Three Different Measures of Graduates’ Instructional Readiness and the Features of Preservice Preparation that Predict Them

Matthew Ronfeldt1, Kavita Kapadia Matsko2, Hillary Greene Nolan1, and Michelle Reininger3

Abstract
This article extends prior research seeking to identify preparation features related to better workforce outcomes. To our knowledge, it is the first to link many dimensions of preparation to graduates’ first-year observation ratings. It follows 305 preservice teachers (PSTs) who student taught in Chicago Public Schools (CPS) in 2014–2015 and were subsequently hired in CPS in 2015–2016. PSTs received stronger observation ratings when their CTs had stronger observation ratings themselves, their CTs reported providing stronger coaching in specific areas, they gained employment in their field placement schools, and they student taught in self-contained elementary classrooms. Finally, we tested whether these same preparation features were associated with two other outcomes—(a) how well prepared PSTs felt after student-teaching and (b) how well prepared their CTs felt their PSTs were—and found they were not. We discuss implications for using workforce and survey-based outcomes to identify promising forms of preparation.

Keywords
student teaching, preservice teacher education, field experiences

Introduction
In recent decades, teacher preparation programs have faced pressures to demonstrate their efficacy, particularly via graduate workforce outcomes. Early efforts tested whether graduates from different pathways or programs had, on average, better value-added to student achievement measures (VAMs) and yielded mixed results. More recently, scholars have shifted focus from comparing pathways and programs to exploring which features of preparation within and across pathways and programs are associated with better instructional quality. Whereas early studies analyzed features of preparation in connection to graduates’ perceived preparedness, recent studies have focused on observed effectiveness (e.g., VAMs). One drawback to this approach is that VAMs are unavailable for many graduates and are indirect measures that make inferences about instructional quality based on student test scores.

We extend prior work by linking features of preparation to graduates’ first-year observational ratings based on their district evaluation rubric, a more widely available and direct measure of instructional performance than VAMs. Moreover, we focus on cooperating teachers (CTs) who have often been overlooked in prior research.

We find that preservice teachers (PSTs) are evaluated as more effective on their first-year observational ratings when (a) their CTs were rated as more effective on the same rubric, (b) their CTs reported providing stronger coaching around the rubric’s instructional domains, and (c) when they were hired into the schools in which they student-taught. However, we find that these same features of preparation were unrelated to (a) how well prepared PSTs felt after student-teaching and (b) how well prepared their CTs felt their PSTs were—and found they were not. We discuss implications for using workforce and survey-based outcomes to identify promising forms of preparation.

1University of Michigan, Ann Arbor, USA
2National Louis University, IL, USA
3Stanford University, CA, USA

Corresponding Author:
Matthew Ronfeldt, University of Michigan, 610 East University Ave., Ann Arbor, MI 48109, USA.
Email: ronfeldt@umich.edu
Literature Review

Identifying features of preparation likely to promote early career teaching quality is the primary goal of our study. Our review focuses on three measures of instructional quality linked to preparation features in prior research: survey-based measures of self-perceived instructional preparedness, VAMs, and observational ratings. Unless otherwise noted, reviewed studies include PSTs from across certification levels and subjects.

Self-Perceived Preparedness

Traditionally, scholars used large-scale surveys of recent graduates to assess preparation. Darling-Hammond et al.'s (2002) survey of 3,000 teachers found that graduates from traditional programs felt better prepared than graduates from alternative programs or teachers without formal training. Although acknowledging that linking preparation to direct measures of effectiveness would be preferable, the authors defended their emphasis on self-perceived preparedness because it was associated with teacher efficacy which, in prior work, predicted student achievement. Subsequently, scholars have found survey-based measures of self-perceived preparedness to be positively related to teachers’ career plans, self-efficacy, and retention (Ronfeldt et al., 2014; Ronfeldt & Reininger, 2012); none, to our knowledge, has linked self-perceived preparedness to more objective measures of instructional quality like VAMs or observational evaluations, a contribution of this study.

Researchers have also investigated which features of preparation predict graduates’ feelings of preparedness and self-efficacy. In Chicago, Ronfeldt and Reininger (2012) examined whether having more or better quality student-teaching experiences predicted feeling better prepared, having stronger self-efficacy, or planning longer teaching careers. While graduates’ reports of better quality clinical experiences positively predicted all outcomes, duration did not. Ronfeldt et al. (2013) found that PSTs who reported better quality CTs, field supervision, and more autonomy over instructional decisions felt better prepared and reported stronger teacher self-efficacy. In addition, Matsko et al. (2018) found PSTs felt better prepared when paired with more instructionally effective CTs whose coaching they felt was stronger and more frequent. Finally, using nationally representative data, Ronfeldt et al. (2014) found that teachers who completed more methods-related courses and weeks of student teaching had higher perceived preparedness and retention.

VAMs

VAM studies have examined average differences between pathways and programs with mixed results—some suggesting few, if any, significant or meaningful differences (Constantine et al., 2009; Goldhaber et al., 2013; Koedel et al., 2015; von Hippel et al., 2016), and others suggesting significant and meaningful differences (Boyd et al., 2009; Darling-Hammond et al., 2005; Glazerman et al., 2006; Henry et al., 2014).

Other studies have considered preparation features that predict graduates’ VAMs. Boyd et al. (2009) found that graduates from elementary programs with more oversight of clinical experiences and with more practice-based teaching opportunities had better VAMs. Ronfeldt (2012, 2015) found that graduates had better VAMs when they had learned to teach in schools with better average teacher retention, collaboration, and school-level achievement gains. Together, these results suggest that schools with supportive environments for teaching and learning to teach make for promising placements. Extending this line of inquiry, we include measures for the “Five Essentials” (Effective Leader, Ambitious Instruction, Involved Families, Collaborative Teachers, Supportive Environment)—known to predict positive school improvement (Bryk et al., 2010) but not yet examined in the student-teaching context.

The studies above also tested whether sociodemographic characteristics of students in field placement schools (FPS) were related to graduate VAMs, generally finding no relationships. In one exception, Ronfeldt (2015) found that candidates prepared in placement schools with higher proportions of Black students also had better VAMs in some specifications. In addition, Goldhaber et al. (2017) found that graduates employed in schools similar to their placement schools had better VAMs.

A growing body of research finds that CTs’ effectiveness—an often overlooked but essential aspect of clinical placements—is related to their mentees’ effectiveness. Ronfeldt et al. (2018) found that candidates in Tennessee with more instructionally effective CTs were themselves more effective as first-year teachers. Goldhaber et al. (2018) found similar positive correlations between VAMs of CTs and candidates in Washington State.

Studies based upon graduates’ VAMs have pointed to relationships between preparation and graduates’ instructional effectiveness. Whereas VAMs exist for a minority of graduates and are indirect measures that infer teachers’ instructional quality through student performance, observational evaluations exist for most or all teachers, and are more direct measures of instructional quality. We discuss observational ratings next.

