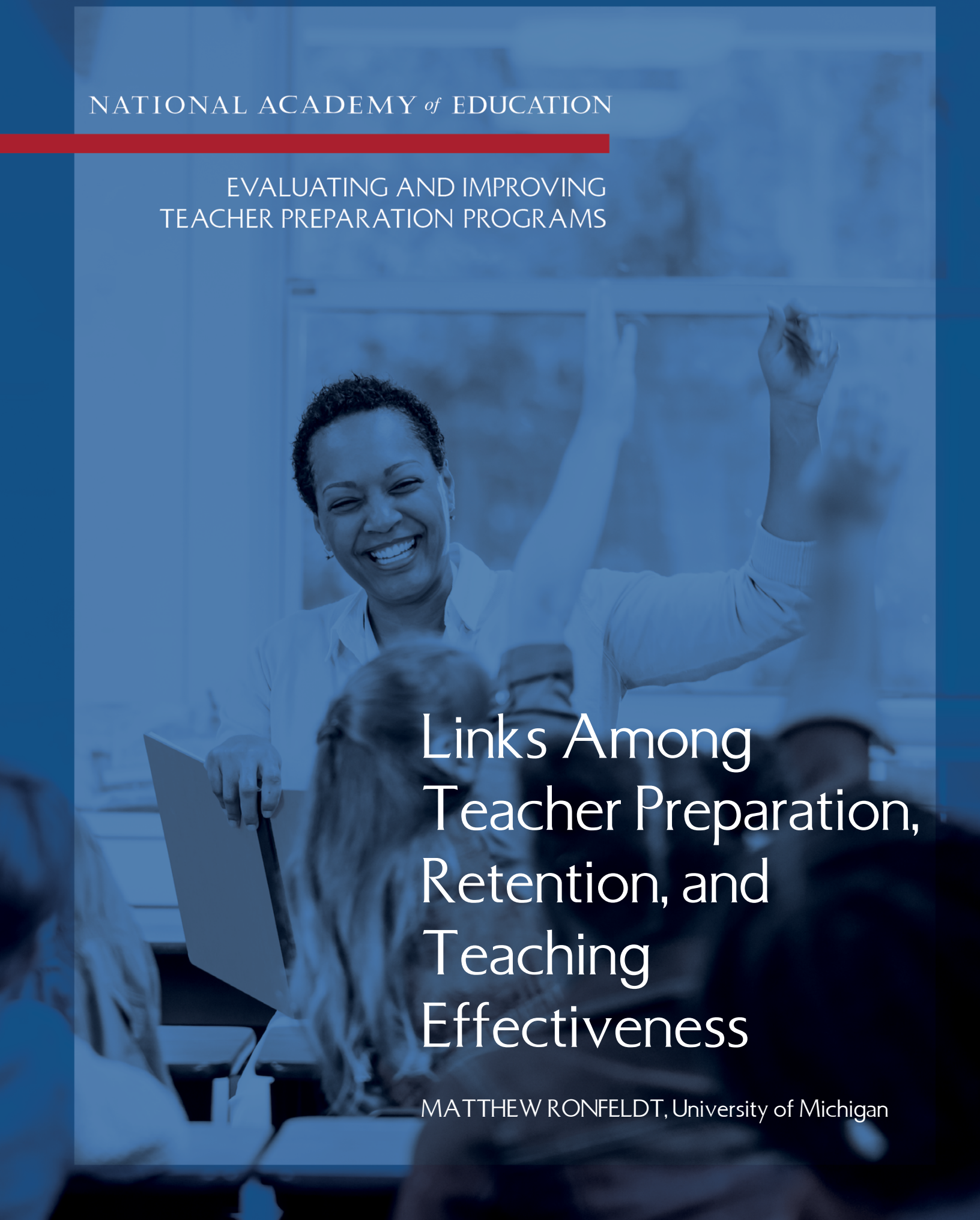
NATIONAL ACADEMY of EDUCATION

EVALUATING AND IMPROVING
TEACHER PREPARATION PROGRAMS



Links Among Teacher Preparation, Retention, and Teaching Effectiveness

MATTHEW RONFELDT, University of Michigan

NATIONAL ACADEMY OF EDUCATION 500 Fifth Street, NW Washington, DC 20001

NOTICE: The project and research are supported by funding from the Bill & Melinda Gates Foundation. This paper was prepared for the National Academy of Education (NAEd) to inform and support the work of the steering committee for *Evaluating and Improving Teacher Preparation Programs*, including the consensus report. The opinions expressed are those of the author and not necessarily those of the NAEd or the Bill & Melinda Gates Foundation.

Digital Object Identifier: 10.31094/2021/3/1

Suggested citation: Ronfeldt, M. (2021). *Links among teacher preparation, retention, and teaching effectiveness*. National Academy of Education Committee on Evaluating and Improving Teacher Preparation Programs. National Academy of Education.

National Academy of Education

Evaluating and Improving Teacher Preparation Programs

Links Among Teacher Preparation, Retention, and Teaching Effectiveness[1]

Matthew Ronfeldt, *University of Michigan*

September 2021

CONTENTS

[1] The author would like to acknowledge extremely helpful feedback on earlier drafts of this paper from Emanuele Bardelli, Kevin Bastian, Eric Brown, Bob Floden, Dan Goldhaber, Kavita Kapadia Matsko, Matthew Truwit, and Peter Youngs. All opinions, perspectives, and errors represented in this paper should be attributed to the author.

INTRODUCTION

With a charge from the U.S. Department of Education, Wilson et al. (2002) put together a technical working group of distinguished scholars to examine the existing research base to identify which claims about teacher education were supported by research. In concluding that the research base on teacher education is "thin," they pointed out that most studies focused on (prospective) teachers' attitudes and beliefs as outcomes rather than teaching practice or retention and were qualitative case studies, often self-studies, of single programs/courses, thus raising concerns over generalizability and objectivity. In summary, the authors made a global observation about methodological trends in teacher education research:

> A decade or two ago, naturalistic or interpretivist inquiry was too seldom found in journals. Its growth has contributed many insights into education, schooling, and teacher preparation. It seems, however, that the pendulum might have swung too far. We found that most scholarship was limited to small-scale interpretivist research. (p. 202)

A number of subsequent initiatives have led to reviews of the teacher education literature reaching similar conclusions. These initiatives included the American Educational Research Association Panel on Research and Teacher Education chaired by Marilyn Cochran-Smith and Ken Zeichner (2009), which determined that there was not enough rigorous research connecting features of teacher education with changes in teacher candidates' practices and skills. Later, Congress commissioned the National Research Council (NRC) to carry out a consensus panel study with a focus on identifying dimensions of teacher preparation related to student learning/achievement. Again, their sobering conclusion was that we could still say little about which preparation features were related to better student outcomes (NRC, 2010).

Since the NRC (2010) report, a substantial number of large-scale, quantitative studies on preparation features associated with teaching effectiveness and retention have emerged. The purpose of this paper is to review this body of evidence and make sense of what it suggests for practitioners and policymakers. To my knowledge, this is the first review focused specifically on large-scale, quantitative studies of teacher education, thus providing a unique opportunity to consider what we can learn from this body of literature.

That said, it is important to acknowledge that the large-scale, quantitative studies reviewed in this paper typically are grounded in and draw motivation from prior qualitative research that has predominated the field of teacher education for decades. Thus, at the beginning of each section, I introduce my review of the quantitative literature by grounding it in foundational qualitative scholarship. By doing so, I intend to acknowledge the debt that quantitative literature has to prior qualitative literature and, more broadly, to build a case for methodological pluralism (Moss & Haertel, 2016) in teacher education research.

To be considered for this paper, studies had to include measures of specific features of teacher education as predictor/independent variables and measures of retention or teaching effectiveness as outcome/dependent variables. Studies typically include many programs and candidates to ensure enough variation in preparation features and outcomes to measure associations among them. However, studies of single programs or

courses were also considered if their designs allowed for comparisons between groups of candidates that experienced different preparation features; for example, experiments are included where candidates were randomly assigned to receive different forms of preparation.

Regarding measures of teaching effectiveness, many studies included in this paper focus on "value-added" to student achievement measures (VAMs) and observation ratings (of both preservice and inservice teachers) by outside evaluators based on rubrics, often aligned to district or state evaluation systems. I also include studies using self-reported (survey) measures of preparedness to teach and planned persistence as there is some evidence that these self-reported measures predict observed teaching effectiveness and retention (Bastian et al., 2019; Ronfeldt et al., 2014),[2] and they are commonly used to evaluate teacher education program quality. Analyses based on self-reported outcomes also typically preceded, and set the table for, studies using observed teaching effectiveness or retention and so they provide important context.

