



Does Cooperating Teachers' Instructional Effectiveness Improve Preservice Teachers' Future Performance?

Matthew Ronfeldt¹, Stacey L. Brockman¹, and Shanyce L. Campbell²

Increasingly, states and teacher education programs are establishing minimum requirements for cooperating teachers' (CTs) years of experience or tenure. Undergirding these policies is an assumption that to effectively mentor preservice teachers (PSTs), CTs must themselves be instructionally effective. We test this assumption using statewide administrative data on nearly 2,900 PSTs mentored by over 3,200 CTs. We find the first evidence, of which we are aware, that PSTs are more instructionally effective when they learn to teach with CTs who are more instructionally effective. Specifically, when their CTs received higher observational ratings and value-added to students' achievement measures (VAMs), PSTs also received higher observational ratings and VAM during their first years of teaching; CTs' years of teaching experience, though, were mostly unrelated to these outcomes. These findings have implications for teacher education program leaders and policymakers who seek to recruit and set requirements for CTs who are more likely to support PSTs' future instructional effectiveness.

Keywords: correlational analysis; hierarchical linear modeling; mentoring; policy; policy analysis; school/teacher effectiveness; student teaching; teacher education/development

Increasingly, state policymakers and education leaders are establishing minimum requirements for cooperating teachers' (CTs)¹ years of experience, tenure, and instructional effectiveness (National Research Council, 2010; NCATE, 2010). Eleven states require teachers to have at least three years of experience to serve as a CT, while two states (Florida, Tennessee) require teachers to be "instructionally effective" according to state evaluation measures (Greenberg, Pomerance, & Walsh, 2011). Undergirding such policies is an assumption that to be effective mentors of preservice teachers (PSTs), CTs must themselves be instructionally effective. If this assumption were true, then we would expect instructional effectiveness, and other workforce outcomes, to be stronger among teachers who were mentored by more instructionally effective CTs.

We are aware of only two large-scale prior studies that link measures of CTs' instructional effectiveness to PSTs' later workforce outcomes (Goldhaber, Krieg, & Theobald, 2014; Matsko, Ronfeldt, Greene, Reininger, & Brockman, in press). Looking across six major education preparation providers in Washington

state, Goldhaber and colleagues (2014) found that CTs' value-added to student achievement measures (VAMs), years of teaching experience, and educational attainment were all unrelated to PSTs' likelihood of gaining employment. Because CTs model teaching and not necessarily employment, one might expect their instructional quality and qualifications to influence PSTs' instructional quality more than their rate of employment. Only one study has tried to link CTs' instructional quality to measures of PSTs' instructional readiness. Drawing on surveys of all PSTs and CTs across Chicago Public Schools, Matsko et al. (in press) found that PSTs reported feeling better prepared in establishing classroom environment (but not in other instructional areas) when their CTs received stronger observational ratings. Additionally, PSTs felt better prepared when they rated the quality of the instruction modeled by their CTs higher. These results provide initial evidence that CTs' instructional quality is

¹University of Michigan, Ann Arbor, MI

²University of California, Irvine, CA

associated with PSTs' readiness to teach. A limitation of that study, however, is that it relied on self-reported outcomes, so although measures of CT effectiveness predicted PSTs feeling better prepared, we do not know if PSTs actually became more instructionally effective. The present study addresses this gap by focusing on observed measures of instructional quality: observational ratings and VAMs.

CT Coaching and PST Instructional Quality

Implicit in policies targeting CT qualifications and instructional quality is an assumption that CTs influence PSTs by modeling effective teaching. Despite this, most literature about CTs focuses on their coaching, rather than modeling, functions (Clarke, Triggs, & Nielsen, 2014). Several studies describe the feedback, instructional scaffolding, autonomy, and support that CTs provide (Grossman, Ronfeldt, & Cohen, 2012; Wilson, Floden, & Ferrini-Mundy, 2002). Importantly, we know of only two prior studies that have linked CT coaching practices to measures of PSTs' instructional readiness; both provide evidence that CTs' coaching practices are related to better PST outcomes. Matsko et al. (in press), summarized previously, found that PSTs felt more prepared in all instructional domains (planning/preparation, instruction, classroom environment, professional responsibility) when they rated their CTs' coaching positively. However, since the outcome in that study was self-reported feelings of preparedness, we do not know if PSTs actually were more instructionally effective as a result of CTs' coaching.

Addressing this limitation, Giebelhaus and Bowman (2002) demonstrated that PSTs who learned to teach with CTs trained to implement a specific coaching model received significantly stronger evaluations of classroom performance by external raters. The authors randomly assigned PSTs from two preparation programs to two groups of CTs: One was trained in the Praxis III/Pathwise coaching model for framing discussions, and the other received no training. Two trained external raters evaluated video of classroom performance in 19 criteria. Controlling for their pretreatment performance, PSTs mentored by CTs who received training on coaching outperformed their peers who worked with CTs in the control condition on 11 out of 19 criteria.

Finally, though few studies link *preservice* coaching to instructional effectiveness, a growing number of studies consider the effects of mentoring for *inservice* teachers. A meta-analysis of 37 studies about the effect of coaching on inservice teachers' instruction and student achievement finds effect sizes of 0.49 standard deviations on the former and 0.18 standard deviations on the latter (Kraft, Blazar, & Hogan, in press).

While good coaching seems to matter, good coaches are not necessarily the most effective teachers of p-12 students. Beyond coaching quality, is it important to have a CT who is also instructionally effective? Given that CTs directly model instruction for PSTs, often co-teach with them, assist with setting up classroom systems, and so on, one would expect that PSTs benefit from sharing classrooms with instructionally effective CTs. The literature is less clear, however, about the effects of having an instructionally effective CT. Though Matsko et al. (in press), described previously, provide some evidence that PSTs feel better prepared

when they learn to teach with more instructionally effective CTs, no study to date has investigated whether PSTs are more instructionally effective. Thus, this study asks:

1. Are PSTs more instructionally effective when they learn to teach with more instructionally effective CTs?
2. Are PSTs more instructionally effective in the same teaching domains in which their CTs excel?

Methods

Data and Sample

This study draws primarily on a statewide data set of PSTs from the Tennessee Department of Education (TDOE). Our data include information about sociodemographics, education preparation, and CT and field-placement school characteristics for approximately 27,000 PSTs who were prepared by 46 education preparation programs (EPPs) during the 2010–2011 through 2014–2015 academic years. Because CT information was unavailable from a number of EPPs, only about 4,700 PSTs could be linked to CTs and administrative data on the characteristics of CTs and the schools in which they worked. Among PSTs subsequently hired in Tennessee public schools during the 2012–2013 through 2015–2016 academic years, we linked state evaluation and school-level data to these PSTs. Our analytic sample includes 2,869 pre-service teachers from 21 EPPs who were mentored by 3,287 CTs within 898 field placement schools and were hired into 1,211 schools.

Table 1 displays summary statistics for the PSTs, CTs, field placement schools, and current schools in our analytic sample. Reflective of teachers nationally, the majority (79%) of the PSTs in our sample were women and White (94%); about 3% were Black, and another 3% were from other racial/ethnic groups, including Native, Hispanic, and Asian. Most PSTs were from Tennessee, with only a small number (4%) of out-of-state residents. Just over half of the PSTs were certified in an undergraduate EPP. In terms of endorsement areas, most PSTs (53%) were elementary endorsed, about 29% were secondary endorsed, and about 6% were endorsed in special education. Very few (less than 1%) had an alternative certification.