Observational Ratings

Despite widespread availability, we have found only three studies linking preparation to graduates’ observational ratings. Ronfeldt and Campbell (2016) found significant and meaningful differences between Tennessee providers in terms of graduates’ average observation ratings. Highest quartile providers graduated teachers who performed as though they had an additional year of teaching experience as compared with lowest quartile providers. Similarly, researchers in North Carolina found that some programs differed significantly from the state mean in terms of graduates’ average observational ratings.
Ronfeldt et al. (Bastian et al., 2018). In Tennessee, most teachers are rated using the Tennessee Educator Acceleration Model (TEAM) rubric which assesses four domains—Instruction, Environment, Planning, and Professionalism. In North Carolina, teachers are evaluated based upon the North Carolina Educator Evaluation System (NCEES), a rubric based upon five standards: Demonstrate Leadership; Establish a Respectful Classroom Environment; Know the Content They Teach; Facilitate Learning for Their Students; and Reflect on Their Practice.

To our knowledge, one prior study has linked features of preparation to observational ratings. Ronfeldt et al. (2018) found that graduates’ observational ratings were positively related to their CTs’ observational ratings but not their VAM scores; likewise graduates’ VAMs were associated with their CTs’ VAMs but not their observation ratings. The authors speculated that these divergent results could suggest that different measures capture different dimensions of instructional quality, consistent with Hill et al. (2011) who suggest that VAMs and observation ratings may capture different dimensions of instructional quality among middle-school mathematics teachers. Shulman (2004) also theorizes teaching quality as a complex construct that requires multiple measures for capturing its complexity, which raises challenges for teacher preparation and related research. A central contribution of this study is to determine whether preparation features predict three measures of instructional readiness similarly.

**Logic Model**

We categorize features of preparation experienced by PSTs into a logic model (Figure 1) consisting of four broad dimensions: (a) CT modeling as captured by instructional quality and professional qualifications (see Table 6 for variable list), (b) CT coaching (see Table 7 for variable list), (c) FPS characteristics (see Table 8 for variable list), and (d) other features of preparation (e.g., program-level structures like timing of coursework versus fieldwork; see Table 9 for variable list). Most features in our conceptual model have been shown to predict graduates’ instructional quality, while others have been identified in the literature as important but have not yet been linked to graduates’ instructional quality. This logic model guided the organization of the PST and CT surveys and the specific questions included on them.

The first two dimensions assess two different mechanisms by which CTs are likely to influence PST quality: (a) modeling effective instruction and (b) providing high-quality coaching (Matsko et al., 2018). Regarding modeling, PSTs learn by observing their CTs and emulating the skills and behaviors, whether efficacious or not, underscoring the importance of recruiting instructionally effective teachers as mentors (Rozelle & Wilson, 2012). As coaches, CTs deliver feedback, deliberately structure learning opportunities, and offer emotional support (Glenn, 2006; Schwille, 2008).

Prior literature also suggests that (c) broader FPS environments can provide more or less supportive contexts for PST learning. We examine measures of school working conditions and school-level student characteristics shown to predict PST outcomes. Finally, we consider additional, structural features of preparation related (d) to student teaching (e.g., how the PST was placed in the school) and (e) to programs more broadly (e.g., number of courses). As Figure 1 shows, we explore the relationships between features of and three outcomes (solid lines): PST-perceived readiness to teach, CT-perceived readiness to teach, and first-year observational ratings. The various outcomes themselves are also likely correlated (dashed lines, Figure 1). Represented by the vertical, dashed line, we anticipated that PSTs would feel better prepared when their CTs

---

**Figure 1. Logic model.**
perceived them as better prepared. We further anticipated that PSTs who felt better prepared and were seen by CTs as more prepared would have better first-year observation ratings (dashed, diagonal lines).

To date, one feature of preparation—the instructional effectiveness of CTs—has been linked to graduates’ observational ratings. A contribution of this article is to identify other preparation features associated with graduates’ observational ratings. In addition, we test whether features associated with observational ratings also predict survey-based measures of instructional performance—PSTs’ and CTs’ perceptions of PSTs’ preparedness to teach. To our knowledge, this latter measure has not previously been used in studies on the effects of preparation. Given self-evaluations of performance are known to be biased and/or inaccurate (Dunning et al., 2003), pairing PSTs’ with CTs’ evaluations seemed worthwhile. Finally, we investigate whether first-year observational ratings are correlated with PSTs’ and CTs’ perceptions of preparedness to evaluate these survey-based measures’ predictive validity.

Thus, we investigate:

1. What features of teacher preparation, and especially clinical preparation, are associated with PSTs’ first-year observational ratings?
2. What features of preparation predict how instructionally effective PSTs (a) feel at the end of preparation and (b) are perceived by their CTs?
3. Are there any features of preparation that predict PSTs’ first-year observational ratings as well as their self-reported and CT-reported preparedness after student teaching?

Method

Setting

This study takes place in Chicago Public Schools (CPS), the third largest school district in the country.2 CPS is also the site of student-teaching placements for nearly 40 university-based teacher preparation programs. PSTs register to student teach in CPS through a centralized registration process that maintains information on PSTs, CTs, and preparation programs.

Data

We administered surveys to PSTs and CTs in 2014–2015. All registered PSTs received surveys via email prior to and following the fall and spring terms.3 Survey administration timelines and response rates are listed in Supplemental Appendix Table 1.4 (Note: Appendices are available in the online version of this article.) Survey completers were offered a US$25 gift card. Using registration data and additional CT data collected by CPS, which were drawn from all “traditional” university-based programs that used CPS for student-teaching placements, we identified the CTs for all registered PSTs and sent them individualized online surveys at the end of the fall and spring terms.5 Survey completers were offered a US$50 gift card. We linked PST and CT survey information to CPS personnel and evaluation data, and to data on their schools.6

Sample

Of our initial population of 1,122 PSTs in CPS during 2014–2015, 305 subsequently gained employment in CPS during the 2015–2016 school year and could be linked to their first-year observational ratings; this constituted our main analytic sample for this study.7 Of these 305 PSTs who could be linked to first-year observational ratings, 225 could be linked to measures of PSTs’ self-perceived preparedness based on post-student-teaching surveys, and 226 could be linked to measures of CTs’ perceptions of PST preparedness based upon CT surveys.8

PST characteristics. Table 1 summarizes characteristics of PSTs. PSTs were mostly female (72%) and White (57%). About 29% graduated from a CPS high school and about one-third (31%) reported either teaching full/part-time or substitute teaching in a school/child care facility prior to beginning their program. Supplemental Appendix Table 2 (top; available in the online version of this article) compares PSTs in our analytic sample with PSTs in the nonanalytic sample (those who student-taught in CPS but not employed there the following year and/or could not be linked to observational ratings). Generally, the analytic sample had more Latinos, more CPS graduates, higher undergraduate GPAs, and fewer “other” race/ethnicity PSTs, suggesting the analytic sample is not representative of Chicago PSTs. Among those in our analytic sample, we also compared the characteristics of PSTs who did (n = 225) and did not (n = 80) complete surveys, finding no differences (Supplemental Appendix Table 3; available in the online version of this article).