Studies included in this review thus vary in how they conceptualize, measure, and study teaching effectiveness. This variability may be an asset, as something as multidimensional as teaching effectiveness likely requires a variety of measures to capture its complexity (Gitomer & Bell, 2013); this variability also reflects the fact that people differ in what they value as educational outcomes (Fenstermacher & Richardson, 2005; Labaree, 1997). At the same time, each of the measures used by studies included in this review has limitations and fails to fully capture the complexity of teaching effectiveness. VAMs have been critiqued for being available only to teachers in tested grades and subjects, based on tests that fail to capture cognitively demanding learning and teaching, related to student and other characteristics beyond the instructional quality that they are meant to signal, and unstable across time (Grossman et al., 2014; Hill et al., 2011; Newton et al., 2010; Rothstein, 2009). Likewise, teachers of color and teachers in classrooms with more lower-achieving students and students of color tend to receive lower observation ratings, but these lower ratings do not appear to be explained by actual differences in teaching effectiveness (Campbell, 2020; Campbell & Ronfeldt, 2018; Grissom & Bartanen, forthcoming; Jiang & Sporte, 2016; Steinberg & Garrett, 2016). Survey-based measures are self-reported, and are thus prone to distortions due to memory limitations and psychological biases (Dunning et al., 2003; Matsko et al., 2020; Ronfeldt et al., 2020b). Despite the limitations of these various measures and disagreements about what they actually capture or signal, in this paper I include observation ratings, VAMs, and self-reported measures of feelings of readiness to teach as measures of "teaching effectiveness." In the concluding section, I return to these limitations and interrogate findings and implications in light of them.

This paper begins by describing the groundbreaking work by the New York City (NYC) Teacher Pathways Project, which identified clinical experiences—including student teaching and pre-student teaching experiences—and its alignment with other aspects of programs generally, as predictive of teaching effectiveness. It then reviews literature on features of clinical experiences, including their duration, the features of the field placement schools in which they occur, and cooperating teacher characteris-

[2] Ronfeldt et al. (2020b), however, found no relationship between recent graduates' self-perceptions of preparedness and first-year observation ratings.

tics. Finally, it focuses on coursework, beginning with new evidence on practice-based course simulations and then the amount of coursework generally. Within each section, I summarize research focused on teaching effectiveness measures followed by retention.

THE CRITICAL ROLE OF CLINICAL EXPERIENCES

Prior to the publication of the 2010 NRC report, only a handful of studies had taken a "birds-eye," labor market perspective on the preparation of teachers by linking features of preparation that vary within and across many programs to graduate workforce outcomes in the schools served by those programs. Perhaps the most ambitious and highest-profile one was the NYC Teacher Pathways Project. One of their earliest analyses on specific *features* of preparation[3] focused on alignment between clinical experiences and other program dimensions including coursework, which prior qualitative literature had suggested to be important (Buchmann & Floden, 1993; Hammerness, 2006). Grossman et al. (2008) found that candidates enrolled in programs that included more structural features meant to promote alignment between clinical experiences and other program dimensions (e.g., more courses with clinical experiences attached, more hours of clinical experiences linked to methods courses, and program leaders selecting cooperating teachers) reported having significantly more aligned and coherent program experiences.

Soon after, the NYC Pathways team published a seminal piece linking extensive program features—based on survey and program review information—to graduates' early-career VAMs (Boyd et al., 2009). They found that graduates from programs scoring higher on their prior study's measure for program–clinical alignment—renamed "oversight of student teaching" measure—had stronger early-career VAMs. The authors also found that graduates had better VAMs when they had capstone experiences (typically clinical-based portfolios), more opportunities for practice-focused coursework (e.g., planning guided reading lessons), more opportunities to study NYC curricular materials, and more congruence between field and inservice placements (grade/subject/training). All of these features converged in suggesting that how programs emphasized and integrated clinical experiences in their program designs and curricula predicted graduates' teaching effectiveness. This led researchers, including me, to investigate whether specific dimensions of clinical experiences were associated with better instructional effectiveness among graduates.

This section reviews the literature examining variation in candidates' clinical experiences, including the quality/kinds of clinical experiences, and specifically field placement school and cooperating teacher characteristics, beginning with an analysis of one of the more fundamental and oft-debated questions: Is having longer clinical experiences better?

Duration of Clinical Experiences

Many assume that having more extensive clinical experiences is better. This assumption is reflected in trends during the 1980s toward requiring early clinical experiences

[3] Their earlier studies mostly compared average workforce outcomes between graduates from alternative versus traditional routes, masking much within- and between-program variation in preparation features (Grossman & Loeb, 2021).

prior to student teaching in the majority of states setting minimum requirements for student teaching duration (Greenberg et al., 2011), and, more recently, in the growing popularity of the residency model. Yet, as John Dewey (1904) warned, experience can be miseducative. Without the proper theoretical training, Dewey argued, clinical experiences can result in unproductive habits among learning teachers. Moreover, multiple prior literature reviews suggest that we do not know enough about the effects of extending student teaching on teaching effectiveness or retention (Clift & Brady, 2005; Grossman et al., 2011; McIntyre et al., 1996).

So, does the more recent, large-scale quantitative literature suggest that extending clinical experiences is better? The results, elaborated below, suggest that the answer depends on the outcome being considered.

Teaching effectiveness. Four large-scale studies have examined the relationship between duration of student teaching and self-reported measures of preparedness to teach, with all four finding positive associations. California State University (2002) found that a higher proportion (79 percent) of teachers who completed student teaching felt "adequately to well-prepared" to teach reading and language arts as compared to emergency-certified and intern teachers (67 percent); however, the latter group differs from the former in many other ways (e.g., having less coursework) that could explain these results. Rather than compare graduates who completed student teaching of any amount to none, a distinction likely associated with the preparation route (traditional, alternative), my collaborator and I considered only traditional route programs and only student teachers in Chicago (Ronfeldt & Reininger, 2012). We found that completing more weeks of student teaching was positively associated with feeling better prepared to teach, but this association was statistically significant in only one out of four model specifications. Examining whether this association observed in Chicago generalized across the United States, my collaborators and I found that recent graduates, especially those who took fewer methods courses, felt better prepared to teach early in their careers when they had completed more weeks of practice teaching (Ronfeldt et al., 2014).

Though candidates who complete longer student teaching tend to feel better prepared, are they also more effective teachers? To test this, my colleagues and I surveyed all student teachers and cooperating teachers in the Chicago area (across alternative, residency, and traditional programs) about preparation experiences and linked this information to three outcome measures: (1) candidates' self-reported preparedness to teach, (2) cooperating teachers' assessment of their candidates' preparedness to teach, and (3) among those graduates hired in the Chicago area, first-year observation ratings based on the district evaluation rubric (Ronfeldt et al., 2020b). Consistent with studies described above, we found that candidates who completed more hours of student teaching felt better prepared to teach; however, they received no better or worse first-year observation ratings and their cooperating teachers felt they were no more or less prepared to teach.

The latter results indicate that completing more student teaching might be associated with candidates feeling better prepared but not necessarily with being better teachers. Three other studies focusing on both mathematics and English language arts (ELA)

VAMs as outcomes reached similar conclusions. Boyd et al. (2009) studied all NYC programs (alternative and traditional) and compared graduates who had any amount of student teaching versus no student teaching; a limitation, though, is that student teaching is much less common among alternative than traditional programs, so it is possible that student teaching duration proxied for route of entry. I tried to improve on the blunt measure (none versus any amount) for student teaching duration by using candidate-level survey information to estimate the number of hours of student teaching across all types of programs supplying teachers to a large, southeastern metropolitan district (Ronfeldt, 2015). Preston (2017) looked across all middle grade certification programs in North Carolina and used a program-level measure for weeks of student teaching. Regarding the relationship between student teaching duration and mathematics VAMs, two studies had mixed but mostly negative results across specifications (Boyd et al., 2009; Preston, 2017) and the third found no relationship (Ronfeldt, 2015). There were no significant relationships between student teaching duration and ELA VAMs across models and studies except for a significant, negative relationship in one of Preston's (2017) model specifications.

Retention. The four studies linking student teaching duration to planned or observed persistence in teaching have found consistently positive results. Ronfeldt and Reininger (2012), introduced above, found that at graduation, candidates who reported completing more weeks of student teaching planned longer teaching careers generally and in Chicago specifically, though results were significant in only one of four model specifications. Going beyond self-reported planned persistence, three studies using nationally representative data considered the relationship between student teaching duration and observed teacher retention. Based on the Baccalaureate and Beyond Longitudinal Study of K-12 teachers (1993-1997), Henke et al. (2000) found that—after 5 years in teaching—twice as many study participants who had never student taught left teaching (29.3 percent) as compared with participants who completed any amount of student teaching (15.3 percent). This study had a number of limitations, though, including that graduates with some versus no student teaching likely differed in many other ways (e.g., route of entry, amount of coursework preparation) and, relatedly, it did not adjust for other characteristics of the programs, graduates, and schools.