We compared the characteristics of PSTs in our analytic sample (those we could link to CTs) to all other PSTs from the state during this period (Appendix Table A2). As the table shows, the PSTs in our sample differ along several dimensions. Those in our analytic sample are, on average, more likely to be women (3 percentage point difference) and White (12 percentage point difference). Our sample has 17 percentage points more elementary endorsed PSTs and fewer PSTs with secondary, special education, or other endorsements. Twenty-five percent of the full sample of PSTs received an alternative certification; however, less than 1% of those in our analytic sample are alternatively certified.² The differences between our analytic sample and other PSTs from the same period are a limitation of this study and suggest that the EPPs that shared CT data with the TDOE are not representative of EPPs in the state; therefore, we must be cautious about generalizing findings to the full population of PSTs in Tennessee.

Table 1
Summary Statistics

	<i>n</i>	Proportion/Mean	<i>SD</i>
Preservice teacher characteristics			
Women	2,839	0.791	0.407
Black PSTs	2,839	0.0296	0.169
Other PSTs of color	2,839	0.0282	0.166
White PSTs	2,839	0.942	0.233
Age	2,551	29.28	6.846
Out-of-state resident	2,838	0.0433	0.203
Undergraduate degree	2,839	0.556	0.497
Elementary endorsed	2,839	0.533	0.499
Secondary endorsed	2,839	0.294	0.456
Special education endorsed	2,839	0.0645	0.246
Other endorsement	2,839	0.167	0.373
Alternative certification	2,838	0.006	0.0772
Not hired year after completion	2,869	0.225	0.418
Had multiple CTs	2,869	0.378	0.485
Overall observation rating	2,766	3.609	0.490
Instruction domain rating	2,492	3.512	0.471
Environment domain rating	2,484	3.981	0.575
Planning domain rating	2,480	3.655	0.526
Professionalism domain rating	2,398	3.889	0.568
Overall VAM score	1,632	-0.125	0.764
Math VAM score	906	-0.178	0.774
ELA VAM score	1,030	-0.127	0.758
Cooperating teacher characteristics			
Women	3,188	0.839	0.367
Black CTs	3,190	0.0451	0.208
Other CTs of color	3,287	0.0453	0.208
White CTs	3,190	0.939	0.240
Years of teaching experience	3,198	13.85	9.168
Overall observation rating	2,480	4.048	0.471
Instruction domain rating	2,137	3.942	0.504
Environment domain rating	2,118	4.378	0.559
Planning domain rating	2,122	3.994	0.634
Professionalism domain rating	2,072	4.294	0.576
Overall VAM score	1,705	0.176	0.689
Math VAM score	1,169	0.151	0.725
ELA VAM score	1,008	0.159	0.720
Field placement school characteristics			
Black students	877	0.181	0.260
Hispanic students	877	0.0664	0.0910
Asian students	877	0.0150	0.0230
Native students	877	0.00185	0.00257
White students	877	0.730	0.288
Students at or above proficiency on statewide tests	852	0.500	0.150
FRPL eligible	877	0.603	0.225
Student enrollment	876	614.7	361.6
Elementary grades	878	0.427	0.495
Middle grades	878	0.130	0.336
High school grades	878	0.183	0.386
Other grades	878	0.261	0.438
Teacher turnover rate	882	0.154	0.0913

(continued)

Table 1. (continued)

	<i>n</i>	Proportion/Mean	<i>SD</i>
Current school characteristics			
Black students	1,168	0.205	0.276
Hispanic students	1,168	0.0805	0.103
Asian students	1,168	0.0170	0.0263
Native students	1,168	0.00276	0.00353
White students	1,168	0.689	0.307
Students at or above proficiency on statewide tests	1,070	0.490	0.155
FRPL eligible	1,168	0.536	0.226
Student enrollment	1,164	594.3	364.9
Elementary grades	1,171	0.368	0.444
Middle grades	1,171	0.144	0.349
High school grades	1,171	0.168	0.374
Other grades	1,171	0.224	0.397
Teacher turnover rate	1,155	0.160	0.0907

Note. The table shows teacher-by-year observations for an analytic sample. PSTs' VAMs were measured after becoming teachers of record. Both CTs' and PSTs' VAM scores are standardized. Other endorsement areas include preK–12 and K–12. CT = cooperating teacher; PST = preservice teacher; FRPL = free and reduced-priced lunch program; VAM = value-added to students' test scores; ELA = English language arts.

Turning to CT characteristics, most CTs (84%) were women and White (94%). On average, CTs had about 14 years of teaching experience and received high observation ratings (e.g., the mean overall rating was 4.1 on a scale of 1 to 5). Compared to other teachers in Tennessee public schools during the same period, those who served as CTs in our sample were more likely to be women and White and less likely to be Black or other races (see Appendix Table A3). However, CTs and non-CTs has similar levels of teaching experience. In terms of instructional effectiveness, CTs received significantly higher observation ratings (overall and domain-specific) and had higher valued-added scores in all content areas than non-CTs.

Measures of Instructional Effectiveness

We investigated two measures of CTs' instructional effectiveness: observation ratings (overall scores and domain-specific scores) and VAMs (overall, math, and ELA).

Observation ratings. 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 the new evaluation system, teachers are evaluated based on (a) student test score growth as measured by the Tennessee Value-Added Assessment System (TVAAS), (b) student achievement on another selected measure, 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; however, the Board approved alternative rubrics to a small number of districts (10%). The TEAM rubric includes four domains: instruction, environment, planning, and professionalism, with several indicators associated with each domain.

Administrators are required to observe multiple domains during a classroom visit, with the exception of the professionalism domain, which is evaluated only at the end of the year. Teachers were rated on a scale of 1 to 5 on each indicator within a domain, with 1 = *significantly below expectations*, 3 = *at expectations*, and 5 = *significantly above expectations*. Overall observation ratings used in our analyses are an average of teachers' ratings across the four domains. We also used averages of each domain as outcomes. For teachers from districts that do not use the TEAM rubric, we only observe an overall score that also ranges from 1 to 5. For CTs, we used observation ratings from the year in which they served as a CT; in alternative specifications, we used ratings from the prior year. For PSTs, we examined the observation ratings they received each year after being hired into a Tennessee public school.

VAMs. In addition to observation ratings, we also observed teachers' VAMs. The TDOE provided us with teacher by subject (exam), ³ grade, and year VAM estimates for each teacher assigned to a tested subject. We standardized these scores within year, grade, and subject (exam) using all teachers in the state so scores are scaled in teacher-level standard deviation units. For teachers who had multiple VAM scores in the same year, we then created a year-specific aggregate score by averaging their different standardized scores. We then created a composite VAM for each teacher in each year by averaging all of his or her VAM scores across subject areas and grade levels. Because it assumes comparability across subject areas, grade levels, and standardized tests, we acknowledge that this approach has limitations. However, it is similar to how Tennessee evaluates teachers who teach multiple subjects and grades. In case CTs' effects on PSTs' instructional effectiveness are subject-specific, we also created separate math and English language arts VAM scores.⁴

Analytic Method

To assess the relationship between CTs' and PSTs' instructional effectiveness, we used four-level multilevel models (MLMs) in which years were nested within PSTs, which were nested within schools and districts. We estimated PSTs' observation ratings (or VAMs) as a function of their CTs' instructional effectiveness. In all models, we controlled for characteristics of PSTs (gender, race, age, residency, undergraduate vs. graduate EPP, endorsement area, and alternative certification) and CTs (gender, race), the schools in which PSTs completed their field placement, and the schools in which they are currently employed (proportion students' race and receipt of free lunch, grade levels served, enrollment, average teacher turnover, proportion of students scoring proficient or above on state tests).

To account for the fact that over one-third of the PSTs in our sample were mentored by multiple CTs, we averaged characteristics and measures of instructional effectiveness across individuals. For dichotomous variables, we created indicators for having ever worked with a CT with that characteristics (e.g., ever had a White CT vs. never had a White CT). We averaged values across continuous variables, such as years of teaching experience, observation ratings, and VAMs. We took the same approach to characteristics of field placement schools. We treated CT and field placement school characteristics as time-invariant, PST-level predictors.