CT characteristics. Table 1 summarizes CT characteristics and demonstrates that CTs were mostly female, White, and tenured. Supplemental Appendix Table 2 (middle) compares CTs in our analytic sample to those in our nonanalytic sample, finding no significant differences. We compared CTs who did (n = 226) and did not (n = 79) complete surveys and found they were similar (see Supplemental Appendix Table 3).

FPS characteristics. FPS characteristics are shown in Table 2. One third of PSTs completed their student teaching in high school (9–12) placements and about two thirds were in elementary (K-8) schools. While almost half of PSTs (46%) completed student teaching in a FPS without a Black or Latino student majority,9 41% were in a majority Latino school and 14% were in a majority Black school. The average school-level prior-year achievement was above average (0.15 SD units). On average, PSTs perceived their working conditions at their FPS positively, with nearly two-thirds of PSTs (63%) rating their FPS working conditions at an average of 3.5 out of 4 or higher.10 The Five Essential11 values
Supplemental Appendix Table 2 (bottom) compares FPS characteristics between PSTs in our analytic sample and those not in our analytic sample. The main difference was that PSTs in our analytic sample completed student teaching in schools with higher average poverty concentration.

**Employed school characteristics.** Table 2 (bottom) summarizes the characteristics of the schools in which PSTs gained employment during school year 2015–2016. First, 53 (17.4%) PSTs had student-taught in the very school in which they became employed. Employed PSTs worked mostly in K-8 primary settings (78%); 39% of PSTs were employed in majority Latino schools and 26% were employed in majority Black schools.

**Student teaching and coaching characteristics**

Table 3 summarizes the characteristics of student teaching and coaching experiences. About three-quarters of PSTs felt they learned a lot from their student teaching placements. On average, PSTs spent 214 hours in their student teaching experiences and completed more than 5 courses prior to these experiences. Less than half of PSTs (43.7%) were primarily lead teachers during their student teaching experiences, while about one-third of PSTs were placed in classrooms where they taught all subjects. Almost three-quarters of CTs had prior mentees but only 29% had received training on how to coach mentees.

**Measures**

In this section, we describe the focal outcome measures: PSTs’ and CTs’ perceptions of PSTs’ preparedness and...
employed PSTs’ 2015–2016 observational ratings. We also describe focal Rasch measures used as predictors.

**Perceived preparedness.** We asked PSTs and CTs a series of survey questions about PSTs’ preparedness to take on the responsibilities of teaching in four domains of instruction aligned with CPS’s teacher evaluation: planning and preparation, instruction, classroom environment, and professional responsibilities. See Supplemental Appendix Table 6 for information on preparedness survey items. We submitted these survey items to Rasch analysis to create domain-level measures and then standardized them for ease of interpretation; we used the standard errors associated with each domain to create precision-weighted mean measures in each domain. To create an overall measure of preparedness across domains, we divided the sum of the four precision-weighted domain-level measures by the sum of the four weights, giving us one precision-weighted measure for PST self-perceptions of preparedness and another for CTs’ perceptions of their PSTs’ preparedness.

**Observational ratings.** REACH Students (Recognizing Educators Advancing Chicago Students) is CPS’s system of educator evaluation and support. A significant component of the REACH evaluation system is its observations of practice scores, which are based on the Danielson-inspired CPS Framework for Teaching. REACH requires that all administrators gain certification by completing a series of training modules; trained specialists provide ongoing support for calibrating their rating performance. All beginning teachers are rated by their principals or assistant principals a total of 4 times (three formal and one informal). Evaluators score teachers from 1 to 4 on 19 components, and each domain score consists of ratings from 4 to 5 components. The overall “observational ratings” measures in this study are the average of the four domain scores.

Prior studies tested the validity and reliability of the CPS Framework for Teaching evaluation ratings by correlating them with student achievement and master observer ratings. Sartain et al. (2011) found that classroom observational ratings were valid measures of teaching practice; teachers with the highest ratings demonstrated the highest growth in student achievement, and teachers with the lowest ratings demonstrated the lowest growth. They also found the observational ratings were reliable measures of teaching practice, where principals and master raters consistently gave the same ratings of the same classrooms. That said, principals tended to assign more “Distinguished” ratings. During the pilot, principals rated 17% of teachers as “Distinguished,” 53% as “Proficient,” 27% as “Basic,” and 2% as “Unsatisfactory.” By contrast, master observers classified only 3% as “Distinguished,” 67% as “Proficient,” 28% as “Basic,” and 2% as “Unsatisfactory” (Sartain et al., 2011). Although principals were more likely than master observers to rate teachers as “Distinguished,” principals were far more discerning during the pilot of the new evaluation system than they had been previously. Previously, 93% of teachers were evaluated as Superior or Excellent, as compared with 70% during the pilot.

### Table 3. Characteristics of Student Teaching and Perceptions of Coaching.

<table>
<thead>
<tr>
<th>Student teaching and coaching characteristics</th>
<th>Percent/M</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student teaching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student-taught all subjects (i.e., elementary)</td>
<td>32.4%</td>
<td>222</td>
</tr>
<tr>
<td>Number of courses taken prior to student teaching</td>
<td>5.4</td>
<td>252</td>
</tr>
<tr>
<td>Number of total hours spent student teaching</td>
<td>214.2</td>
<td>222</td>
</tr>
<tr>
<td>PST strongly agrees she or he learned a lot from student teaching</td>
<td>74.1%</td>
<td>220</td>
</tr>
<tr>
<td>TEP chose the placement school</td>
<td>56.3%</td>
<td>222</td>
</tr>
<tr>
<td>Primarily lead teacher during student teaching</td>
<td>43.7%</td>
<td>222</td>
</tr>
<tr>
<td>PST-Perceived field instructor supporta</td>
<td>-0.00</td>
<td>217</td>
</tr>
<tr>
<td>PST perceptions of CT coaching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain-specific conversationsa</td>
<td>-0.06</td>
<td>219</td>
</tr>
<tr>
<td>Coaching relationship and feedbacka</td>
<td>0.02</td>
<td>220</td>
</tr>
<tr>
<td>Assistance with job searcha</td>
<td>0.06</td>
<td>220</td>
</tr>
<tr>
<td>CT perceptions of own coaching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain-specific coachinga</td>
<td>-0.01</td>
<td>227</td>
</tr>
<tr>
<td>Frequency of feedbacka</td>
<td>-0.08</td>
<td>227</td>
</tr>
<tr>
<td>Assistance with job searcha</td>
<td>0.07</td>
<td>226</td>
</tr>
<tr>
<td>CT’s teaching strongly aligns with TEP</td>
<td>36.7%</td>
<td>205</td>
</tr>
<tr>
<td>CT has had training in coaching</td>
<td>29.0%</td>
<td>231</td>
</tr>
<tr>
<td>CT has had prior mentees</td>
<td>74.7%</td>
<td>233</td>
</tr>
</tbody>
</table>

Note. For more information about the coaching (Rasch) measures, see Supplemental Appendix Table 7. Unlike the other items in this table which are from the post-survey only, we asked about number of courses taken prior to student teaching on both the pre- and the post-surveys which is why the sample is greater for this item.

