0.185)
Factor: Opportunities to practice planning	0.261	0.341+	0.242	0.322+
	(0.170)	(0.188)	(0.170)	(0.196)
Panel B. Practice Establishing a Learning Environment				
Establish a classroom environment	0.357*	0.391*	0.360*	0.354+
	(0.168)	(0.174)	(0.167)	(0.195)
Establish a respectful culture in the classroom	0.0314	0.0614	0.0294	0.00977
	(0.135)	(0.139)	(0.136)	(0.180)
Motivate students	0.0589	0.202*	0.0492	0.0211
	(0.124)	(0.0957)	(0.131)	(0.165)
Implement student grouping strategies in the classroom	0.0160	0.0463	0.0243	0.217
	(0.180)	(0.192)	(0.178)	(0.198)
Factor: Opportunities to establish a learning environment	0.120	0.213+	0.115	0.0956
	(0.128)	(0.127)	(0.131)	(0.159)

Table 10A Continued

Panel C. Practice Delivering Instruction				
Set academic expectations for all students	-0.0722	-0.0246	-0.0700	0.0969
	(0.158)	(0.173)	(0.159)	(0.198)
Manage student behavior	0.146	0.208	0.147	0.140
	(0.186)	(0.206)	(0.184)	(0.204)
Present instructional content	0.122	0.128	0.134	0.131
	(0.181)	(0.202)	(0.180)	(0.231)
Structure and pace lessons	0.273	0.369+	0.267	0.299
	(0.178)	(0.198)	(0.175)	(0.216)
Plan instructional activities and materials	0.141	0.213	0.115	0.226
	(0.170)	(0.188)	(0.169)	(0.185)
Provide academic feedback	-0.0824	-0.00668	-0.0774	-0.143
	(0.184)	(0.199)	(0.184)	(0.209)
Implement student grouping strategies in the classroom	0.0160	0.0724	0.0243	0.217
	(0.180)	(0.200)	(0.178)	(0.198)
Factor: Opportunities to deliver instruction	0.170	0.267	0.168	0.261
	(0.175)	(0.192)	(0.174)	(0.208)
Panel D. Opportunities to Differentiate Instruction				
Understand students' instructional needs	-0.0447	0.367*	-0.0314	-0.267
	(0.186)	(0.179)	(0.187)	(0.177)
Provide opportunities for students to think about material on multiple levels	0.160	-0.122	0.162	0.151
	(0.173)	(0.149)	(0.173)	(0.211)
Implement activities that teach and reinforce multiple problem-solving strategies	0.311+	0.267	0.317+	0.263
	(0.172)	(0.184)	(0.173)	(0.209)
Factor: Opportunities to differentiate instruction	0.276	0.367*	0.279	0.248
	(0.172)	(0.179)	(0.172)	(0.206)
N	126	116	126	126

Note. The first row reports the treatment effect on the main factor and the other rows report effects on each item raw score.
*Robust standard error in parentheses. *** p < 0.001 ** p < 0.01 * p < 0.05 + p < 0.10*

Table A11: Treatment effects on Feelings of Preparedness - General factor

How prepared do you feel to … [not at all; slightly; moderately; very; exceptionally]	Intent to Treat (IPW Adjusted) (1)	Local Average Treatment Effect (IPW Adjusted) (2)	Intent to Treat (OLS) (3)	Intent to Treat (Block Fixed Effects) (4)
Handle a range of classroom management or discipline situations	0.326+	0.406*	0.330+	0.248
	(0.178)	(0.195)	(0.176)	(0.191)
Use a variety of instructional methods	0.279	0.370+	0.249	0.226
	(0.180)	(0.193)	(0.178)	(0.196)
Teach your subject matter	0.262	0.231	0.232	0.163
	(0.178)	(0.200)	(0.178)	(0.181)
Use computers in classroom instruction	0.286	0.318	0.279	0.319+
	(0.174)	(0.195)	(0.173)	(0.179)
Assess students	-0.0353	-0.000337	-0.0296	-0.0480
	(0.179)	(0.199)	(0.179)	(0.215)
Differentiate instruction in the classroom	0.179	0.214	0.185	0.131
	(0.179)	(0.195)	(0.179)	(0.206)
Use data from student assessments to inform instruction	0.168	0.232	0.148	0.161
	(0.183)	(0.203)	(0.183)	(0.211)
Teach to state content standards	0.274	0.261	0.254	0.121
	(0.186)	(0.207)	(0.181)	(0.212)
Factor: General preparedness	0.310+	0.363+	0.293+	0.231
	(0.179)	(0.198)	(0.177)	(0.191)
N	122	113	122	122