Across the measures of instructional effectiveness that we use as outcomes, we focus on four model specifications. Our first model specification is a four-level, MLM with observation-years at Level 1, PSTs at Level 2, employment schools at Level 3, and district at Level 4:

$$PST_{ist} = \beta_{0000} + \beta_1 CT_{ist} + \chi_{isd} + \gamma_{ist} + FPS_{isd} + CS_{ist}, \\ + \varepsilon_{ist} + \tau_{0isd} + u_{00sd} + r_{000d}$$

where PST_{ist} and CT_{ist} are measures of the instructional effectiveness of PSTs and CTs, respectively, for PST i , in year t , teaching in school s , and district d ; χ_{isd} represents a vector of PST characteristics; γ_{ist} represents CT characteristics; FPS_{isd} is a vector of field placement school controls; and CS_{ist} is a vector of current school controls. Our multilevel approach allows us to account for the nested structure of our data; we do so by including mutually independent random effects associated with time, ε_{ist} , teachers, τ_{0isd} , schools, u_{00sd} , and districts, r_{000d} . Across models, we report the coefficient β_1 on CT_{ist} , which represents the association between measures of PSTs' and CTs' effectiveness. Because our remaining three model specifications use an ordinary least squares (OLS) framework, we include Model 2—which is the same as Model 1 but uses OLS instead of MLM—in order to check whether this change results in similar estimates.

We find evidence from Model 1 and 2 results that there is a positive association between PST and CT instructional effectiveness. Because this is a correlational study, we cannot necessarily conclude that instructionally effective CTs are causing PSTs to be more instructionally effective; we must consider threats to a causal interpretation—especially different kinds of selection that are likely occurring. First, it is possible that more instructionally

promising PSTs sort into EPPs that tend to recruit more instructionally effective CTs. To test whether this might be the case, in Model 3, we reestimated Model 2 but also included EPP fixed effects. Second, it is possible that PSTs who are more likely to gain employment in schools in which teachers tend to get stronger average observation ratings also tend to be placed with more instructionally effective CTs. In other words, having an instructionally effective CT is not causing a PST's instruction to improve but is instead predicting the PST will end up being employed in a school where he or she is more likely to receive stronger observation ratings (e.g., because the school supports instructional improvement or the evaluator is more lenient). Thus, in Model 4, we add school fixed effects to Model 3 to adjust for this form of selection.⁵

Results

PSTs who were mentored by more effective CTs generally earned higher observation ratings and higher VAMs; we describe results related to these two outcomes separately. For each outcome, we assessed the extent to which PSTs' effectiveness is associated with their CTs' effectiveness as measured by overall and domain observation ratings, VAMs (overall, math, and ELA), and years of teaching experience.

Preservice Teachers' Observation Ratings

Table 2 displays results from models with PSTs' overall observation ratings as the outcome, while Table 3 shows results from models with domain scores as outcomes; each coefficient is from a separate model. Generally, CTs' observation ratings—both overall and domain-specific scores—were significant predictors of PSTs' future ratings. In terms of overall observation ratings, a 1 point increase in CTs' ratings was associated with between a 0.07 and 0.1 point increase⁶ in PSTs' overall ratings, equivalent to about half a year of initial teaching experience in our models, or about one-third of PSTs' first-year gain.⁷ Thus, compared with PSTs whose CTs had average ratings of 3.0, PSTs whose CTs had average ratings of 5.0 performed as though they had been teaching an additional year. When investigating CTs' domain scores, we found CTs' instruction and planning domains to positively and significantly predict PSTs' overall ratings across model specifications; CTs' environment and professionalism domains trended positive across models and were significant in Models 1 and 4.

Though CTs' observation ratings were associated with increases in PSTs' observation ratings, CTs' VAM scores (overall, math, and ELA) were not. Across models, estimates on CTs' years of teaching experience trended negative and were statistically significant, though still small in magnitude, in Models 1 and 4. These results provide some evidence that PSTs received worse observation ratings when their CTs had more years of experience.⁸ It is important here to remind readers that the average experience of CTs was almost 14 years. We are thus mostly comparing experienced teachers to other experienced teachers. It is possible that results might look different if we were comparing inexperienced with experienced CTs.

Table 2
Preservice Teachers' Observation Ratings as a Function of CT Instructional Effectiveness

	PSTs' Overall Observation Ratings			
	(1)	(2)	(3)	(4)
CT years of teaching experience	-0.0016* (0.00096)	-0.0013 (0.0012)	-0.0017 (0.0012)	-0.0035*** (0.0010)
CT Overall observation rating score	0.086*** (0.022)	0.098*** (0.025)	0.088*** (0.025)	0.071** (0.024)
CT instruction domain score	0.089*** (0.022)	0.091*** (0.025)	0.085*** (0.025)	0.072** (0.025)
CT environment domain score	0.045** (0.019)	0.034 (0.022)	0.028 (0.022)	0.052** (0.024)
CT planning domain score	0.046** (0.017)	0.047** (0.019)	0.043** (0.019)	0.045** (0.018)
CT professionalism domain score	0.042** (0.019)	0.032 (0.021)	0.031 (0.021)	0.049** (0.021)
CT overall VAM score	-0.0042 (0.013)	-0.0080 (0.015)	-0.0054 (0.015)	-0.0019 (0.015)
CT math VAM score	-0.027 (0.018)	-0.032 (0.021)	-0.033 (0.020)	-0.038 (0.027)
CT ELA VAM score	-0.017 (0.017)	-0.013 (0.020)	-0.016 (0.020)	0.012 (0.023)
PST characteristics	x	x	x	x
CT characteristics	x	x	x	x
FPS characteristics	x	x	x	x
CS characteristics	x	x	x	
EPP fixed effects			x	x
School fixed effects				x

Note. Standard errors shown in parentheses. Each row and column represents a separate regression. All models control for PST, CT, and FPS characteristics; Models 1 through 3 also control for CS characteristics. Model 1 is a four-level multilevel regression (time nested within PSTs, schools, and districts). Models 2 through 4 are ordinary least squares regressions in which standard errors are clustered at the PST level. Models 3 and 4 contain EPP fixed effects. Model 4 contains school fixed effects and omits controls for CS characteristics. CT = cooperating teacher; FPS = field placement school; PST = preservice teacher; EPP = education preparation programs; VAM = value-added to students' test scores; ELA = English language arts; CS = current school.

* $p < .05$. ** $p < .01$. *** $p < .001$.

If the instructional effectiveness of CTs is directly related to PSTs becoming more instructionally effective, then we would expect PSTs to be most instructionally effective in those instructional domains in which their CTs particularly excel. Thus, we cycled through each PST domain score (e.g., instruction) to test whether it was associated with CT scores in the same domain (e.g., instruction), other domains (e.g., environment), and overall. Table 3 summarizes results.

Are PSTs more instructionally effective in the same domains in which their CTs excel? In general, our results are mixed. Our most robust finding is that CTs' scores in the instruction domain positively predicted PSTs' scores in all domains. Consistent with expectations, CTs' scores in the planning domain positively predicted PSTs' scores in the planning domain; however, CTs' scores in planning also predicted PSTs' scores in the instruction domain. Contrary to expectations, CTs' environment domain scores were not significantly associated with PSTs' environment domain scores but were positively associated with PSTs' professionalism domain scores.

Across Table 3, CTs' VAM (overall, math, and ELA) scores are mostly unrelated to PSTs' domain scores, with one exception. In three out of four model specifications, CT math VAM scores

were negatively and significantly related to PSTs' scores in the instruction domain. CT math VAM also negatively predicted PSTs' scores in the environment and planning domains but only in models with school fixed effects. Finally, CTs' years of experience trend negatively across PSTs' domain scores; however, estimates are larger in magnitude and statistically significant only for PST scores in instruction and environment.