Two more recent studies using the nationally representative Schools and Staffing Surveys—Teacher Follow-Up Survey (SASS-TFS) addressed these limitations by comparing graduates with different amounts of student teaching and by adjusting regression models for extensive preparation, teacher, and school characteristics (Ingersoll et al., 2014; Ronfeldt et al., 2014). Using the 2003-2004 SASS administration, Ingersoll et al. (2014) found that graduates who completed at least a semester of practice teaching were significantly more likely to remain in teaching than graduates who completed less than a semester and that these relationships were significantly stronger for mathematics teachers. My collaborators and I used both the 2003-2004 and 2007-2008 SASS administrations, a continuous measure for weeks of student teaching, and somewhat different modeling approaches than Ingersoll et al. (Ronfeldt et al., 2014). Though results trended positive, we found no statistically significant correlation between increasing weeks of practice teaching and retention rates for the full sample. That said, in comparison to

completing no weeks of practice teaching, completing 8 to 11 weeks doubled the odds of remaining in teaching. Moreover, among graduates who completed few/no methods courses, weeks of practice teaching significantly predicted retention. Among teachers who completed no methods coursework, completing 1 additional week of practice teaching increased the odds of staying in teaching by 5 to 10 percent.

Summary. Results here provide suggestive evidence that candidates who complete longer durations of student teaching feel better prepared to teach and are more likely to persist in teaching, but are no more or less instructionally effective.

Kinds/Quality of Clinical Experiences

While quantity of clinical experiences seems to matter for some outcomes, what about its quality? This section considers the evidence from large-scale, quantitative research about the quality or kinds of clinical experiences that are related to instructional effectiveness and retention. It begins with a review of the literature on field placement school working conditions followed by a focus on field placement school student demographics. Finally, I consider the literature on cooperating teacher characteristics.

Field Placement School Working Conditions

Based on qualitative evidence, a number of scholars have argued that, in order to support student learning, schools must also be organized for teacher learning (Feiman-Nemser, 1983; Lightfoot, 1986; Little, 1982). Large-scale, quantitative studies have since provided extensive support for this argument, demonstrating that schools marked by positive working environments, shared commitments to teaching and learning among faculty, high-quality teacher collaboration and professional development (PD), relational trust, and supportive leadership demonstrate not only higher levels of student learning but also increased instructional effectiveness and retention among teachers (Allensworth et al., 2009; Bryk et al., 2010; Goddard et al., 2007; Kraft & Papay, 2014; Ronfeldt et al., 2015). Having the opportunity to engage with highly skilled colleagues also seems to matter. Instructional quality improves most when teachers are able to collaborate with and learn from more instructionally effective teacher peers (Jackson & Bruegmann, 2009; Loeb et al., 2012; Papay et al., 2020). Just as employing instructionally effective teachers without promoting collaboration is unlikely to promote teacher learning, so too is promoting collaboration without employing instructionally effective teachers; schools likely need both to function as organizations for teacher learning.

Given that schools with strong professional learning environments benefit inservice teachers, one would expect these same kinds of schools to have especially strong effects on prospective teachers when they are first learning to teach. For decades, the professional development school (PDS) literature has suggested this to be the case. Rooted in Dewey's ideas about the teaching laboratory, PDSs are P-12 schools that partner with universities and function as sites for preservice candidate learning alongside P-12 student and inservice teacher learning (Darling-Hammond, 1994; Stallings & Kowalski, 1990). In our review of the literature on student teaching prior to 2008, my collaborators and I noted that PDS studies "stand apart from other literature on preservice field

experience" in moving "away from case studies on teacher beliefs and toward large scale, comparative, and quantitative studies" (Grossman et al., 2008, p. 318). Compared to their peers, PDS-prepared teachers tended to have better self-reported outcomes (efficacy, beliefs), observation ratings, and retention. More recent studies have continued to find PDS-prepared candidates to outperform other candidates in terms of teaching effectiveness (Castle et al., 2008) and retention (Latham et al., 2015). However, study limitations constrain what we can conclude from the PDS literature. First, across studies, candidates self-selected into PDS settings and experienced preparation that differed from their peers in other ways beyond having PDS placements (e.g., number of clinical hours, coursework sequence) that could explain better outcomes among PDS-prepared teachers. Second, there is much variation in what constitutes a PDS, making it difficult to ascertain which PDS features matter.

Furthermore, the vast majority of candidates today are prepared in more typical field placement schools. Though these schools may not be PDSs per se, some make more promising contexts for candidate/teacher learning. Until relatively recently, though, no large-scale studies had linked field placement school characteristics to graduates' teaching effectiveness or retention in order to identify promising characteristics.

Teaching effectiveness. Drawing on the extensive NYC Teacher Pathways data, I linked early-career teachers to their field placement schools (Ronfeldt, 2012). Though I did not have direct measures of field placement school professional learning environments, based on prior literature, I constructed school-level "stay-ratio" measures of average prior teacher retention as a proxy. By correlating this measure to (student) teacher survey information, I demonstrated construct validity: teachers in schools with higher stay-ratios (less turnover) reported experiencing better staff collegiality and administrative quality and support, observing excellent teachers and role models more frequently, being observed more regularly, receiving more useful feedback, and having more opportunities to experiment with coursework strategies. Moreover, candidates placed in field placement schools with higher stay-ratios had significantly better mathematics VAMs after graduating and becoming early-career teachers.

In a follow-up study in a different, large metropolitan area, I found further evidence that stronger field placement school professional learning environments are associated with better teaching effectiveness (Ronfeldt, 2015). In addition to school stay-ratio, I used teacher survey information to construct school-level measures of teacher collaboration quality. I also constructed measures for school-level VAMs, hypothesizing, based on prior literature (Jackson & Bruegmann, 2009), that higher-VAM schools would provide prospective teachers more opportunities to learn from instructionally effective colleagues. Indeed, candidates who learned to teach in field placement schools with stronger school-level VAMs, better collaboration quality, and, to a lesser degree, higher stay-ratios had better mathematics VAMs as early-career teachers; none of these field placement school characteristics were related to ELA VAMs.

Contrary to prior evidence, in their study of six large teacher preparation programs in Washington, Goldhaber et al. (2016) found student teaching in schools with higher stay-ratios to be unrelated to graduates' VAMs. More recently, Bastian et al. (2020) examined the relationships between field placement school characteristics and graduates'

teaching effectiveness, as measured by VAMs and observation ratings by principals, among six large preparation providers in North Carolina. Like Goldhaber et al. (2016), Bastian et al. (2020) found field placement school stay-ratio to be unrelated to graduates' instructional effectiveness on either measure. Consistent with Ronfeldt (2015), Bastian et al. also reported that graduates who student taught in field placement schools with better school-level VAMs had better VAMs as early-career teachers, and graduates who student taught in field placement schools with better teacher collaboration received stronger observation ratings. Consistent with the claim that these schools function as professional learning environments, Bastian et al. also found that candidates with lower grade point averages benefited most from placements in schools with stronger collaboration and achievement gains.

Retention. To my knowledge, only three studies have examined the relationship between specific field placement school characteristics and retention (Goldhaber et al., 2016, 2020b; Ronfeldt, 2012). Two of these studies found that graduates who learned to teach in higher stay-ratio field placement schools stayed in teaching longer than their peers (Goldhaber et al., 2016; Ronfeldt, 2012), while the third found no relationship (Goldhaber et al., 2020b).

Summary. Taken together, large-scale, quantitative studies indicate that teacher candidates benefit from learning to teach in field placement schools with strong professional learning environments—schools with better quality teacher collaboration, histories of producing strong achievement gains and employing instructionally effective faculty, and higher rates of teacher retention. Student teaching in schools with these characteristics predicts better later teaching effectiveness, suggesting that these kinds of contexts function as organizations for professional learning where collaborating with instructionally effective teacher peers in supportive (teacher) learning environments builds instructional skills that candidates carry into their inservice years. Evidence suggests that these same contexts might also predict better retention, but more research is needed.

Field Placement Student Demographics

Much has been written about the importance of designing clinical experiences in schools with racially and socioeconomically diverse students. Literature on culturally relevant pedagogy demonstrates the importance of learning about and leveraging the rich cultural knowledge and practices specific to the communities with which teachers work (Ladson-Billings, 1995; Lee, 2007), so it seems intuitive to incorporate clinical experiences that require candidates to engage with and learn about students and families from different cultural and racial backgrounds. However, in our review of prior literature (Grossman et al., 2008), some studies showed that placing student teachers, who are typically White, in field placements with many Black and Brown students sometimes reinforced stereotypes and deficit views. However, clinical experiences in communities of color with deliberately designed pedagogical/curricular approaches to interrogating these experiences typically improved candidates' attitudes and beliefs

(Grossman et al., 2008). No studies examined whether improved attitudes translated into improved teaching effectiveness or retention.

Thus, what do we learn from large-scale, quantitative research about the relationships between the sociodemographic characteristics (including racial/ethnic, socioeconomic, and linguistic backgrounds) of student populations in field placement schools and graduates' teaching effectiveness and retention? Generally, the evidence is mixed, though there is some indication that the demographic match between field placement and inservice schools where graduates are eventually employed is positively associated with teaching effectiveness.