*Standardized Rasch measures; as these were standardized based on Rasch measures in the full sample, the means and standard deviations differ slightly from 0 to 1 for the analytic sample, shown here. PST = preservice teacher; CT = cooperating teacher; TEP = teacher education program.
Recently, Jiang and Sporte (2016) found that schools with the most marginalized students had an overrepresentation of teachers with the lowest observational and value-added scores. Moreover, teachers of color and male teachers tended to get lower ratings. These results are consistent with studies outside of Chicago that have found observational ratings to be associated with characteristics of teachers and their students (Campbell & Ronfeldt, 2018; Steinberg & Garrett, 2016). Given these concerns, when modeling effects on first-year observation ratings, we include employment school characteristics as covariates in all models; while these adjustments may not remove all forms of bias, they likely minimize them.

**Predictor Rasch measures.** We used eight Rasch measures as predictors (see Supplemental Appendix Table 7). Five Rasch predictors included measures of PST-perceived: (a) CT teaching effectiveness, (b) field instructor helpfulness, (c) instructional domain-specific conversations, (d) coaching relationship and feedback, and (e) job search support. For CTs, we used three Rasch measures of CT-perceived: (a) instructional domain-specific coaching, (b) frequency of feedback, and (c) job search support.

**Analytic Method**

**Research Question 1:** To estimate PSTs’ observational ratings in 2015–2016 as a function of features of preparation, we used two-level hierarchical linear models with PSTs at Level 1 and employment school at Level 2. The general form of the model is summarized in Equation 1:

\[
\text{RATING}_{ij} = \gamma_{00} + \gamma_{10} \text{Feat}_{ij} + \gamma_{20} \text{ES}_{ij} + \gamma_{30} \text{Z}_{ij} + u_{0j} + r_{ij}
\]  

where the 2015–2016 observational rating of PST \(i\) in school \(j\) is a function of an intercept (\(\gamma_{00}\)), \(\text{Feat}_{ij}\) (features of preparation) is the focal predictor, \(\text{ES}_{ij}\) is a vector of employment (2015–2016) school characteristics (school/grade level, school area poverty concentration, student race composition, school prior achievement, and whether the PST student-taught in the current school), \(\text{Z}_{ij}\) (controls) is a vector of characteristics of the PST (race/ethnicity, gender) and FPS (school/grade level, school-level prior achievement), a school random effect is denoted as \(u_{0j}\), and a PST-level residual is denoted as \(r_{ij}\).

In Tables 6–9, we include two model specifications. Given prior research that observation ratings are associated with student/school characteristics, we include employment school characteristics in all models with observation ratings as outcomes. In Model A, we enter each preparation feature (\(\text{Feat}_{ij}\)) independently (each estimate is from a different model). Model B includes all preparation features together in the same model as well as controls for PST and FPS characteristics.

Model B is our preferred model as it adjusts for PST and FPS characteristics, as well as other preparation features, that could otherwise bias estimates. However, we present Model A estimates for many reasons. First, the inclusion of many covariates in Model B, from different data sets (each with limited sample coverage), substantially reduces sample. In addition to reducing statistical power, a risk is that the subsample included in Model B differs systematically from that in Model A. Second, including Model A (as a subset of Model B) allows us to check whether and how estimated parameters for program features change with the addition of PST/FPS characteristics and program features. Third, some focal preparation features included together in Model B are conceptually similar to one another, thus presenting possible collinearity concerns and challenges in interpreting estimates; for example, CT qualifications such as observational ratings, experience, and National Board status are closely related to one another. Finally, including Model A allows us to consider estimates on preparation features that are of interest to many readers but not included in Model B; for example, estimates on CT VAM scores are not presented in Model B due to poor sample coverage but are of interest to many researchers, policymakers, and practitioners.

**Research Questions 2 and 3:** To estimate PSTs’ and CTs’ perceptions of PST preparedness (survey-based measures) as a function of features of preparation, we use Equation 1 but replace observational ratings with these survey-based outcome measures. As all data collection for these analyses occurred prior to PSTs completing preparation, we nest PSTs in teacher education programs instead of employment schools and drop employment school characteristics (\(\text{ES}\)) from all models.

**Results**

This study’s primary goal is to identify which features of preparation predicted graduates’ first-year observational ratings and whether those features also predict PSTs’ and their CTs’ perceptions of PST preparedness. We have organized the results according to the dimensions of preparation that are the focus of our Logic Model (left side): (a) CT instructional quality/professional qualifications, (b) CT coaching, (c) FPS characteristics, and (d) other features. For each dimension, we begin by describing which features predict first-year observational ratings; we then consider

| Table 4. Correlation Matrix of PSTs’ Overall Observational Ratings and Perceptions of Overall Preparedness. |
|---------------------------------------------------------|---------------------------------|---------------------------------|
| Outcome measures                                      | 1                              | 2                              | 3                              |
| 1. Overall Self-Perceived PST Preparedness            | 1.00                           |                                 |                                |
| 2. Overall CT-Perceived PST Preparedness              | .06                            | 1.00                           |                                |
| 3. Overall PST Observational Rating                   | .03                            | .24                            | 1.00                           |

Note. We also examined disattenuated correlations and results were very similar. PST = preservice teacher; CT = cooperating teacher.
First, we examine simple correlations between our three outcome measures: PSTs’ and CTs’ perceptions of PSTs’ preparedness and first-year observational ratings of PSTs. In particular, we examine whether the hypothesized relationships in our Logic Model (dashed lines) held (a) that PSTs would feel better prepared when their CTs perceived them as better prepared and (b) that PSTs would have better first-year observation ratings when they and their CTs felt they were better prepared. In fact, PSTs’ self-perceived preparedness had little association with either CTs’ perceptions of PSTs’ preparedness (0.06) or with their first-year observational ratings (0.03; see Table 4). However, CTs’ perceptions of PSTs’ preparedness were more moderately correlated (0.24) with PSTs’ first-year observational ratings.