Preservice Teachers' VAMs

Just as working with a more instructionally effective CT was mostly associated with increased instructional effectiveness in terms of observation ratings, the same was generally true for PSTs' VAM scores as well. Table 4 displays results for PSTs' overall VAM scores, and Table 5 shows results for math and ELA. In Models 1 through 3 in Table 4, a 1 standard deviation increase in CTs' overall VAM scores was associated with an average increase of 5% to 6% of a standard deviation increase in PSTs' overall VAM scores, equivalent to the gains in VAM associated with about one-half of a year of initial teaching experience, or about one-third of the gain associated with the first year.⁹ However, estimates are smaller in magnitude

Table 3
Preservice Teachers' Observation Domain Scores as a Function of CT Instructional Effectiveness

	PSTs' Instruction Domain Scores				PSTs' Environment Domain Scores			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
CT years of teaching experience	-0.0030** (0.0010)	-0.0027** (0.0012)	-0.0027** (0.0013)	-0.0035** (0.0011)	-0.0024* (0.0013)	-0.0022 (0.0015)	-0.0025* (0.0015)	-0.0021 (0.0014)
CT overall observation rating score	0.090*** (0.024)	0.085** (0.028)	0.081** (0.029)	0.084** (0.027)	0.083** (0.030)	0.098** (0.035)	0.092** (0.036)	0.066* (0.036)
CT instruction domain score	0.093*** (0.023)	0.085** (0.027)	0.086** (0.028)	0.085** (0.027)	0.084** (0.029)	0.091** (0.034)	0.086** (0.035)	0.069* (0.036)
CT environment domain score	0.035* (0.020)	0.017 (0.023)	0.016 (0.023)	0.044* (0.026)	0.039 (0.026)	0.037 (0.030)	0.032 (0.030)	0.041 (0.032)
CT planning domain score	0.047** (0.018)	0.040* (0.021)	0.041* (0.021)	0.047** (0.020)	0.037 (0.023)	0.032 (0.025)	0.033 (0.025)	0.044* (0.026)
CT professionalism domain score	0.044** (0.020)	0.028 (0.022)	0.030 (0.022)	0.057** (0.023)	0.041* (0.025)	0.023 (0.027)	0.023 (0.028)	0.064** (0.029)
CT overall VAM score	0.00076 (0.014)	-0.0039 (0.017)	-0.0031 (0.017)	-0.0057 (0.018)	-0.017 (0.018)	-0.016 (0.019)	-0.018 (0.019)	-0.015 (0.022)
CT math VAM score	-0.027 (0.019)	-0.039* (0.021)	-0.035* (0.021)	-0.052* (0.030)	-0.021 (0.024)	-0.040 (0.026)	-0.040 (0.025)	-0.100** (0.037)
CT ELA VAM score	0.0069 (0.018)	0.020 (0.021)	0.018 (0.021)	0.050* (0.026)	-0.015 (0.024)	0.0068 (0.027)	0.0018 (0.027)	0.048 (0.035)
PST characteristics	x	x	x	x	x	x	x	x
CT characteristics	x	x	x	x	x	x	x	x
FPS characteristics	x	x	x	x	x	x	x	x
CS characteristics	x	x	x		x	x	x	
EPP fixed effects			x	x			x	x
School fixed effects				x				x

	PSTs' Planning Domain Scores				PSTs' Professionalism Domain Scores			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
CT years of teaching experience	-0.0016 (0.0012)	-0.0019 (0.0014)	-0.0020 (0.0014)	-0.0017 (0.0012)	-0.0018 (0.0013)	-0.0011 (0.0016)	-0.0010 (0.0015)	-0.0017 (0.0014)
CT overall observation rating score	0.077** (0.028)	0.068** (0.034)	0.056 (0.035)	0.083** (0.030)	0.090** (0.029)	0.10** (0.035)	0.10** (0.036)	0.064** (0.032)
CT instruction domain score	0.077** (0.026)	0.070** (0.031)	0.064** (0.032)	0.089** (0.029)	0.087** (0.028)	0.094** (0.034)	0.093** (0.035)	0.083** (0.032)
CT environment domain score	0.034 (0.024)	0.020 (0.028)	0.010 (0.028)	0.051* (0.028)	0.076** (0.025)	0.070** (0.028)	0.067** (0.028)	0.085** (0.030)
CT planning domain score	0.053** (0.021)	0.048* (0.026)	0.047* (0.026)	0.053** (0.023)	0.028 (0.022)	0.037 (0.028)	0.032 (0.029)	0.027 (0.023)
CT professionalism domain score	0.026 (0.023)	0.0048 (0.026)	0.0064 (0.027)	0.048* (0.026)	0.047* (0.024)	0.045 (0.028)	0.035 (0.028)	0.039 (0.028)
CT overall VAM score	0.00035 (0.016)	-0.010 (0.018)	-0.0092 (0.019)	0.0064 (0.021)	0.0050 (0.017)	-0.0068 (0.020)	-0.0084 (0.020)	-0.0051 (0.019)
CT math VAM score	-0.011 (0.022)	-0.015 (0.026)	-0.015 (0.026)	-0.055* (0.032)	0.0033 (0.024)	-0.024 (0.029)	-0.025 (0.028)	-0.022 (0.039)
CT ELA VAM score	-0.0027 (0.021)	0.0082 (0.025)	0.0033 (0.025)	0.033 (0.035)	-0.022 (0.023)	-0.036 (0.025)	-0.028 (0.026)	0.036 (0.031)
PST characteristics	x	x	x	x	x	x	x	x
CT characteristics	x	x	x	x	x	x	x	x
FPS characteristics	x	x	x	x	x	x	x	x
CS characteristics	x	x	x		x	x	x	
EPP fixed effects			x	x			x	x
School fixed effects				x				x

Note. Standard errors shown in parentheses. Each row and column represents a separate regression. All models control for PST, CT, and FPS characteristics; Models 1 through 3 also control for CS characteristics. Model 1 is a four-level multilevel regression (time nested within PSTs, schools, and districts). Models 2 through 4 are ordinary least squares regressions in which standard errors are clustered at the PST level. Models 3 and 4 contain EPP fixed effects. Model 4 contains school fixed effects and omits controls for CS characteristics. CT = cooperating teacher; FPS = field placement school; PST = preservice teacher; EPP = education preparation programs; VAM = value-added to students' test scores; ELA = English language arts; CS = current school.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 4
Preservice Teachers' VAMs as a Function of CT Instructional Effectiveness

	PSTs' VAMs			
	(1)	(2)	(3)	(4)
CT years of teaching experience	-0.00026 (0.0023)	0.00016 (0.0024)	0.00014 (0.0023)	0.0011 (0.0028)
CT overall observation rating score	0.035 (0.052)	0.025 (0.059)	0.012 (0.064)	0.014 (0.081)
CT instruction domain score	0.067 (0.051)	0.068 (0.061)	0.053 (0.062)	0.049 (0.080)
CT environment domain score	-0.027 (0.046)	-0.055 (0.054)	-0.060 (0.054)	-0.066 (0.071)
CT planning domain score	-0.018 (0.041)	-0.038 (0.050)	-0.040 (0.050)	-0.047 (0.066)
CT professionalism domain score	-0.014 (0.045)	-0.027 (0.050)	-0.041 (0.050)	0.0032 (0.063)
CT overall VAM score	0.052* (0.027)	0.059* (0.033)	0.062* (0.033)	0.025 (0.046)
CT math VAM score	-0.032 (0.038)	-0.041 (0.044)	-0.051 (0.045)	-0.10 (0.079)
CT ELA VAM score	0.067* (0.038)	0.075* (0.039)	0.066* (0.039)	0.055 (0.052)
PST characteristics	x	x	x	x
CT characteristics	x	x	x	x
FPS characteristics	x	x	x	x
CS characteristics	x	x	x	
EPP fixed effects			x	x
School fixed effects				x

Note. Standard errors shown in parentheses. Each row and column represents a separate regression. All models control for PST, CT, and FPS characteristics; Models 1 through 3 also control for CS characteristics. Model 1 is a four-level multilevel regression (time nested within PSTs, schools, and districts). Models 2 through 4 are ordinary least squares regressions in which standard errors are clustered at the PST level. Models 3 and 4 contain EPP fixed effects. Model (4) contains school fixed effects and omits controls for CS characteristics. CT = cooperating teacher; FPS = field placement school; PST = preservice teacher; EPP = education preparation programs; VAM = value-added to students' test scores; ELA = English language arts; CS = current school.