Teaching effectiveness. Three studies revealed no relationship between average sociodemographic characteristics of students in field placement schools and graduates' teaching effectiveness. In Chicago, Ronfeldt et al. (2013) found average student characteristics (race/ethnicity, income status, linguistic backgrounds) to be unrelated to candidates' feelings of preparedness to teach and self-efficacy. Likewise, two other studies considering graduates' VAMs or observation ratings found average field placement school student sociodemographic characteristics to be unrelated (Ronfeldt, 2012; Ronfeldt et al., 2020b).

Three other studies, though, established positive relationships between average sociodemographic characteristics and teaching effectiveness. Bastian et al. (2020) found that candidates placed in field placements with more students of color had significantly better early-career VAMs though no better or worse observation ratings; the proportion of lower-income students was unrelated to either outcome. Similarly, I found that graduates who student taught in schools with more Black students had better mathematics VAMs but similar reading VAMs; the proportion of students who were lower-income, limited English proficient, or receiving exceptional student education services was unrelated to VAMs (Ronfeldt, 2015). Finally, Goldhaber et al. (2017) found that graduates from field placements with more "under-represented" students, which they defined as racial minority and lower-income students, had better mathematics VAMs but only in some models. Looking deeper, the authors found positive effects to be driven by graduates who were subsequently employed in the three most racially diverse districts suggesting that learning to teach in settings with more under-represented students benefited graduates employed in similar settings. This finding is consistent with scholarship from various disciplinary perspectives indicating that learning to teach with students representing specific cultural and economic backgrounds has the potential to support the development of knowledge, skills, and human capital specific to serving these populations (Goldhaber et al., 2017; Haberman, 1995).

The latter point also suggests that researchers consider a match in student demographics between field placement and employment schools, which three studies have done. I used the absolute difference between employment and field placement student demographics and found little evidence for demographic match to be related to graduates' VAMs (Ronfeldt, 2015). Goldhaber et al. (2017), though, improved on my approach by using a more flexible proxy for demographic match, finding that graduates had better early-career mathematics VAMs when employed in schools with a closer match to their field placements in terms of the proportion of under-represented (lower-income, racial minority) students. This means, for instance, that among graduates who begin

their inservice teaching careers in schools that enroll 75 percent under-represented students, those who completed student teaching in schools also with 75 percent under-represented students would have significantly better VAMs than those who student taught in schools with 75 percent affluent White students. The authors argued that the positive match effects likely indicate that recent graduates developed human capital specific to the populations of students they taught as preservice student teachers.

If so, then we would expect classroom match effects to be even stronger than school match effects, given that the former describes the population of students with whom teachers are most directly engaged. This is precisely what Krieg et al. (2020) found. While the school-level match on student income status predicted better mathematics VAMs, classroom-level match did so more strongly. The authors also found that graduates who were employed in the same grade and school levels (e.g., elementary) as their field placement schools had significantly better VAMs than other teachers; the grade-level match finding is consistent with Henry et al. (2013) while both the grade- and school-level match findings are consistent with Boyd et al. (2009), described earlier. Finally, among elementary teachers, the authors also found that graduates had better mathematics VAMs when employed in the same schools where they student taught. This latter result is also consistent with Ronfeldt et al. (2020b), which found observation ratings were greater among graduates employed in their field placement schools; it is inconsistent, though, with Henry et al. (2013), which found teachers employed in their field placement schools to have similar VAMs as other teachers.

Retention. Four studies have linked field placement school characteristics to teacher retention (Goldhaber et al., 2016, 2020b; Ronfeldt, 2012; Ronfeldt et al., 2013). All found student sociodemographic characteristics in field placement schools to be unrelated to planned years in teaching, planned years in the same district, and early-career retention among graduates. Goldhaber et al. (2016) found the match on proportion of under-represented students between field placement and employment schools to be unrelated to retention, unrelated to retention. On the other hand, Goldhaber et al. (2020b) found the match between field placement and employment schools on both the proportion of under-represented students and school type/level to predict retention but grade and district matches to be unrelated.

Summary. The evidence for an association between average field placement school student sociodemographic characteristics and instructional effectiveness is mixed. Though a number of studies found no relationship, others found that teachers have better observation ratings or VAMs when they student taught in schools with more students of color and more lower-income students. There is evidence that recent graduates are more instructionally effective when the student demographics of their employment schools more closely match those of their field placement schools, as student teachers may develop population- or context-specific knowledge and skills (human capital) that they carry into their early careers. There is related evidence that getting employed in the same grade level, school level, or school as one's field placement school is also related to better outcomes. Regarding studies on retention, none found relationships with field placement school student sociodemographic characteristics, while the evi-

dence is mixed for the match between field placement and employment school student sociodemographic characteristics.

The above sections suggest that field placement schools can function as organizations for teacher learning. One way that organizations can support teacher learning is by having faculty directly mentor student teachers. The next section reviews literature on cooperating teachers who provide some of the most direct and consistent mentoring support during preparation.

Cooperating Teachers' Teaching and Coaching Effectiveness

During student teaching and residency experiences, cooperating teachers host candidates in their classrooms, structure opportunities to practice teach, and provide feedback. In their seminal review of the teacher socialization literature, Zeichner and Gore (1990) conclude that cooperating teachers specifically, and field placement schools generally, are powerful forces of socialization—often away from what they termed as the more "progressive"/student-centered forms of teaching promoted by teacher education programs and toward more "conservative"/teacher-centered approaches that typify U.S. schools. However, the vast majority of research on cooperating teachers reviewed by Zeichner and Gore, and in subsequent reviews, involve small-scale qualitative case studies focused on the beliefs and attitudes among candidates, making it difficult to assess the pervasiveness of cooperating teacher impacts across contexts and on teaching effectiveness and retention outcomes specifically.

Teaching effectiveness. Though the majority of states in the United States set minimum requirements for years of experience in order to serve as a cooperating teacher, a few states now set minimum requirements for performance on state teaching evaluation measures. These policies assume that more effective teachers of P-12 students will be better mentors to adult learners. But is there evidence to support such policies? In short, yes. Ten studies have linked the teaching effectiveness of cooperating teachers to measures of teaching effectiveness or perceived readiness to teach candidates or recent graduates and found consistently positive and significant relationships (Bastian et al., 2020; Goldhaber et al., 2020a, 2020b, 2020c; Matsko et al., 2020; Ronfeldt et al., 2013, 2018a, 2018b, 2020a, 2020b). These studies, elaborated below, span five states and use various teaching quality measures and analytic strategies but have remarkably consistent findings.

As is true in prior sections, the large-scale, quantitative studies relating cooperating teacher and candidate teaching effectiveness began with studies using self-reported preparedness outcomes. Using surveys of all student teachers in the Chicago area, my colleagues and I used factor analysis to construct a measure of "cooperating teacher quality" based on 10 survey items about cooperating teachers, including some about the quality of teaching modeled by cooperating teachers and others about the quality of coaching/feedback/support (Ronfeldt et al., 2013). Student teachers who rated their cooperating teachers higher on this measure reported feeling better prepared, stronger teaching efficacy, and planning longer teaching careers in Chicago. A limitation, though, is that the measure for cooperating teacher quality conflated coaching with modeling and thus failed to reveal whether one or both mattered.

In a subsequent study, my collaborators and I tried to disentangle the effects of having an instructionally effective cooperating teacher from having a high-quality coach (Matsko et al., 2020). To do so, we surveyed all cooperating teachers and candidates across the Chicago area and linked surveys to district administrative data. We found both to matter. Student teachers felt better prepared to teach at graduation when their cooperating teachers provided more instructional support, frequent and adequate feedback, collaborative coaching, job search support, and a balance of autonomy and encouragement. Student teachers also felt better prepared to teach at the end of preparation when they reported that their cooperating teachers modeled more effective instruction and when their cooperating teachers had received stronger observational ratings on classroom management based on the district evaluation rubric.

Recent graduates felt better prepared when they learned to teach with more instructionally effective cooperating teachers; but were they actually more instructionally effective themselves? Four studies across four states have subsequently employed regression-based approaches to demonstrate that cooperating teachers' instructional effectiveness (as measured by observation ratings or VAMs) is positively and significantly related to the instructional effectiveness of early-career teachers they mentored (Bastian et al., 2020; Goldhaber et al., 2020a; Ronfeldt et al., 2018a, 2020b). Three studies focused on graduates' early-career observation ratings as outcomes, and all showed positive and significant associations with cooperating teachers' observation ratings but no relationship to cooperating teachers' VAMs (Bastian et al., 2020; Ronfeldt et al., 2018a, 2020b). All three studies focused on graduates' early-career VAMs as outcomes showed positive relationships with cooperating teachers' VAMs but relationships were statistically significant in only two (Bastian et al., 2020; Goldhaber et al., 2020a; Ronfeldt et al., 2018a). Graduates' early-career VAMs were unrelated to their cooperating teachers' observation ratings in the two studies that considered these cross-measure relationships (Bastian et al., 2020; Ronfeldt et al., 2018a). The above set of results suggests that recent graduates' instructional quality is related to their cooperating teachers' instructional quality, but that this relationship may be domain-specific—graduates excel on those measures of teaching effectiveness in which their cooperating teachers excelled but not on others. These trends are consistent with a causal story; to the degree that instructionally effective cooperating teachers cause candidates to be more instructionally effective, we would expect candidates to excel in the same areas of instruction as their cooperating teachers.