Given that this study presents some of the first evidence for whether or not survey-based feelings of preparedness predict first-year teaching performance, we examined these relationships further in a regression framework so we could adjust for school and PST characteristics that could explain observed relationships. Table 5 summarizes results from regression models estimating PSTs’ first-year observational ratings as a function of the two focal self-perceived preparedness measures—PSTs’ perceptions (Row 1) and CTs’ perceptions (Row 2). Results suggest that PSTs were significantly more effective as first-year teachers when their CTs rated them as more prepared at the end of student teaching, both overall and in individual instructional domains. A 1 SD increase in how prepared CTs evaluate their PSTs as being (on a survey-based factor) is associated with an increase of 0.06 to 0.07 on PSTs’ first-year observation ratings. Based upon Jiang and Sporte (2016), in Chicago, this equates to slightly less than half the difference in observation ratings between a first year teacher and the average teacher with between 2 and 5 years of experience (β = .16). However, PSTs who felt better prepared were no more or less effective as first-year teachers—true across model specifications.

Having examined the degree to which these three measures of PSTs’ instructional quality are related to one another, we next investigate which features of preparation predict them. Tables 6 to 9 are organized the same way—with features of preparation presented as rows and outcome measures presented as columns.
Table 7. Perceived Preparedness and First-Year Effectiveness as a Function of CT Coaching.

<table>
<thead>
<tr>
<th>CT coaching measures</th>
<th>Self-perceived PST preparedness</th>
<th>CT-perceived PST preparedness</th>
<th>First-year observational rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>PST perceptions of CT coaching</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain-specific conversations (Rasch)</td>
<td>0.22***</td>
<td>0.04</td>
<td>0.08</td>
</tr>
<tr>
<td>Coaching relationship/feedback (Rasch)</td>
<td>0.33***</td>
<td>0.28***</td>
<td>0.21**</td>
</tr>
<tr>
<td>Assistance with job search (Rasch)</td>
<td>0.24***</td>
<td>0.11</td>
<td>0.33***</td>
</tr>
<tr>
<td>N</td>
<td>211</td>
<td>163</td>
<td>210</td>
</tr>
<tr>
<td>CT perceptions of own coaching</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domain-specific coaching (Rasch)</td>
<td>−0.02</td>
<td>−0.14</td>
<td>0.45***</td>
</tr>
<tr>
<td>Frequency of feedback (Rasch)</td>
<td>−0.04</td>
<td>−0.01</td>
<td>0.14*</td>
</tr>
<tr>
<td>Assistance with job search (Rasch)</td>
<td>−0.00</td>
<td>0.08</td>
<td>0.32***</td>
</tr>
<tr>
<td>CT’s teaching strongly aligns with TEP</td>
<td>0.05</td>
<td>0.12</td>
<td>0.61***</td>
</tr>
<tr>
<td>CT coaching experience</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT had training in coaching</td>
<td>−0.02</td>
<td>−0.08</td>
<td>−0.06</td>
</tr>
<tr>
<td>CT had prior mentees</td>
<td>0.08</td>
<td>0.07</td>
<td>0.19</td>
</tr>
<tr>
<td>N</td>
<td>149</td>
<td>197</td>
<td>198</td>
</tr>
</tbody>
</table>

Note. Model A includes each focal predictor (row) independently (each coefficient is from a different model). Model B includes predictors (rows) together in same model and also includes PST and FPS characteristics. All models with first-year observation ratings also include employment school characteristics as covariates. See the “Methods” section for details. CT = cooperating teacher; PST = preservice teacher; TEP = teacher education programs; FPS = field placement school.

Table 8. Perceived Preparedness and First-Year Effectiveness as a Function of Field Placement School (FPS) Characteristics.

<table>
<thead>
<tr>
<th>Field placement school characteristics</th>
<th>Self-perceived PST preparedness</th>
<th>CT-perceived PST preparedness</th>
<th>First-year observational rating</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>Employed in field placement school</td>
<td>0.15**</td>
<td>0.15**</td>
<td></td>
</tr>
<tr>
<td>Concentration of poverty</td>
<td>−0.13</td>
<td>−0.19</td>
<td>−0.19</td>
</tr>
<tr>
<td>High-school level (9–12)</td>
<td>−0.14</td>
<td>0.04</td>
<td>−0.03</td>
</tr>
<tr>
<td>Majority Black students</td>
<td>0.11</td>
<td>0.36</td>
<td>−0.21</td>
</tr>
<tr>
<td>Majority Latino students</td>
<td>−0.06</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>School prior achievement</td>
<td>0.13</td>
<td>0.14</td>
<td>−0.11</td>
</tr>
<tr>
<td>PST-perceived working conditions (3.5 or above, out of 4)</td>
<td>0.51***</td>
<td></td>
<td>0.28†</td>
</tr>
<tr>
<td>5E: Involved families</td>
<td>0.01</td>
<td></td>
<td>−0.00</td>
</tr>
<tr>
<td>5E: Collaborative teachers</td>
<td>−0.00</td>
<td></td>
<td>−0.01</td>
</tr>
<tr>
<td>5E: Ambitious instruction</td>
<td>0.00</td>
<td></td>
<td>−0.01†</td>
</tr>
<tr>
<td>5E: Supportive environment</td>
<td>−0.00</td>
<td></td>
<td>−0.01</td>
</tr>
<tr>
<td>5E: Effective leader</td>
<td>0.00</td>
<td></td>
<td>−0.01</td>
</tr>
<tr>
<td>N</td>
<td>217</td>
<td>220</td>
<td>289</td>
</tr>
</tbody>
</table>

Note. Model A includes each focal predictor (row) independently (each coefficient is from a different model). Model B includes predictors (rows) together in same model and also includes PST and FPS characteristics. All models with first-year observation ratings also include employment school characteristics as covariates. See the “Methods” section for details. CT = cooperating teacher; PST = preservice teacher.

as columns. Model A includes each feature independently as a predictor; Model B includes a set of features together (those with coefficients listed), and also controls for PST and FPS characteristics. Both models (A and B) of first-year observation ratings follow the same logic and also control for employment school characteristics. See “Analytic Method” section for rationale.

CT Instructional Quality/Qualifications

Table 6 summarizes results from models estimating our three outcomes as a function of CTs’ instructional quality/professional qualifications. Beginning with first-year observational ratings (far right), PSTs whose CTs were rated as more instructionally effective themselves had significantly stronger observational ratings.
Every additional point (on a scale of 1–4) in CTs’ observational ratings was associated with a 0.16 point gain for PSTs’ observational ratings, equivalent to the average difference between a first-year teacher and a teacher with between 2 and 5 years of experience (Jiang & Sporte, 2016).17

Other CT professional qualifications were either unrelated or negatively related to PSTs’ first-year effectiveness. Having a CT with tenure, more experience, or National Board certification was mostly unrelated to PSTs’ first-year ratings, though the latter was negative and significant in Model B. We caution about over-interpreting these results, given that Model B includes predictors simultaneously, including ones (e.g., observation ratings) which likely explain some of the effect of National Board. Finally, when PSTs rated their CTs as more effective teachers, their own first-year performance was significantly lower.