* $p < .05$.

and not statistically significant in models with school fixed effects (Model 4). CTs' VAM scores in ELA were also predictive of higher overall VAM for PSTs; a 1 standard deviation increase in CTs' scores was associated with 6% to 8% of a standard deviation increase, equivalent to about a year of early career experience, or the average gain associated with almost one-half of the initial year of experience. CTs' math VAM trended negative but was not significantly associated with PSTs' overall VAM scores. When we considered PSTs' VAM scores in math and ELA as outcomes separately (see Table 5), we found that CTs' higher overall VAM scores and CTs' ELA VAM scores were both significantly associated with increases in ELA but not math. The effect sizes for PSTs' ELA VAM scores were larger in magnitude than for PSTs' overall VAM scores, ranging from 0.1 to 0.17 standard deviation units, which is roughly equivalent to the difference in average VAM scores between a first- and second-year teacher. CTs' math VAM scores were unrelated to both PSTs' math and ELA VAM scores.¹⁰

For the most part, CTs' observation ratings were unrelated to PSTs' VAM scores. CTs with higher overall observation ratings had PSTs whose overall, math, and ELA VAM scores were statistically indistinguishable from other PSTs. The same was also

mostly true for CT domain scores; while a few estimates on CT domain scores were statistically significant, there was no clear pattern, and none were significant across model specifications. CTs' years of teaching experience were generally unrelated to PSTs' VAM scores as well.

Sensitivity Tests

We ran several other checks to test the sensitivity of our main findings. First, we re-ran our preferred models with a subsample of the data using only PSTs' first and second years of teaching. The fact that the results were quite similar suggests that the possible effect of working with a more effective CT continues beyond PSTs' first years of teaching. Additionally, we looked to see whether the positive effects associated with CT effectiveness were driven by PSTs who were hired into their field placement schools, thus giving them an advantage over their peers who changed schools. We found that including an indicator for being hired by the FPS did not alter our main findings.

Also, we replaced the observation ratings and VAMs of CTs in the year in which they served as CT with their scores from the

Table 5
Preservice Teachers' Math and ELA VAM Scores as a Function of CT Instructional Effectiveness

	PSTs' Math VAM Scores				PSTs' ELA VAM Scores			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
CT years of teaching experience	-0.0039 (0.0033)	-0.00059 (0.0042)	-0.00020 (0.0039)	-0.0045 (0.0050)	-0.00061 (0.0027)	-0.0023 (0.0026)	-0.0021 (0.0025)	-0.0064* (0.0036)
CT overall observation rating score	0.051 (0.070)	0.098 (0.082)	0.12 (0.085)	-0.085 (0.092)	0.099 (0.066)	0.087 (0.069)	0.053 (0.073)	0.070 (0.13)
CT instruction domain score	0.056 (0.071)	0.11 (0.082)	0.13 (0.084)	0.0074 (0.12)	0.11* (0.063)	0.12* (0.069)	0.086 (0.070)	0.0017 (0.12)
CT environment domain score	0.038 (0.065)	0.025 (0.076)	0.038 (0.077)	-0.086 (0.094)	-0.0030 (0.058)	-0.021 (0.066)	-0.057 (0.067)	-0.079 (0.12)
CT planning domain score	-0.071 (0.056)	-0.089 (0.065)	-0.072 (0.068)	-0.17** (0.085)	-0.044 (0.052)	-0.022 (0.055)	-0.044 (0.055)	-0.11 (0.085)
CT professionalism domain score	0.033 (0.062)	0.028 (0.069)	0.024 (0.072)	0.017 (0.098)	0.089 (0.057)	0.050 (0.059)	0.047 (0.059)	0.029 (0.095)
CT overall VAM score	0.039 (0.038)	0.031 (0.046)	0.017 (0.044)	0.048 (0.080)	0.10** (0.035)	0.12*** (0.036)	0.12*** (0.036)	0.16** (0.060)
CT math VAM score	-0.062 (0.043)	-0.073 (0.051)	-0.082 (0.052)	-0.12 (0.13)	0.032 (0.056)	0.0082 (0.066)	0.016 (0.064)	0.017 (0.12)
CT ELA VAM score	0.066 (0.053)	0.055 (0.053)	0.034 (0.052)	0.055 (0.096)	0.098** (0.043)	0.13** (0.052)	0.11** (0.053)	0.17** (0.077)
PST characteristics	x	x	x	x	x	x	x	x
CT characteristics	x	x	x	x	x	x	x	x
FPS characteristics	x	x	x	x	x	x	x	x
CS characteristics	x	x	x		x	x	x	
EPP fixed effects			x	x			x	x
School fixed effects				x				x

Note. Standard errors shown in parentheses. Each row and column represents a separate regression. All models control for PST, CT, and FPS characteristics; Models 1 through 3 also control for CS characteristics. Model 1 is a four-level multilevel regression (time nested within PSTs, schools, and districts). Models 2 through 4 are ordinary least squares regressions in which standard errors are clustered at the PST level. Models 3 and 4 contain EPP fixed effects. Model 4 contains school fixed effects and omits controls for CS characteristics. CT = cooperating teacher; FPS = field placement school; PST = preservice teacher; EPP = education preparation programs; VAM = value-added to students' test scores; ELA = English language arts; CS = current school.

* $p < .05$. ** $p < .01$. *** $p < .001$.

previous year. We did this in case working with a student teacher affected their scores. This strategy greatly reduced our sample because no prior evaluation data were available for the CTs who mentor our first cohort of PSTs; in some models, the sample was reduced by one-half to two-thirds. Another limitation of this approach is that it assumes CTs' current performance is equivalent to their performance in the prior year even though evidence suggests that performance, as measured by VAM scores, can vary from year to year (Loeb & Candelaria, 2012). Estimates on lagged CT observation ratings (lagged $B = 0.063$; original $B = 0.088$ ***; see Table 2, Model 3)¹¹ and lagged CT VAM scores (lagged $B = 0.03$; original $B = 0.062$ **; see Table 4, Model 3) still trended positive, though were somewhat smaller in magnitude. This reduction in effect size, though, appears to be explained by the reduction in sample.¹²

To test whether a potentially more stable estimate of CTs' performance yielded different results, we replaced the observation ratings and VAMs of CTs in the year in which they served as CT with their scores averaged across all years of data in which we observe their scores. We found that CTs' average observation ratings predicted PSTs' future ratings in ways similar to our

original results (average $B = 0.12$ ***; original $B = 0.09$ ***; see Table 2, Model 3). For VAM, the result still were positive but smaller in magnitude and no longer significant (average $B = 0.03$; original $B = 0.062$ **; see Table 4, Model 3). Taken together, these checks suggest that our main results for observation ratings are not the result of PSTs' contributions to their CTs' scores. Robustness checks for VAM were less conclusive since estimates were still positive but smaller in magnitude and nonsignificant.