Despite the consistently positive correlational evidence above, we still cannot conclude that instructionally effective cooperating teachers necessarily *caused* their student teachers to become more instructionally effective. The main reason is the potential selection of candidates and cooperating teachers to one another; pairs were not randomly assigned. For example, candidates who were already predisposed to become more instructionally effective may have sought to work with more instructionally effective cooperating teachers. Experimental designs with random assignment of candidates to cooperating teachers are needed to rule out selection threats.

Thus, my colleagues and I proceeded to design and implement the Improving Student Teaching Initiative (ISTI), a field experiment utilizing within-program randomization of candidates. Using characteristics of clinical experiences shown in prior research to be related to graduates' teaching effectiveness—cooperating teacher obser-

vation ratings, VAMs, and experience; field placement school stay-ratio and school-level VAMs—we constructed a placement quality index to predict more (higher-index) and less (lower-index) promising placements among those that programs had selected for their candidates. We then randomly assigned candidates to one condition or the other. Based on post-student teaching survey data, candidates assigned to higher-index placements reported better quality instruction modeled by their cooperating teachers and more frequent and stronger coaching (Ronfeldt et al., 2018b). Though effects were not as large or consistently significant, we also found that candidates in higher-index placements felt better prepared to teach, had more opportunities to learn to teach, experienced better quality collaboration with teachers, and reported better school working conditions. In a subsequent study, we found that candidates randomly assigned to higher-index placements received better student teaching evaluations overall (observation ratings by university supervisors) and improved performance across student teaching at faster rates (Goldhaber et al., 2020c). A limitation of the ISTI studies is that the index used to identify more/less promising placements included both cooperating teacher and field placement school characteristics, making it impossible to disentangle which one drove results.

In collaboration with Tennessee Technological University and the Tennessee Department of Education, my colleagues and I developed, implemented, and evaluated the Mentors Matter Recruitment initiative to increase the average teaching effectiveness and experience of those serving as cooperating teachers (Ronfeldt et al., 2020a). We drew on the ISTI study but focused only on teacher characteristics (observation ratings, VAMs, years of experience), thus removing field placement school characteristics. Using historical course placement and administrative data, we identified potential cooperating teachers in placement districts, subjects, and grades, and created recommendation lists targeting the most instructionally effective and experienced ones. We then randomly assigned partner districts to use these recommendation lists to guide their recruitment or to use business-as-usual recruitment strategies. Across 2 years of implementation, the districts receiving recommendation lists recruited cooperating teachers with significantly and meaningfully (by one-third to two-thirds of a standard deviation) greater observation ratings, VAMs, and experience. Candidates placed with cooperating teachers in these same districts also reported feeling significantly better prepared to teach. Moreover, as in ISTI, we found some evidence that more instructionally effective and experienced cooperating teachers provided better coaching—candidates assigned to districts that received recommendation lists reported more frequent coaching, especially data-driven coaching, though differences were not always significant.

The results from the above studies suggest the relationship between cooperating teachers' and candidates' instructional effectiveness to be causal (i.e., instructionally effective cooperating teachers cause candidates to become more instructionally effective). But how? Many assume modeling as the mechanism—candidates observe, learn from, and emulate the effective instruction modeled by their cooperating teachers (Bandura, 1977; Rozelle & Wilson, 2012). However, the ISTI and Mentors Matter initiatives both suggest coaching as an alternative mechanism—teachers who are more instructionally effective with P-12 students also appear to be more effective coaches of adult learners.

Cooperating Teacher Coaching. Regardless of whether modeling or coaching explains the relationship between cooperating teachers' and candidates' teaching effectiveness—an area in need of further study—a number of studies summarized above demonstrate that cooperating teachers' coaching frequency/quality is related to candidates' self-perceived and early-career teaching effectiveness (Matsko et al., 2020; Ronfeldt et al., 2018b, 2020b). Going beyond these naturalistic, descriptive studies showing correlations between cooperating teachers' coaching and candidates' teaching effectiveness, four studies have developed, implemented, and studied coaching PD programs for cooperating teachers. All provide evidence that cooperating teachers' coaching can improve candidates' teaching effectiveness, including three studies using experimental designs to demonstrate that these relationships are likely causal.

Giebelhaus and Bowman (2002) randomly assigned 28 cooperating teachers across two programs to either receive coaching PD or business-as-usual supports. The 14 teachers assigned to coaching PD attended 10 training sessions (30 hours), which focused on general principles and practices from the *Praxis III/Pathwise* framework, analysis of videotaped lessons, and role playing. Trained, external evaluators rated the instruction of candidates paired with trained cooperating teachers as stronger. Following a similar model, McQueen (2018) randomly assigned coaches in an alternative certification program to receive PD on how to provide focused and choice-based coaching. Mentees assigned to coaches who attended the PD reported better quality coaching and received better observation ratings themselves.

Becker et al. (2019) randomly assigned 130 cooperating teachers to one of three coaching PD groups or a control group. However, substantial noncompliance issues post-randomization resulted in only 59 participating cooperating teachers and raised concerns over whether results can be interpreted as causal. The authors found evidence that the cooperating teachers who participated in PD changed their coaching practice in ways that aligned with the PD, that their candidates reported better collaborative exchanges and constructive feedback during coaching conversations, and that their candidates were more successful at addressing disruptive behaviors in their classrooms. Finally, a non-experimental study by Gareis and Grant (2014) provides descriptive evidence consistent with the aforementioned experimental studies—cooperating teachers who received PD felt more efficacious about their coaching roles and their candidates were rated as more instructionally effective by university supervisors. The findings from these studies are consistent with the extensive literature demonstrating the positive effects of mentor coaching on inservice teachers' instructional performance (Kraft et al., 2018).

Retention. As mentioned above, my colleagues and I found that student teachers in Chicago who rated their cooperating teachers higher on a "cooperating teacher quality" factor—combining survey items about coaching and modeling by cooperating teachers—planned longer teaching careers in Chicago but not longer careers generally (Ronfeldt et al., 2013). Goldhaber et al. (2020b) is the only study linking cooperating teacher instructional effectiveness (using VAMs) to graduates' observed retention, finding no relationship.

Summary. Candidates who learn to teach with instructionally effective cooperating teachers are more instructionally effective themselves. This relationship has been reproduced across multiple geographic and subject matter contexts, using various methodological approaches and measures. Recent studies using experimental designs show the relationship to be causal and that it is possible to substantially improve the instructional effectiveness of teachers serving as cooperating teachers by using administrative data to inform recruitment decisions. Studies reviewed in this section suggest that cooperating teachers likely impact candidates' instructional effectiveness through both modeling and coaching. Regarding the latter, experimental evidence has shown that PD can improve coaching practice among cooperating teachers and, in turn, improve candidates' teaching effectiveness.

Practice teaching need not occur only in field settings and coaching need not be provided only by cooperating teachers. A number of teacher educators have developed teaching simulations to provide structured opportunities during coursework for candidates to rehearse and receive feedback on developing practice. As I turn to the literature on coursework, I begin with studies about these practice-based efforts.

COURSEWORK

In this section, I review the literature linking coursework during preservice preparation to teaching effectiveness and retention. I begin with recent practice-based efforts to integrate teaching simulations into coursework. After, I turn to the more extensive literature on coursework quantity.

Practice-Based Coursework Simulations with Coaching

Over the past two decades, a number of scholars have developed a strong theoretical basis for practice-based teacher education (PBTE), as well as innovative pedagogical and curricular approaches aligned with theory (Ball & Cohen, 1999; Grossman et al., 2009). In addition to *decomposing* teaching into a set of specific "core" or "high-leverage" practices and *representing* (e.g., video exemplars) these practices to learning teachers, practice-based approaches also include approximations of practice, including teaching simulations, combined with coaching. Simulations provide candidates with designed opportunities to enact practice in settings where they can experiment and even fail without harming children. Only recently have studies examined the impact of PBTE on teaching effectiveness. That said, today's PBTE reforms have some theoretical, pedagogical, and empirical overlap with "microteaching," a practice-focused movement from the 1950s-1970s (Zeichner, 2012), which also prioritized the use of simulations in preparing preservice teachers and similarly found coaching and feedback to be critical in promoting pedagogical skills (Cooper, 1967; Joyce & Showers, 1981).

Teaching effectiveness. In three studies, the authors required all candidates within

participating programs to engage with simulations but randomly assigned them to receive coaching/feedback or not, generally finding positive effects of pairing coaching with simulations. Bardach et al. (2020) developed online teaching scenarios with various possible teaching responses that participants could select and then randomly assigned candidates in two programs to three conditions—control, feedback-only, and feedback-plus-reflection. Compared to the control, candidates in both feedback conditions reported feeling better prepared to teach, while candidates in the feedback-plus-reflection condition also reported higher self-efficacy.