Turning from first-year observational ratings, we considered PSTs’ and CTs’ perceptions of PST preparedness (Table 6, left and middle). When PSTs rated CTs as more instructionally effective in observation-aligned domains, they felt significantly better prepared; though they felt better prepared, these PSTs actually received worse first-year ratings (see above). PSTs felt better prepared when their CTs had taught longer in CPS, though results were nonsignificant in Model B.

CT Coaching

Table 7 summarizes results from models investigating CT coaching measures as predictors, including PSTs’ perceptions of CTs’ coaching (top) and CTs’ perceptions of their own coaching, as well as their coaching experience and training (bottom). In terms of PSTs’ first-year observational ratings (Table 7, right), only CTs’ coaching aligned with the district observational rubric was significantly (positively) related. Otherwise, perceptions of CTs’ coaching and their actual training and experience as coaches were mostly unrelated. One possible exception is that PSTs received lower observational ratings when they reported that their CTs had provided better job support; however, estimates were significant only after adjusting for coaching in other areas.

Regarding CTs’ perceptions of PSTs’ preparedness (middle columns), generally CTs rated their PSTs as better prepared when either they or their PSTs reported stronger coaching. By contrast, while PSTs’ perceptions of coaching predicted their own levels of self-perceived preparedness, CTs’ perceptions of coaching were unrelated to PSTs’ self-perceptions of preparedness. Finally, CTs having more training or experience in coaching were unrelated to all perceptions of preparedness. Because both predictors and outcomes described above are based upon survey measures, they are prone to many forms of subjectivity and bias, and thus may not reflect truly causal effects. For example, CTs who are placed with more promising PSTs may feel they provide better coaching whether or not they do; if so, the relationship between CTs’ perceptions of their coaching and their PSTs’ readiness to teach could reflect this bias rather than a causal effect of coaching on PST readiness.

FPS Characteristics

As summarized in Table 8, we found that PSTs were significantly more effective as first-year teachers when they were hired into the schools in which they student-taught. Being a first-year teacher in a familiar FPS was associated with a 0.15 point increase in first-year observational ratings. Although prior research has indicated FPS achievement and working conditions positively predict graduates’ VAM, we found PSTs (a) received lower observational ratings (in Model B) when their FPS had better average achievement and (b) received no higher or lower observational ratings based on their perceptions of FPS working conditions. FPS
grade level, socioeconomic, and racial composition appeared unrelated to first-year teaching effectiveness.

The majority of FPS characteristics we tested were unrelated to all perceptions of PST preparedness, with one exception: PSTs’ perceptions of working conditions. Although unassociated with first-year observational ratings, PSTs felt better prepared when they perceived better FPS working conditions. Finally, CTs felt their PSTs were less prepared when their FPS had better average achievement, but estimates were significant only in Model B. None of the 5E measures significantly predicted any focal outcome.

Other Features of Preparation

Table 9 summarizes results from models examining associations between other features of preparation and the three focal outcomes. First, elementary (or all-subject) PSTs had higher observational ratings across models. In one specification, PSTs also received higher observational ratings when their programs were primarily responsible for selecting their FPS. Finally, in one model specification, PSTs who reported taking more courses before student teaching had lower observational ratings.

On the whole, CTs did not feel PSTs were better or worse prepared as a function of these other features of preparation (Table 9, middle). However, some of these features significantly explained PSTs’ own feelings of preparedness. Taking more courses prior to student teaching, spending more hours in the placement, feeling that student teaching was instructive, primarily being a lead teacher during student teaching, and finding the field instructor to be helpful were all associated with PSTs feeling better prepared.

Do the same features of preparation that predict one outcome also predict the others? Ideally, promising features of preparation that make PSTs feel better prepared would also predict better CT ratings and first-year observational evaluations. However, we found no feature of preparation positively associated with all three outcomes used in this study. Furthermore, only one feature of preparation positively predicted PSTs’ observational ratings and either survey-based outcome. Namely, when CTs felt they provided stronger coaching in specific instructional domains (including those evaluated on the district rubric), they also rated their PSTs as being better instructionally prepared and their PSTs received stronger first-year observational ratings. And, only two predictors were positively related to both PSTs’ and CTs’ perceptions of PST preparedness: PSTs reporting better coaching relationships and better job support from CTs.

Feeling more prepared, but being less effective. A few features of preparation were associated with PSTs feeling significantly better prepared after student teaching but receiving significantly lower observational ratings as first-year teachers. First, PSTs who thought their CTs were more effective teachers felt better prepared but also had significantly lower first-year observation ratings. Similarly, taking more courses before student teaching helped PSTs feel more prepared but predicted significantly lower first-year ratings.

Seeming more prepared, but being less effective. One feature was associated with higher CT ratings of preparedness but lower first-year teaching effectiveness. When CTs reported providing more job search assistance, they thought their PSTs were better prepared, but their PSTs had lower first-year observational ratings.

Limitations

It is important to acknowledge that the estimates reported in this article are correlational in nature and may not reflect causal effects. Many forms of selection could explain the relationships we observe, including selection of PSTs to certain kinds of (a) preparation programs, (b) FPS, (c) CTs, and (d) employment schools.

These forms of selection might introduce bias into our estimates for the relationships between CT and PST instructional effectiveness, for instance, and potentially explain the positive associations we observe. It could be that some teacher education programs (TEPs), or FPS they use, produce more instructionally effective teachers; if these same TEPs (or FPS) happen to use more instructionally effective CTs—even where these CTs are not actually causing gains in PST performance—then the positive relationship between CT and PST instructional effectiveness might be explained by the effects of TEPs (or their FPS). By adjusting our models for TEP and FPS characteristics, we try to account for these alternative explanations; however, it is possible that unobserved characteristics of TEPs and FPS could account for differences we observe.

It is also possible that within TEPs, the most promising PSTs sort to the most instructionally effective CTs. Thus, the association we observe between PST and CT instructional effectiveness could simply reflect this initial sorting rather than more instructionally effective CTs causing PSTs to improve. However, we are not aware of any way to—prior to student teaching—measure the potential for PSTs to become instructionally effective once employed. Thus, we have no way to satisfactorily test or adjust for this explanation. We do the best we can by controlling for undergraduate GPA, prior teaching experience, and other PST characteristics. However, prior literature suggests that these are weak predictors of later performance; also, unobserved characteristics (e.g., innate potential) could still explain observed relationships.