For individuals with more than one CT, in our main specifications, we average CT observation ratings or VAM scores. As a sensitivity check, we reproduced our analyses using two alternative approaches: (1) We constrained the sample to only those PSTs who had one CT, and (2) we used the CT with the highest observation rating (or VAM score), assuming that PSTs are learning most from the CT who is most instructionally effective. One limitation of constraining the sample only to individuals with one CT is doing so reduced our analytic sample by about one-half. Even so, the estimates on CTs' observation ratings predicting PSTs' future ratings were similar (one CT $B = 0.11$ ***; original $B = 0.088$ ***; see Table 2, Model 3). For VAM, the results were still positive but smaller in magnitude and no longer significant (one CT $B = 0.033$; original

$B = 0.062^{**}$; see Table 4, Model 3). As an alternative check, we used the most instructionally effective CT instead of average CT scores. The strategy assumes that PSTs are likely learning most from their most instructionally effective mentor; it also has the advantage of retaining the original analytic sample. Estimates were very similar to original ones. For observation ratings, we found: highest CT $B = 0.098^{***}$; original $B = 0.088^{***}$; see Table 2, Model 3. For VAM, we found highest CT $B = 0.059^*$; original $B = 0.062^{**}$; see Table 4, Model 3.

Finally, to ensure that the subject matter-specific nature of teachers' value-added to student test scores at the secondary level did not influence our main results, we estimated our main models using a subsample of PSTs who were endorsed as elementary teachers. The results were quite similar to our main findings. We also tried reproducing the analyses with only the subsample of teachers endorsed in non-elementary areas. This greatly reduced our sample. Estimates were unstable across model specifications and standard errors often quite large, so we did not have confidence in these results.

Discussion and Implications

This study provides the first evidence, of which we are aware, that PSTs are more instructionally effective when they learn to teach with CTs who are more instructionally effective. More specifically, when their CTs received higher observational ratings, PSTs also received higher observational ratings during their first years of teaching. Likewise, when CTs had higher VAM scores, so too did the PSTs they mentored. Additionally, the magnitudes of the relationships appear to be practically meaningful. Compared to a PST whose CT received an observational rating of 3.0, for instance, one whose CT received an observational rating of 5.0 performed as though they had completed an additional year of teaching. Likewise, a one standard deviation increase in CTs' overall or ELA VAM was associated with the average difference between a first- and second-year teacher on ELA VAM.

Due to the correlational nature of this study, we advise caution in drawing causal conclusions. In particular, there are many forms of selection that could explain the relationships between CTs' and PSTs' instructional effectiveness that we observe, including: (a) more instructionally promising PSTs may sort into programs that have more instructionally effective CTs on average, (b) more instructionally promising PSTs may tend to sort to schools (for student teaching and then for employment) where faculty tend to receive better evaluation scores (e.g., because the schools provide better support or have more lenient evaluators), and (c) more instructionally promising PSTs may sort to, or be selected by, CTs who tend to be more instructionally effective. Because we have no good measures for how instructionally promising PSTs are prior to student teaching, we are unable to test these forms of selection directly. While our use of EPP and school fixed effects go a long way in addressing (a) and (b), respectively, we are unable to adequately investigate or adjust for (c). More research is needed, including studies that randomly assign PSTs within EPPs to more and less effective CTs.

To this end, the first author, in collaboration with Dan Goldhaber and other colleagues, has partnered with eight teacher education programs across three states on the Improving Student Teaching Initiative (ISTI). As part of ISTI, some partnering programs are

randomly assigning PSTs to CTs and field placement schools. Initial evidence from the pilot year with one large program suggests that PSTs assigned to a combination of more instructionally effective CTs and better functioning field placement schools report (on surveys) not only better instruction modeled by their CTs but also more and better quality feedback and coaching as well as more opportunities to learn specific teaching skills (Ronfeldt et al., 2018). These preliminary findings suggest that more instructionally effective CTs not only model better instruction but also provide better coaching to learning teachers and are thus not only consistent with a causal explanation but are indicative of possible causal mechanisms. In future work, we will test if PSTs are also more instructionally effective in their first year of teaching.

One pattern we observe in the present study is that observation ratings of PSTs tend to be associated with the observation ratings but not the VAM scores of their CTs. Likewise, PSTs' VAM scores tend to be associated with the VAM scores but not the observational ratings of their CTs. If the instructional effectiveness of CTs is actually causing PSTs to become more instructionally effective, then we might expect the pattern to hold regardless of how "instructional effectiveness" is measured. While these results may seem counter to a causal explanation, they might also be supportive of one. Assuming that (a) observational ratings and VAM scores measure different dimensions of teaching quality¹³ and (b) CTs cause PSTs to improve most in the dimensions of instructional quality in which they perform best, then we would expect observed relationships to be within, and not necessarily between, measures of instructional quality. If these are indeed causal relationships, then it would suggest that districts and states can improve instructional effectiveness by using evaluation data (observation ratings or VAMs) to identify mentors or coaches who are likely to model more effective instruction (at least on the dimensions of instructional quality that these measures capture).

Though many states and districts have established minimum qualifications to serve as CTs, prior empirical support for these policies is thin at best. In providing initial evidence for an association between PST and CT instructional effectiveness, this study offers the strongest empirical support to date for such policies. However, our findings also suggest that such policies should likely focus on direct measures of instructional effectiveness, like observational ratings and VAMs, rather than years of experience, which tends to be more common. To our knowledge, only two states (Tennessee, Florida) currently require that teachers demonstrate a minimum level of instructional effectiveness (on state evaluations) in order to serve as CTs; our findings suggest more states follow their lead.

This study also has implications for school and district leadership, who face many realities that likely discourage them from wanting teachers, especially those that are most instructionally effective, to serve as CTs. In particular, students, as well as their families, want to be able to learn from the best and most seasoned teachers rather than prospective teachers with little to no experience. This study, though, suggests that schools and districts can benefit from tapping their most instructionally effective teachers to serve as CTs since doing so promises to boost the instructional effectiveness of the new teacher supply. Underscoring this point, we find that 45% of the PSTs who are subsequently employed in Tennessee are hired into the district in which they completed their student teaching

experiences.¹⁴ Moreover, nearly one in five PSTs is hired by the school where he or she student-taught, and this group of PSTs significantly outperforms other new hires. Allowing the most instructionally effective teachers to mentor PSTs may benefit schools not only by improving the instructional quality of potential hires but also by effectively providing a semester-long (or longer) and authentic job application period during which leaders can identify and recruit the most promising candidates.

NOTES

The authors would like to acknowledge the U.S. Department of Education, Institute of Education Sciences (IES) for their support of this work through the Statewide, Longitudinal Data Systems Grant (PR/Award R372A150015) and through predoctoral training support for Stacey Brockman (PR/Award R305B150012). The authors also thank the Tennessee Department of Education and participating Educator Preparation Programs for their support of this work and their ongoing commitment to improving teacher education in Tennessee. Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of our funders, collaborators, or those we acknowledge here.

¹A list of acronyms used throughout the paper is provided in Appendix Table A1.

²Given that alternative certification programs do not generally include traditional student teaching experiences under the mentorship of a cooperating teacher (CT), it is perhaps unsurprising that few pre-service teachers (PSTs) in this study were alternatively certified.

³Value-added to students' test scores (VAM) estimates were taken from three different standardized tests: SAT10, TCAP, and end-of-course exams.

⁴Subjects that comprised the math VAM scores include math, Integrated Math I, Integrated Math II, Integrated Math III, Algebra I, and Algebra II. Subjects that comprised the English language arts VAM scores include reading, language, English I, English II, and English III.