Cohen et al. (2020) created online and interactive mixed-reality simulations aiming to assess and develop classroom management ("redirection") skills in which candidates interact virtually with avatar "students" controlled by "interactors" in prespecified ways based on candidates' actions. Candidates were randomly assigned to either (1) reflect on their practice without coaching, (2) receive coaching on their performances, or (3) receive coaching both during ("bug-in-ear") and after their simulations. During follow-up simulations, candidates in the two coaching conditions demonstrated much stronger skills at redirection than candidates that only reflected on their performance; they also rated "student" behavior less harshly.

The prior study mostly demonstrated that individualized feedback makes a difference. In a follow up study, Cohen and Wiseman (under review) tested whether practice-based coaching did so; trained coaches identified a skill in need of improvement and designed role plays for their candidates to practice that skill while providing feedback and support. Candidates who were randomized to receive practice-based coaching outperformed candidates in the reflection-only condition by more than one standard deviation in the quality of feedback they provided "students" during subsequent simulated text-based discussions.

The experimental studies described above demonstrate that practice-based interventions with coaching improve teaching in a simulator, but do they lead to better performance in "real" classrooms? One study suggests they can. Instead of randomizing candidates within the same course or program, Kang and Windschitl (2018) studied the effects of an entire practice-based science methods course on two cohorts of graduates' early-career performance. The course was designed to develop candidates' abilities to successfully enact a set of "core" practices in simulations accompanied by feedback from course instructors. The authors then followed 41 graduates of the core practices group (CPG) over 2 years, observing them in their classroom settings. They also recruited 13 first-year science teachers from various college-recommending (traditional) programs as a comparison group. The authors found that the CPG teachers outperformed the comparison teachers on four metrics associated with their students' opportunities to learn, though without randomization we cannot confidently attribute observed differences to the practice-based course; for example, baseline differences between groups could explain these results. However, new experimental evidence provides causal evidence that practice-based, inservice PD combining mixed-reality simulations with coaching positively impacts instruction in real classroom settings (Garrett & Smith, 2020).

Retention. I am not aware of any studies linking practice-based interventions to graduates' retention.

Summary. The emerging line of research on PBTE and, specifically, the use of simulations (including mixed-reality) with coaching demonstrates strong, positive impacts on feelings of preparedness, efficacy, and candidate performance in simulated settings. One study has suggested that positive effects on teaching effectiveness carry into real inservice classrooms as well. This body of research also demonstrates that simulated practice opportunities with only self-reflection is insufficient; they must be paired with individualized coaching/feedback from teacher educators.

Coursework Quantity

Prior literature reviews suggest that both content and pedagogy coursework is positively associated with teaching effectiveness (Floden & Meniketti, 2005; Wilson et al., 2002). Floden and Meniketti (2005), the most recent of these reviews, concluded that the literature generally suggests that completing more content and pedagogy coursework is related to stronger teaching effectiveness among science and mathematics teachers.[4] Floden and Meniketti (2005) also reviewed a number of studies on foundation courses, finding positive relationships with teachers' knowledge but none linked to teaching effectiveness. The authors did not report on, and I am not aware of, any studies linking coursework to retention prior to 2005.

Teaching effectiveness. Though the pre-2005 literature suggests generally positive associations between coursework completion and teaching effectiveness, newer studies have found null or mixed results. Harris and Sass (2011) found undergraduate content and pedagogy coursework to be mostly unrelated to teachers' VAMs, with pedagogy coursework sometimes being negatively related. This study's only positive association was that high school mathematics teachers performed better when they completed more content coursework. Constantine et al. (2009), who used an experimental design to compare traditionally certified with alternatively certified teachers, found in secondary, correlational (regression-based) analyses that the amount or kind (e.g., pedagogical, content) of coursework did not contribute significantly to differences in instructional effectiveness. A limitation of both of these studies is that their analytic methods did not adjust for other preparation features that might be correlated with coursework and could explain observed relationships.

Three other studies focused on both graduates' mathematics and ELA VAMs as outcomes and included extensive controls for preparation features and other factors. Boyd et al. (2009) examined all programs preparing teachers in NYC and studied only content courses.[5] The other two studies were in North Carolina and considered content,

[4] In mathematics, relying on Monk (1994) and Monk and King (1994), they say the evidence is stronger for pedagogy than content coursework. In science, they refer to a meta-analysis by Druva and Anderson (1983).

[5] Boyd et al. (2009) constructed a number of other measures for preparation in pedagogy, mathematics, ELA, and other areas but used survey items about "opportunities to practice" that did not differentiate coursework from clinical opportunities.

pedagogy, foundations, and technology coursework, with Henry et al. (2013) examining all programs and Preston (2017) only middle school programs. Regarding mathematics VAMs as outcomes, two studies found content coursework to be positively related (Boyd et al., 2009; Henry et al., 2013) and the other found it to be mostly negatively related (Preston, 2017). The two studies that considered other kinds of courses found mostly negative relationships for pedagogy courses and null relationships for foundations and technology courses.

Turning to ELA VAMs, results for ELA content courses were mixed. Henry et al. (2013) found null results, Preston (2017) found null results in two specifications and positive results in the third, and Boyd et al. (2009) found negative associations with first-year VAMs but positive associations with second-year VAMs. Henry et al. (2013) found null relationships for pedagogy coursework, while Preston (2017) found null relationships in two models and positive relationships in the third. Regarding foundations and technology courses, Henry et al. (2013) found positive results while Preston (2017) found mixed results.

In contrast to studies using graduates' early-career VAMs, those focused on graduates' self-perceived preparedness revealed mostly positive associations with amount of coursework. Using a nationally representative data set, Ronfeldt et al. (2014) found that early-career teachers felt better prepared when they had completed more methods-related coursework during initial preparation. Likewise, in their analysis of Chicago-area programs, Ronfeldt et al. (2020b) found that completing more coursework prior to student teaching was positively associated with candidates' own feelings of preparedness but unrelated to cooperating teachers' evaluations of candidates' preparedness and negatively related to graduates' first-year observation ratings.

Retention. To my knowledge, only one study has directly examined the association between the amount of preparation coursework and retention. Using multiple waves of nationally representative SASS-TFS data, Ronfeldt et al. (2014) found that graduates who completed more methods-related coursework were somewhat more likely to stay in teaching. Although the correlation between retention and the number of methods courses completed was not statistically significant overall, completing three to four methods courses approximately doubled the odds of staying in teaching relative to completing no methods coursework. Moreover, the relationship between coursework completion and retention was most pronounced for individuals who completed no practice teaching—completing an additional methods course increased the odds of staying in teaching between 15 and 20 percent. Using a single SASS-TFS wave and focusing specifically on mathematics and science teachers, Ingersoll and May (2012) combined methods-related coursework, duration of student teaching, and other variables into a single measure for extent of pedagogical preparation and found that measure to positively predict retention.

Summary. Candidates that complete more preparation coursework tend to feel better prepared to teach. But are they more instructionally effective? It does not appear to be the case. Studies linking coursework completion to VAMs suggest null or mixed results with findings varying by type of coursework and subject area in ways that were not

consistent across studies. The one study linking coursework to observation ratings suggests a negative relationship, though the latter only considered the amount of coursework completed prior to student teaching. The two studies linking coursework to teacher retention provide suggestive evidence for a positive relationship, though more studies are needed.

CONCLUSION

After decades of initiatives and literature reviews concluding that we know little about features of preparation related to teaching effectiveness and retention, this paper suggests that recent scholarship has made meaningful progress. At the most general level, large-scale quantitative studies suggest that the quality, more than the quantity, of preparation makes a difference. Though candidates that complete more courses and more weeks of student teaching appear to have stronger retention and feel better prepared to teach, there is little evidence that they are more instructionally effective. On the other hand, better quality clinical experiences are consistently associated with stronger retention, feelings of preparedness, and observed teaching effectiveness. What makes for "better quality" clinical experiences? The literature suggests clinical experiences that (1) are aligned with other program dimensions including coursework (program coherence); (2) occur in field placement schools with strong professional learning environments and that match employment schools on student demographics, school, and grade levels; and (3) include instructionally effective cooperating teachers who also provide high-quality coaching. Additionally, emerging evidence suggests that coursework quality also likely matters: practice-based courses including carefully-designed simulations that pair opportunities to rehearse teaching with individualized coaching have the potential to improve teaching effectiveness.

Recommendations for Policy and Practice

What does all of this mean for policymakers and practitioners? As a general principle, the findings of this paper suggest a need to place more emphasis on the quality than the quantity of preparation experiences. The implication is not, though, to reduce the amount of coursework and clinical experiences, as these are related to retention and feelings of preparedness. Rather, it is that increasing the amount of coursework or clinical experiences alone, without simultaneous and deliberate attention to their quality, is unlikely to improve observed teaching effectiveness.