Note. The first row reports the treatment effect on the main factor and the other rows report effects on each item raw score.
*Robust standard error in parentheses. *** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$ + $p < 0.10$*

Table A12: Treatment effects on Feelings of Preparedness - Sub-factors

How prepared do you feel to … [not at all; slightly; moderately; very; exceptionally]	Intent to Treat (IPW Adjusted) (1)	Local Average Treatment Effect (IPW Adjusted) (2)	Intent to Treat (OLS) (3)	Intent to Treat (Block Fixed Effects) (4)
Panel A. Planning Instruction				
Design instructional plans	0.287	0.289	0.278	0.282
	(0.179)	(0.200)	(0.178)	(0.207)
Create student work assignments	0.0676	0.0596	0.0556	-0.0689
	(0.181)	(0.201)	(0.179)	(0.204)
Design assessment plans	-0.0436	-0.0609	-0.0334	-0.0894
	(0.176)	(0.197)	(0.177)	(0.211)
Plan instructional activities and materials	0.202	0.268	0.201	0.209
	(0.183)	(0.202)	(0.183)	(0.212)
Factor: Preparedness to plan	0.136	0.158	0.128	0.0551
	(0.182)	(0.203)	(0.181)	(0.206)
Panel B. Design instructional environment				
Set academic expectations	0.0186	0.0296	-0.0189	-0.216
	(0.176)	(0.193)	(0.173)	(0.207)
Manage student behavior	0.128	0.125	0.109	-0.0456
	(0.181)	(0.196)	(0.180)	(0.200)
Establish a respectful culture in the classroom	0.213	0.190	0.176	-0.119
	(0.180)	(0.200)	(0.178)	(0.181)
Set learning objectives and standards	0.0347	0.0405	0.0270	-0.239
	(0.174)	(0.189)	(0.174)	(0.189)
Motivate students	0.237	0.302+	0.209	0.0586
	(0.169)	(0.181)	(0.168)	(0.199)
Factor: Preparedness to establish a learning environment	0.154	0.158	0.124	-0.0741
	(0.171)	(0.188)	(0.171)	(0.186)

Table A12 Continued

Panel C. Instruction				
Present instructional content	0.153	0.0771	0.138	0.138
	(0.181)	(0.192)	(0.180)	(0.196)
Structure and pace lessons	0.122	0.185	0.125	0.123
	(0.176)	(0.196)	(0.176)	(0.196)
Provide academic feedback	0.00607	0.0151	0.00915	0.0422
	(0.181)	(0.201)	(0.181)	(0.213)
Group students	0.0190	-0.0161	0.0292	0.0551
	(0.176)	(0.196)	(0.177)	(0.197)
Demonstrate content knowledge	0.0620	0.0567	0.0561	0.122
	(0.174)	(0.191)	(0.174)	(0.195)
Understand students' instructional needs	-0.0534	-0.0394	-0.0647	-0.161
	(0.175)	(0.194)	(0.174)	(0.201)
Provide opportunities for students to think about material on multiple levels	0.184	0.243	0.181	0.124
	(0.179)	(0.198)	(0.180)	(0.221)
Factor: Preparedness to deliver instruction	0.0873	0.0889	0.0746	0.0293
	(0.174)	(0.194)	(0.174)	(0.190)
N	122	113	122	122

Note. The first row reports the treatment effect on the main factor and the other rows report effects on each item raw score.
*Robust standard error in parentheses. *** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$ + $p < 0.10$*