⁵As another check, we also calculated the propensities for PSTs to become employed in schools with higher average observation ratings and then adjusted for these propensities when estimating the effects of CT instructional effectiveness on PST instructional effectiveness. Specifically, for each school, we calculated the average observation ratings of teachers in that school; based on this, we constructed a binary measure for whether (=1) or not (=0) a school, on average, had teachers with observation ratings of 4 (on a 1–5 scale) or greater (high-rating school). We then used the characteristics of PSTs and their field placement schools to estimate the propensities that PSTs gain employment in high-rating schools. With the inverse of these propensities as probability weights, we used weighted least squares regression to estimate PSTs' observation ratings (or VAMs) as a function of their CTs' observation ratings (or VAMs) and current school covariates. Results were similar, suggesting that this kind of sorting is unlikely to explain observed relationships between CT and PST instructional effectiveness.

⁶Since the standard deviation on PST observation ratings is 0.49, these results equate to between one-fourth and one-fifth of a standard deviation unit.

⁷The coefficient for PSTs' years of teaching experience ranged from 0.16 to 0.18 across model specifications. The coefficient for PSTs' first year of teaching ranged from 0.24 to 0.27.

⁸Given that results for years of experiences were somewhat unexpected, we tried some alternative specifications. First, we entered CT experience on its own (without other CT characteristics). Next, we estimated a model that included a binary measure for CTs in the top three-quarters of the distribution in term of experience (equal to 1 if a CT had at least 7 years of teaching experience). Results were consistent with the original findings.

⁹Coefficients on PST experience in our model suggests that each year of early-career ("initial") teaching experience is associated with an increase in VAM between 0.05 and 0.1 standard deviation units; this estimate is based on the sample of PSTs during their first five years of teaching. Specifically, in the first year, PSTs in our sample gained between 0.15 and 0.22 standard deviation units.

¹⁰We reproduced results for only the subsample of PSTs who were receiving an elementary endorsement, and the results were very similar. We also tried reproducing the analyses with only the subsample of teachers endorsed in non-elementary areas. This greatly reduced our sample; for example, Model 3 estimates on PST math, VAM scores typically used 150 to 500 observations as compared with 900 to 1,200 observations from original models. Estimates were also quite unstable across model specification, so we did not have confidence in reporting these results.

¹¹We focus on our estimates from Model 3 because these models adjust for all covariates and EPP fixed effects. In some of our models with PST VAM as the outcome, our sample were small, and there were only one to two teachers in each current school. In these cases, standard errors were large, and some of the assumptions for using school fixed effects were likely not being met. Thus, for the specification checks in this section, we focus on Model 3.

¹²Our original models included $n = 3,169$ teacher-year observations, whereas our models using lagged CT VAM included $n = 905$ teacher-year observations. When focusing on only the subsample of CTs with lagged VAM scores, estimates on current and lagged VAM were almost identical, suggesting results are similar when using lagged or current VAM scores. This suggests that the subsample with lagged CT VAM scores may not be representative of our analytic sample.

¹³While prior literature has found observational ratings and VAMs to be significantly associated, correlations tend to be small in magnitude.

¹⁴Out of the 2,869 PSTs in the sample, 1,306 (45%) are initially hired into the same district in which they completed their student teaching. Out of 2,896 PSTs, there are 513 (17%) who are first hired by their field placement school.

REFERENCES

- Clarke, A., Triggs, V., & Nielsen, W. (2014). Cooperating teacher participation in teacher education: A review of the literature. *Review of Educational Research, 84*(2), 163–202.
- Giebelhaus, C. R., & Bowman, C. L. (2002). Teaching mentors: Is it worth the effort? *The Journal of Educational Research, 95*(4), 246–254.
- Goldhaber, D., Krieg, J., & Theobald, R. (2014). Knocking on the door to the teaching profession? Modeling the entry of prospective teachers into the workforce. *Economics of Education Review, 43*, 106–124.
- Greenberg, J., Pomerance, L., & Walsh, K. (2011). *Student teaching in the United States*. Washington, DC: National Council on Teacher Quality.
- Grossman, P., Ronfeldt, M., & Cohen, J. J. (2012). The power of setting: The role of field experience in learning to teach. In K. R. Harris, S. Graham, T. Urdan, A. G. Bus, S. Major, & H. L. Swanson (Eds.), *American Psychological Association (APA) educational psychology handbook* (Vol. 3, pp. 311–334). Washington, DC: American Psychological Association.
- Kraft, M. A., Blazar, D., & Hogan, D. (in press). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*.
- Loeb, S., & Candelaria, C. A. (2012). *How stable are value-added estimates across years, subjects and student groups?* Retrieved from <http://carnegieknowledgegenetwork.org/briefs/value-added/value-added-stability/>
- Matsko, K. K., Ronfeldt, M., Greene, H., Reininger, M., & Brockman, S. (In press). The role of cooperating teachers in preparing pre-service teachers: A district-wide portrait. *Journal of Teacher Education*.

National Research Council. (2010). *Preparing teachers: Building evidence for sound policy*. Washington, DC: National Academies Press.

NCATE. (2010). *Transforming teacher education through clinical practice: A national strategy to prepare effective teachers*. Washington, DC: Author.

Ronfeldt, M., Goldhaber, D., Cowan, J., Bardelli, E., Johnson, J., Tien, C. D. (2018). *Identifying promising clinical placements using administrative data: Preliminary results from ISTI placement initiative pilot* (CALDER Working Paper No. 189). Retrieved from: <https://caldercenter.org/publications/identifying-promising-clinical-placements-using-administrative-data-preliminary-results>

Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2002). Teacher preparation research: An insider's view from the outside. *Journal of Teacher Education*, 53(3), 190–204.

SHANYCE L. CAMPBELL, PhD, is an assistant professor at the University of California, Irvine, School of Education, 3200 Education, Irvine, CA 92697; shanyce.campbell@uci.edu. Her research focuses on understanding how policies and practices influence access to quality learning opportunities for students marginalized by the educational system.

Manuscript received October 6, 2017
Revisions received March 7, 2018, and May 8, 2018
Accepted May 22, 2018

AUTHORS

MATTHEW RONFELDT, PhD, is associate professor of educational studies at the University of Michigan, Ann Arbor, School of Education 610 E. University Ave., Ann Arbor, MI 48109-1259; ronfeldt@umich.edu. His research focuses on improving teaching quality through research, practice, and policy, particularly in schools and districts serving marginalized students.

STACEY L. BROCKMAN, MA, is a doctoral candidate at the University of Michigan, Ann Arbor, School of Education 610 E. University Ave., Ann Arbor, MI 48109-1259; sbrock@umich.edu. Her research focuses on the effect of educational programs and practices on students' academic and social-emotional growth and the equality of student outcomes.

Appendix

Table A1
Acronyms and Descriptions

Acronym	Full Phrase
CS	Current school
CT	Cooperating teacher
EPP	Education preparation program
FPS	Field placement school
OR	Observation rating
PST	Preservice teacher
TDOE	Tennessee Department of Education
TEAM	Tennessee Educator Acceleration Model
TVAAS	Tennessee Value-Added Assessment System
VAM	Value-added to students' test scores