Though states commonly have teacher education policies regarding the number of credits for different kinds (e.g., methods, content, foundations) of courses and duration (e.g., 12 weeks) of student teaching, less common are policies that target the quality of these experiences. One implication of this paper is to focus reforms on the design of high-quality clinical experiences with a cornerstone being the recruitment of instructionally effective and experienced cooperating teachers. Though many states already have minimum years of experience (quantity) requirements for teachers to serve as cooperating teachers, only a handful of states have minimum requirements for teaching effectiveness measures like observation ratings. Policymakers could consider set-

ting ambitious minimum teaching effectiveness requirements for teachers to serve as cooperating teachers and then use historical administrative data to help identify for program and district recruiters those teachers who qualify.

Additionally, we know that coaching and feedback are critical to prospective teacher learning, yet cooperating teachers rarely receive PD in coaching practices. Less than 10 percent of cooperating teachers affiliated with traditional route programs in Chicago, for instance, reported receiving coaching PD (Matsko et al., 2021). There is evidence that coaching PD can improve cooperating teachers' coaching practices and, in turn, their mentees' performance. Policymakers might then consider investing in coaching PD for cooperating teachers, and offering compensation and incentives (e.g., course releases, PD credits, certification) for their strongest teachers and coaches to serve, as doing so promises to improve the instructional effectiveness of the next generation of teachers. Policymakers might also consider requiring that, to serve as cooperating teachers, teachers complete coaching PD or receive some kind of coaching certification. The Louisiana Department of Education, for example, has invested in state-approved coaching programs that, since 2017, have recruited and trained about 2,000 cooperating teachers; completion of these coaching programs results in certification—a new requirement for teachers to serve as cooperating teachers in the state. In addition, the Louisiana Department of Education is providing $1,000 per year for each certified cooperating teacher that mentors preservice candidates. Pairing minimum evaluation and experience requirements with required coaching certification is one way that policymakers can ensure that cooperating teachers are both effective models and coaches.

Designing high-quality clinical experiences also means placing candidates in field placement schools with strong professional learning environments, where teacher learning is known to be a critical part of student learning (Feiman-Nemser, 1983; Lightfoot, 1986; Little, 1982). The recent literature suggests that program and district leaders in charge of selecting field placement schools should target schools with high-quality collaboration among teachers, low teacher turnover, and faculty with a track record of learning and achievement gains; as with cooperating teachers, recruiters can use administrative and, where available, survey data to target placement schools with these characteristics. One implication is that federal and state agencies consider providing more funding to support schools with strong professional learning environments to continue to offer promising field placement opportunities for student teachers and residents.

Graduates are also more instructionally effective when they are employed in schools and classrooms that more closely match the student sociodemographic characteristics, school level (e.g., elementary), and grade level of their placement schools and classrooms. The implications are many. Employers could recruit and hire recent graduates who student taught in schools and classrooms with characteristics that match their own schools and classrooms, while program leaders could select placements based on the kinds of schools and classrooms in which candidates plan to eventually seek employment.[6] In addition, policymakers can play a role in promoting and incentivizing

[6] One potential limitation of the latter approach is that it might constrain candidates' opportunities to experiment with teaching in other kinds of schools than those that initially appeal to them. Moreover, some teacher educators and programs may want to promote a wide range of school experiences so as to ensure graduate flexibility.

residency and other programs that prepare candidates and secure their employment in the district contexts in which they intend to teach but also have a promising empirical basis.[7]

More broadly, these findings suggest promoting policies that establish partnerships between teacher education programs and surrounding districts. Such partnerships promise to be mutually beneficial in that they will likely strengthen the field placements that teacher education programs can provide for their candidates, which, in turn, will strengthen the teaching effectiveness of graduates that local districts can hire. The latter point is bolstered by evidence that at least 40 percent of candidates are hired in the same districts in which they completed their student teaching experiences (Krieg et al., 2020; Ronfeldt et al., 2018a). Establishing district partnerships also means that teacher education programs can work toward being responsive to local district needs by selecting placements that align with district needs in terms of grade level, school level, and student demographics. Doing so not only promises to offer candidates preparation experiences that build population- and context-specific knowledge and skills (human capital) that will make them more employable and, once hired, more instructionally effective, but will also help districts to address local shortages.

Finally, designing high-quality field placements means ensuring that they align with other program dimensions such as coursework. An implication is for program leaders to incorporate more structures designed to promote program–clinical alignment, including integrating more hours of clinical experience deliberately into coursework, requiring supervisors to observe candidates and meet with other program faculty more often, and having program leaders, who know more about program goals and needs than district and school leaders or candidates themselves, take primary responsibility for selecting placements (Grossman et al., 2008).

Given that the relationship between the amount of coursework completion and teaching effectiveness is mixed, one might be tempted to conclude that practitioners and policymakers deemphasize coursework in favor of clinical experiences. Such a conclusion would be premature, though, not only because coursework quantity is related to other outcomes (retention, feelings of preparedness) but also because not enough research has considered whether the quality and content (e.g., Youngs & Qian, 2013) of coursework, as opposed to quantity, matters. To this point, recent studies have found that integrating practice-focused opportunities, including simulations with individualized coaching, into coursework has promise. Related, there is also some evidence that practice-focused and fieldwork-aligned coursework is associated with better prepared teachers (Boyd et al., 2009; Grossman et al., 2008; Kang & Windschitl, 2018). If future research continues to suggest that the quality and content more than quantity of courses is associated with better graduate outcomes, then the policy implications might be to prioritize coursework quality and content as much as, if not more than, the number of course credits.

[7] Policymakers might consider other kinds of context-specific programs, like "grow-your-own" programs that prepare candidates in and for specific district context, especially because these programs typically target paraprofessionals and other staff already working in local schools who have already developed knowledge of skills specific to these communities. Though an evidentiary base for grow-your-own programs is beginning to emerge (Gist et al., 2019), we need more studies linking these programs to graduate workforce outcomes including teaching effectiveness and retention.

It is notable that evidence from studies linking preparation to teaching effectiveness and retention overwhelmingly identifies features of high-quality clinical experiences as making a difference; even the types of coursework shown to be related to these outcomes were built around simulated clinical practice opportunities with feedback. Taken together, this evidence seems to support models of teacher preparation that center practice opportunities for candidates to learn about teaching while engaged in the enactment of teaching in supportive contexts with feedback and modeling from cooperating teachers or other teacher educators who are skilled teachers and coaches.

Recommendations for Future Research

This review indicates a number of gaps in the current literature and corresponding recommendations for future research. First, more studies are needed that link preparation features to retention and observation ratings or other measures of teaching effectiveness beyond VAMs and self-perceived preparedness. Second, while research about clinical experiences moved from a focus on quantity to quality, most of the large-scale studies about coursework continue to focus on the number of courses; we need more research on the quality and content of courses. One promising future direction might be to build on studies that have gathered information from candidates about different kinds of "opportunities to learn" that they have experienced (Boyd et al., 2009; Youngs et al., under review); though these prior studies asked about opportunities to learn across preparation experiences, future studies could examine those opportunities specific to coursework. Among studies that have previously examined coursework quality, those focused on coursework simulations and other practice-based reforms have shown promise, though we need more evidence for effects on teaching effectiveness outside of the simulator and inside real classrooms.

Large-scale studies have considered a relatively small number of possible aspects of preparation. Future large-scale studies should consider many more preparation features, as well as the ways in which different features work together (and do not). As an example, there is growing evidence that residency programs are related to positive workforce outcomes (Guha et al., 2016). Though there is not full consensus on what defines the residency model, residency programs tend to integrate many of the features identified as promising in this review, including high-quality cooperating teachers combined with longer clinical experiences and program-district partnerships that ensure a match between the characteristics of the schools in which residents complete their residency experiences and the schools in which they are eventually employed. Future research should consider whether specific features that make up residency programs are driving positive outcomes or whether the combination of features is more than the sum of its parts.

Justice-oriented teacher preparation is another area in need of further study. Though many scholars have contributed to its strong theoretical and pedagogical base, few studies have linked justice-oriented reforms to candidates' teaching and retention outcomes (Sleeter, 2001). Enterline et al. (2008) used surveys to develop a "learning to teach for social justice" beliefs scale and found that recent graduates from a social justice program scored significantly higher than incoming candidates, and that graduates maintained

scores 1 year after graduating, but they did not go beyond belief outcomes. Future large-scale studies should consider features of justice-oriented preparation that predict better retention and teaching effectiveness; building off of Enterline et al. (2008), studies may want to broaden teaching effectiveness measures to better align with justice-oriented goals (more on this below). A major obstacle, though, in pursuing research in these proposed areas, and in teacher education generally, is that few states have developed data systems linking features of preparation to measures of teaching effectiveness and retention (National Academies of Sciences, Engineering, and Medicine, 2020). Such links may require stronger partnerships between preparation programs and districts.