Other literature suggests that there is a lot of within- and between-program variation in how PSTs are matched to CTs, even within Chicago (Boyd et al., 2008; Matsko et al., under review; Mullman & Ronfeldt, under review; Ronfeldt et al., 2013; St. John et al., 2018). For example, some programs leave
selection to PSTs themselves, others take primary responsibility for selection, and still others depend upon district and school leaders. If PSTs with the most potential sort to the most instructionally effective CTs, this would require that various stakeholders (a) evaluate future potential prior to student teaching and (b) use their assessments of potential in the same ways—by placing the most promising PSTs with the most instructionally effective CTs. Some qualitative and anecdotal evidence suggest that the latter is unlikely as school leaders are sometimes inclined to place the most promising PSTs with less effective teachers as a way to offer them an “extra pair of hands” (Mullman & Ronfeldt, under review). Even so, we cannot rule out that this form of selection may explain the relationships we observe.

Finally, it could be that certain PSTs are inclined to seek employment in schools that are more supportive of teacher learning and performance. When these PSTs enter the job market, they apply for (and thus are more likely to obtain) positions in schools that will likely boost their performance relative to their peers. In this scenario, the employment schools—and not CTs (or any preservice factors)—are actually causing differences in performance. This scenario could explain the observed association between PST and CT instructional effectiveness if these same PSTs also sort to more supportive schools during student teaching—schools that, being more supportive, also tend to have more instructionally effective CTs. In our main model specifications, we adjust for numerous employment school characteristics to control for their effects and better isolate the effects of preservice features; it is still possible that unobserved characteristics of employment schools boosts inservice teacher performance while also happening to be associated with CT instructional effectiveness.

Furthermore, our results may be biased in that PSTs who responded to surveys had different characteristics than those that did not (see Supplemental Appendix Table 2). In addition, individuals employed in CPS (who, thus, have first-year observation ratings) are likely to differ from those who did not gain employment.

**Discussion**

This is the first study to our knowledge that links features of preparation to graduates’ first-year instructional effectiveness as measured by observational evaluations, identifying a number of features positively associated with first-year effectiveness. Although we acknowledge that the associations we have identified may not be causal, we also discuss implications (in the following) assuming they may be. Future studies should focus on establishing causal estimates, including studies that randomly assign PSTs to features we have found to be associated with instructional quality.

Of note, this study suggests that the instructional effectiveness of CTs is positively associated with the instructional effectiveness of their PSTs. Assuming future studies confirm this relationship to be causal, then this finding offers support for policies that set minimum requirements for instructional effectiveness for CTs. This conclusion is bolstered by the fact that two recent studies have found similar, positive correlations between the instructional effectiveness of recent graduates and their CTs (Goldhaber et al., 2018; Ronfeldt et al., 2018).

This finding also suggests that recruiting instructionally effective teachers to serve as CTs should be a priority of preparation programs and school and district leaders. Having more effective CTs could benefit programs by improving clinical experiences they offer. Because roughly two out of five new hires are employed in the districts in which they completed their student teaching (Ronfeldt et al., 2018), and one in five in the specific schools in which they student-taught (see above), school and district leaders may have access to a more effective supply of new teachers should they recruit their instructionally effective teachers to serve as CTs. In addition to being strong models for PSTs, school and district leaders should likely seek CT input when considering hiring the PSTs with whom they worked as CTs’ perceptions of PSTs’ instructional effectiveness were positively associated with PSTs’ first-year performance.

Consistent with Matsko et al. (2018), this study also suggests that, while being a model of effective instruction likely matters, it is important that CTs be able to provide quality coaching too. Specifically, we find that PSTs had better first-year observational ratings when their CTs reported providing stronger coaching in specific instructional areas aligned with the district’s observational rubric. A likely implication is then that CTs receive training in how to provide coaching in specific instructional areas that are included in state/district evaluation rubrics.

Given that coaching PSTs is challenging work, that only 15% of traditional pathway CTs in Chicago receive financial compensation, and that those who receive compensation receive only US$226 on average (Matsko et al., under review), leaders might consider ways to incentivize and support more instructionally effective teachers to serve. Possibilities include release time for coaching responsibilities, additional compensation, promotion to a full or part-time position as a district mentor for inservice and PSTs, or relief from being evaluated while serving as CTs.

The finding that graduates hired into the schools in which they student-taught received higher first-year observational ratings suggests that school leaders can benefit from allowing PSTs to learn to teach in their schools and use clinical placements as potential hiring opportunities. Given that nearly 20% of first-year teachers student-taught in their school, it seems many school leaders are already capitalizing on this opportunity. It is possible that PSTs learn context-specific knowledge/skills in their FPS and that leaders noticed (or talked to CTs about) particularly promising PSTs, both of which would make PSTs more appealing for hiring.
The final feature of preparation that positively predicted observational ratings across model specifications was student teaching in an elementary, self-contained placement. We initially speculated that this unexpected relationship may reflect PST selection rather than a causal story. In particular, prior research has found that elementary teachers receive better observational ratings than high-school teachers (Harris et al., 2014). Thus, we hypothesized that the positive estimate on elementary/self-contained placements might simply proxy for eventually being employed as an elementary teacher. However, the results were robust to a number of alternative specifications that attempted to adjust for selection (e.g., adding fixed effects for being employed in an elementary school and a variety of PST and school covariates). If this result does reflect a causal relationship, one possible explanation is that PSTs benefit by having training in multiple subject areas. For example, teachers of mathematics likely will benefit from also being effective teachers of English (Master et al., 2017). Meanwhile, others argue that specialized and subject-specific training may be more beneficial than general training (Ost, 2014; Shulman, 1987).

While Ronfeldt (2012, 2015) found that graduates had better VAM when they learned to teach in FPS with better working conditions and school-level achievement gains, we found FPS working conditions as measured by the Five Essentials and based on PSTs’ own perceptions to be unrelated to graduates’ first-year observational ratings. Moreover, FPS achievement was negatively related to first-year observational ratings, though at significant levels in only one specification. It is possible that these mixed results are due to the fact that these studies focused on different measures of instructional quality—VAM scores versus observational evaluations. Another possibility is that these features of preparation function differently in different labor markets. More research is needed to interrogate these mixed findings.

This study also extends prior research that uses PSTs’ self-perceived preparedness to identify promising features of preparation. Consistent with prior work, it finds that PSTs felt better prepared when they rated their CTs as being strong models of instruction and as providing stronger coaching (Matsko et al., 2018). Also consistent with prior work, PSTs felt better prepared when they reported more student teaching, more coursework prior to student teaching, better field instructor support, better student-teaching experiences generally, and better FPS working conditions (Ronfeldt, 2012; Ronfeldt et al., 2013, 2014). Yet, none of these features of preparation predicted stronger first-year observational ratings. To the degree that program leaders and policymakers intend to use outcomes-based research to identify promising levers for program improvement, an implication is that they may first determine which outcome they want to impact. Rather than certain features positively predicting all outcomes, different features predicted different outcomes.