Table A2
Comparison of PST Analytic Sample to Full Sample of PSTs

	PSTs Linked to CTs		PSTs Not Linked to CTs		Difference
	n	Proportion/Mean	n	Proportion/Mean	
Women	2,839	0.791	13,993	0.760	0.031***
Black PSTs	2,839	0.0296	13,994	0.120	-0.090***
Other PSTs of color	2,839	0.0282	13,994	0.0589	-0.031***
White PSTs	2,839	0.942	13,994	0.821	0.12***
Age	2,551	29.29	10,873	32.51	-3.22***
Undergraduate degree	2,839	0.556	13,994	0.433	0.12***
Elementary endorsed	2,839	0.534	13,957	0.361	0.17***
Secondary endorsed	2,839	0.294	13,957	0.307	-0.013
Special education endorsed	2,839	0.0645	13,994	0.0804	-0.016**
Other endorsement	2,839	0.167	13,957	0.326	-0.16***
Alternative certification	2,838	0.00614	13,753	0.250	-0.24***
Out of state	2,838	0.0434	13,989	0.0799	-0.036***
Not hired year after completion	2,869	0.226	13,994	0.177	0.049***
Overall observation rating	2,766	3.609	13,159	3.686	-0.077***
Instruction domain score	2,492	3.512	10,630	3.589	-0.078***
Environment domain score	2,484	3.982	10,562	4.012	-0.030**
Planning domain score	2,480	3.655	10,320	3.682	-0.027**
Professionalism domain score	2,398	3.889	10,417	3.952	-0.064***
Overall VAM score	1,632	-0.125	6,887	-0.0693	-0.056**
Math VAM score	906	-0.178	3,644	-0.200	0.022
ELA VAM score	1,030	-0.127	3,819	-0.139	0.012

Note. Table summarizes teacher-by-year statistics for graduates of TN TEPs who are mentored by a CT in our analytic sample, TN TEP graduates who were not in our analytic sample, and the difference between the analytic and full sample. Differences are the result of regressions of each teacher characteristic as a function of a time invariant indicator for being in the analytic sample. CT = cooperating teacher; PST = preservice teacher; VAM = value-added to students' test scores; ELA = English language arts; TN TEP = Tennessee teacher education program.

** $p < .01$. *** $p < .001$.

Table A3
Comparison of CTs to Non-CTs

	Cooperating Teachers		All other Teachers		Difference
	<i>n</i>	Proportion/Mean	<i>n</i>	Proportion/Mean	
Women	4,353	0.844	557,774	0.790	0.054***
Black teachers	4,355	0.0393	564,731	0.124	-0.085***
White teachers	4,355	0.945	564,731	0.829	0.12***
Other teachers of color	4,459	0.0386	583,848	0.0708	-0.032***
Age	4,305	42.99	480,285	43.31	-0.32*
Years of teaching experience	4,358	13.94	506,862	13.06	0.88***
Observation rating	3,057	4.057	325,543	3.903	0.15***
Instruction domain score	2,604	3.953	256,342	3.810	0.14***
Environment domain score	2,580	4.390	251,489	4.209	0.18***
Planning domain score	2,584	4.003	234,981	3.824	0.18***
Professionalism domain score	2,516	4.306	252,675	4.113	0.19***
Overall VAM score	2,075	0.217	139,711	0.0165	0.18***
Math VAM score	1,094	0.219	66,801	0.00597	0.17***
ELA VAM score	1,249	0.183	74,797	0.0279	0.15***

Note. Table summarizes teacher-by-year statistics for teachers who serve as CTs during the year in which they mentor a PST, all non-CT teachers, and the difference between CTs and non-CTs. Differences are the result of regressions of each teacher characteristics as a function of a time-varying indicator for “CT versus non-CT.” CT = cooperating teacher; PST = preservice teacher; VAM = value-added to students’ test scores; ELA = English language arts;
p* < .05. **p* < .001.

Table A4
Comparison of Main Results With and Without Current School Average Proficiency

	PSTs’ Overall Observation Ratings					
	(1)	(2)	(3)	(4)	(5)	(6)
CT years of teaching experience	-0.0015 (0.00096)	-0.0014 (0.0010)	-0.0013 (0.0010)	-0.0015 (0.0011)	-0.0011 (0.0012)	-0.0011 (0.0012)
CT overall observation rating score	0.083*** (0.022)	0.082*** (0.023)	0.081*** (0.023)	0.088*** (0.025)	0.083** (0.027)	0.078** (0.027)
CT instruction domain score	0.085*** (0.022)	0.077*** (0.023)	0.075** (0.023)	0.084*** (0.025)	0.072** (0.027)	0.069** (0.027)
CT environment domain score	0.044** (0.019)	0.043** (0.020)	0.041** (0.021)	0.030 (0.021)	0.027 (0.024)	0.025 (0.024)
CT planning domain score	0.045** (0.017)	0.043** (0.018)	0.043** (0.018)	0.041** (0.019)	0.045** (0.021)	0.043** (0.021)
CT professionalism domain score	0.041** (0.019)	0.031 (0.020)	0.031 (0.020)	0.030 (0.021)	0.026 (0.022)	0.025 (0.022)
CT overall VAM score	-0.00098 (0.013)	-0.0063 (0.013)	-0.0070 (0.013)	-0.0038 (0.015)	-0.0100 (0.016)	-0.011 (0.016)
CT math VAM score	-0.027 (0.018)	-0.023 (0.019)	-0.023 (0.019)	-0.032 (0.020)	-0.036* (0.021)	-0.036* (0.021)
CT ELA VAM score	-0.0098 (0.017)	-0.018 (0.018)	-0.020 (0.018)	-0.015 (0.019)	-0.022 (0.022)	-0.025 (0.022)
PST characteristics	x	x	x	x	x	x
CT characteristics	x	x	x	x	x	x
FPS characteristics	x	x	x	x	x	x
Reduced set of CS characteristics	x	x		x	x	
Full set of CS characteristics			x			x
EPP fixed effects				x	x	x

Note. Standard errors shown in parentheses. Each row and column represents a separate regression. All models control for PST, CT, and FPS characteristics. Models 1 and 4 reproduce results shown in Table 2 (Columns 1 and 3). Models 2 and 5 restrict the sample to observations for which information about CS average proficiency is available. Models 3 and 6 add a control for CS average proficiency. Models 1 through 3 are four-level multilevel regression (time nested within PSTs, schools, and districts). Models 4 through 6 are ordinary least squares regressions in which standard errors are clustered at the PST level. CT = cooperating teacher; FPS = field placement school; PST = preservice teacher; EPP = education preparation programs; VAM = value-added to students’ test scores; ELA = English language arts; CS = current school.
p* < .05. *p* < .01. ****p* < .001.

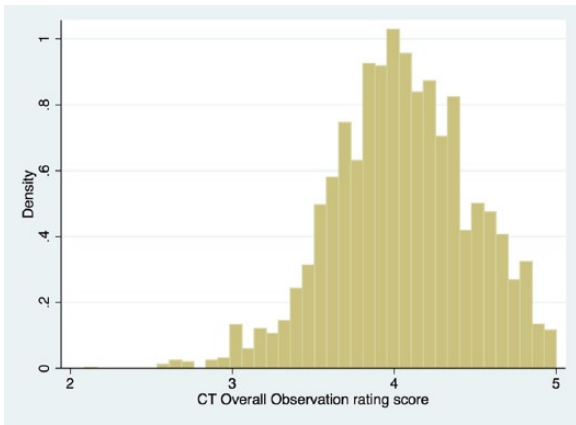


FIGURE A1. *Distribution of cooperating teacher overall observation ratings.*

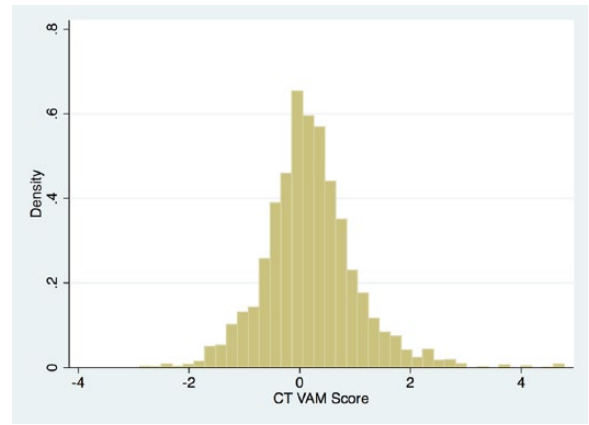


FIGURE A2. *Distribution of cooperating teacher overall value-added to students' test scores.*