Finally, this is the first analysis to our knowledge testing whether PSTs who report feeling better prepared are actually more effective first-year teachers, and we find this to not be the case. Our findings raise questions about using perceived preparedness as proxies for instructional readiness, but these measures may still have some place in program evaluation given prior evidence that they predict other outcomes we care about (e.g., self-efficacy and retention). However, we find CTs’ perceptions of their PSTs’ preparedness to have some predictive validity. This may suggest that teacher educators who rely on measures of PST perceptions of preparedness to assess graduate readiness to teach may gain more insights from the perceptions of their CTs.

Findings from this study offer evidence in support of many aspects of our Logic Model. In particular, our study offers evidence that each of the domains of preparation (CT modeling, CT coaching, FPS characteristics, and other/structural features) contribute to readiness to teach—at least one measure from each domain of preparation was positively and significantly related to at least one of the focal outcomes. While this may suggest that these specific preparation domains have some relationship to instructional readiness, we cannot conclude from this work that some domains matter more than others. In fact, it appears that the individual features within domains are a more meaningful level to study. Future research should examine whether the same or different features predict better PST outcomes to build a robust evidence base on which features of preparation contribute most to PST readiness.

Author Note
Kavita Kapadia Matsko is now affiliated to Northwestern University, Evanston, IL, USA.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
We are grateful to the Spencer Foundation for their generous support of this research.

ORCID iDs
Matthew Ronfeldt https://orcid.org/0000-0003-1702-1812
Kavita Kapadia Matsko https://orcid.org/0000-0003-1653-6750

Supplemental Material
Supplemental material for this article is available online.

Notes
1. By “direct” we mean that it is a measure based upon direct observation of teaching, as compared with VAMs which infer teaching quality based upon learning outcomes. This distinction
is important because, as Fenstermacher and Richardson (2005) note, high-quality teaching may not always yield high-quality learning. Likewise, high-quality learning may result from factors beyond high-quality teaching. Thus, it is not always appropriate to infer teaching quality based upon learning quality.

2. CPS at a glance (website) http://cps.edu/About_CPS/At-a-glance/Pages/Stats_and_facts.aspx

3. To accommodate multiple start dates and placement lengths, CPS administered pre-practicum teaching surveys twice during the year.

4. PSTs who had placements longer than a term or with two different CTs received the survey twice.

5. CTs who worked with multiple PSTs received separate surveys for each PST.

6. We used registration information to link some PSTs with missing CT information to CPS data on their FPs.

7. N = 321 of the original sample (n = 1,122) could be linked to observational ratings; however, we had no information about preservice preparation (including surveys) for 16 of these individuals.

8. In some cases, PSTs had multiple CTs. Of the 305 PSTs in our analytic sample, 27 PSTs matched to 2 CTs. In these cases, for continuous CT variables, we averaged the values. For dichotomous CT variables, we counted all responses; for example, a PST with a male CT one term and a female CT the next was counted as having a male CT and a female CT. We treated school characteristics in the same ways.

9. We define “majority” Black (or Latino) schools as having at least 70% of students who are Black (or Latino).

10. Our working conditions measure is the mean of four survey items about how strongly PSTs agreed with statements about the working conditions of their FPs; these items were on a 4-point scale where 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree.

11. Through an extensive research program, Tony Bryk and colleagues identified the “Five Essentials” which are characteristics of schools important for supporting school improvement.

12. Both item difficulties and respondent abilities are placed on the same scale and expressed in logits. Most measures in this study met minimum reliability thresholds (0.8).

13. See http://www.cps.edu/ReachStudents/Pages/AtaGlance.aspx for more information for CPS “REACH” system.

14. We multiplied each standardized domain-specific measure by 1/standard error^2_domain or its weight.

15. Seven PSTs were missing observational ratings on the planning and preparation and professional responsibilities domains, so we imputed mean values in these rare cases.

16. We retained this final set of covariates after testing other potential covariates (e.g., GPA, FPS race composition) that either reduced the sample, did not contribute to model fit, and/or were conceptually redundant of retained covariates.

17. See Jiang and Sporte (2016) Table 3, page 18. Ronfeldt and Campbell (2016) find the average difference between first and second year teachers on the TEAM rubric is 33% of a standard deviation which is consistent with the difference of 40% of a standard deviation unit (0.16 points) found here.

18. CT observational ratings are positively related to both PST first-year evaluation scores and PST self-perceived preparedness, though the latter is nonsignificant in models using our analytic sample. In alternative specifications including all PSTs employed in CPS (whether or not they have first-year observational ratings), we find the relationship to be statistically significant; moreover, in prior work with the full population of survey respondents, we also find this relationship to be significant (Matsko et al., 2018).

19. Completing more hours of student teaching positively predicted PSTs’ self-perceived preparedness. This predictor also trended positive on CTs’ perceptions of PST preparedness, though was not statistically significant. However, in alternative specifications that included all PSTs employed in CPS (whether or not they had first-year evaluation scores) and in prior work (Matsko et al., under review), results were positive and significant at the p < .05 level.

20. In alternative model specifications, we also added TEP fixed effects that should adjust for unobserved characteristics of TEPs; results were similar. We did not have enough observations per school to justify the addition of field placement or employment school fixed effects.

21. Higher percentages of CTs in alternative (33%) and residency (44%) pathways reported receiving financial compensation; they also received more, on average (about US$321 and US$1,793, respectively).

References


Author Biographies

Matthew Ronfeldt is Associate Professor of Educational Studies at the University of Michigan. His scholarship seeks to understand how to improve teaching quality. His research sits at the intersection of educational practice and policy and focuses on preservice and inservice teacher education, teacher retention, working conditions in schools, and the assessment of teaching and teacher education programs.

Kavita Kapadia Matsko is Associate Professor and Associate Dean for Teacher Education at Northwestern University’s School of Education and Social Policy. Her interests focus on the study and design of teacher preparation, mentoring, and induction, with a particular interest in the features of clinical preparation that promote teacher development.

Hillary Greene Nolan completed her PhD in Educational Studies at the University of Michigan in 2019 with a concentration in Teaching and Teacher Education. Her research focuses on the relational aspects of teaching, how teachers balance students’ academic and other needs, and how teachers develop their relational capacity. She currently teaches middle school in Rhode Island.

Michelle Reininger is the former Executive Director of Stanford University’s Center for Education Policy Analysis. Her research is largely focused on the dynamics of teacher labor markets including preparation, recruitment, and retention.