The vast majority of studies that focused on teaching effectiveness used VAMs, observation ratings, or self-perceived preparedness (survey-based). Identifying and promoting preparation features correlated with these outcomes will likely increase graduates' teaching effectiveness on these same metrics. We must then take seriously whether existing measures represent the kinds of teaching we want to reproduce. There exists some evidence that these measures are valid, reliable, and signal dimensions of teaching worth promoting. At the same time, they may fail to reflect aspects of teaching that we care about, like teaching that promotes equity, well-being, high-cognitive demand, critical consciousness, and compassion among learners. Moreover, there is evidence that Black and male teachers, as well as teachers who teach in classrooms with more Black, Hispanic, lower-achieving, and lower-income students receive lower observation ratings for reasons other than instructional quality (Campbell, 2020; Campbell & Ronfeldt, 2018; Grissom & Bartanan, forthcoming; Jiang & Sporte, 2016; Steinberg & Garrett, 2016; Whitehurst et al., 2014). Likewise, there is evidence that certain kinds of VAMs are related to student and classroom characteristics, including being lower in some cases for teachers in classrooms with more students who are lower-income, Hispanic, and English language learners (Hill et al., 2011; Newton et al., 2010; Rothstein, 2009). As such, we risk perpetuating systemic inequities by privileging these measures. This is not to suggest that we throw out existing measures or conclusions from this paper, but instead it is a call to acknowledge their limitations while we work to develop and create policies to promote other measures of teaching effectiveness that are likely more equitable and more fully capture its complexity, including ones focused on non-academic outcomes for students such as critical consciousness and well-being. As we develop new measures, this paper offers guidance on preparation features to consider associating with them, as well as methods for doing so.

A related consideration is the growing evidence demonstrating benefits of being taught by teachers of color for students' learning, socioemotional, and other outcomes, especially for same-race students (Bristol & Martin-Fernandez, 2019; Grissom et al., 2015). Students of color taught by teachers with similar racial or ethnic backgrounds are less likely to experience exclusionary discipline (Holt & Gershenson, 2015; Hughes et al., 2020), have better learning outcomes (Dee, 2004; Egalite & Kisida, 2018), are more likely to progress to advanced courses (Grissom et al., 2020), and have better school attendance and graduation rates (Gershenson et al., 2018; Holt & Gershenson, 2015). This evidence base suggests that programs that are successfully recruiting and employing teachers of color—an area of success among many residency and alternative route programs, for example (Grossman & Loeb, 2021; Guha et al., 2016)—are effectively increasing teaching effectiveness particularly for students of color. Given its emphasis

on features of preparation rather than features of recruitment, this paper did not review studies about the strategies used by programs that have had success recruiting candidates of color; however, this is a critical line of inquiry for future study.

A final implication of this paper for research is that solving anything as complex as improving teacher education and teaching quality will likely require a methodologically pluralistic approach (Moss & Haertel, 2016). This paper privileges large-scale, quantitative studies that are needed to reveal which preparation features are associated with teaching effectiveness and retention. However, as I tried to underscore in each section above, theory and research produced by qualitative studies typically identified the features that large-scale, quantitative studies subsequently linked to retention and teaching effectiveness. Additionally, experimental and causal inference methodologies are needed to ensure that associations between preparation features and focal outcomes, identified through large-scale correlational studies, are truly causal. The research on cooperating teachers is a case in point. Qualitative studies recognized cooperating teachers as powerful socializing agents that motivated large-scale, descriptive studies correlating the instructional effectiveness of cooperating teachers and candidates. In turn, experimental research found this relationship to be causal, but has been unable to establish the mechanism (modeling or coaching); this will hopefully motivate new, likely qualitative studies, to investigate the mechanism. Through methodological pluralism we know much more now about which aspects of teacher preparation predict teaching effectiveness and retention; it is likely that only through methodological pluralism will our understanding continue to grow.

REFERENCES

Allensworth, E., Ponisciak, S., & Mazzeo, C. (2009). *The schools teachers leave: Teacher mobility in Chicago public schools.* Consortium on Chicago School Research.

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession* (pp. 3-32). Jossey Bass.

Bandura, A. (1977). *Social learning theory*. Prentice Hall.

Bardach, L., Klassen, R., Durksen, T. L., Rushby, J., Bostwick, K., & Sheridan, L. (2020). *The power of feedback and reflection: A brief online scenario-based learning activity designed to increase student teachers' self-efficacy and classroom readiness.* PsyArXiv. http://doi.org/10.31234/osf.io/whsny.

Bastian, K. C., Patterson, K. M., & Carpenter, D. (2020). Placed for success: Which teachers benefit from high-quality student teaching placements? *Educational Policy*. https://doi.org/10.1177/0895904820951126.

Bastian, K. C., Sun, M., & Lynn, H. (2019). What do surveys of program completers tell us about teacher preparation quality? *Journal of Teacher Education*, *72*(1), 11-26.

Becker, E. S., Waldis, M., & Staub, F. C. (2019). Advancing student teachers' learning in the teaching practicum through Content-Focused Coaching: A field experiment. *Teaching and Teacher Education*, *83*, 12-26.

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

Bristol, T. J., & Martin-Fernandez, J. (2019). The added value of Latinx and Black teachers for Latinx and Black students: Implications for policy. *Policy Insights from the Behavioral and Brain Sciences*, *6*(2), 147-153.

Bryk, A., Sebring, P., Allensworth, E., Easton, J., & Luppescu, S. (2010). *Organizing schools for improvement: Lessons from Chicago.* University of Chicago Press.

Buchmann, M., & Floden, R. (1993). Coherence: The rebel angel. In M. Buchmann & R. Floden (Eds.), *Detachment and concern: Conversations in the philosophy of teaching and teacher education* (pp. 222-235). Cassell Press.

California State University. (2002). First system wide evaluation of teacher education programs in the California State University: Summary report. Long Beach, CA: Office of the Chancellor.
Campbell, S. L. (2020). Ratings in Black and White: A quantcrit examination of race and gender in teacher evaluation reform. *Race Ethnicity and Education*. https://doi.org/10.1080/13613324.2020.1842345.
Campbell, S., & Ronfeldt, M. (2018). Observational evaluations of teachers: Measuring more than we bargained for? *American Educational Research Journal*, *55*(6), 1233-1267.
Castle, S., Arends, R. I., & Rockwood, K. D. (2008). Student learning in a professional development school and a control school. *Professional Educator*, *32*(1), 1-15.
Clift, R. T., & Brady, P. (2005). Research on methods courses and field experiences. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education*. Lawrence Erlbaum Associates.
Cochran-Smith, M., & Zeichner, K. M. (Eds.). (2009). *Studying teacher education: The report of the AERA panel on research and teacher education*. Routledge.
Cohen, J., & Wiseman, E. (under review). *Approximating complex practice: Teacher simulation of text-based discussion*.
Cohen, J., Wong, V., Krishnamachari, A., & Berlin, R. (2020). Teacher coaching in a simulated environment. *Educational Evaluation and Policy Analysis*, *42*(2), 208-231.
Constantine, J., Player, D., Silva, T., Hallgren, K., Grider, M., & Deke, J. (2009). *An evaluation of teachers trained through different routes to certification, final report* (Report No. NCEE 2009-4043). National Center for Education Evaluation and Regional Assistance.
Cooper, J. (1967). Developing specific teaching skills through micro-teaching. *The High School Journal*, *51*(2), 80-85.
Darling-Hammond, L. (1994). *Professional development schools: Schools for developing a profession*. Teachers College Press.
Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of economics and statistics*, *86*(1), 195-210.
Dewey, J. (1904). The relation of theory to practice in teacher education. In M. Borrowman (Ed.), *Teacher education in America: A documentary history* (pp. 140-171). Teachers College Press.
Druva, C. A., & Anderson, R. D. (1983). Science teacher characteristics by teacher behavior and by student outcome: A meta-analysis of research. *Journal of Research in Science Teaching*, *20*(5), 467-479.
Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own competence. *Current Directions in Psychological Science*, *12*(3), 83-87.
Egalite, A. J., & Kisida, B. (2018). The effects of teacher match on students' academic perceptions and attitudes. *Educational Evaluation and Policy Analysis*, *40*(1), 59-81.
Feiman-Nemser, S. (1983). Learning to teach. In L. S. Shulman & G. Sykes (Eds.), *Handbook of teaching and policy* (pp. 150-170). Longman.
Fenstermacher, G. D., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record*, *107*(1), 186-213.
Floden, R. E., & Meniketti, M. (2005). *Research on the effects of coursework in the arts and sciences and in the foundations of education.* In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA Panel on Research and Teacher Education* (pp. 261-308). Lawrence Erlbaum Associates Publishers; American Educational Research Association.
Gareis, C. R., & Grant, L. W. (2014). The efficacy of training cooperating teachers. *Teaching and Teacher Education*, *39*, 77-88.
Garrett, R., & Smith, T. (2020). *SIM PD study*. American Institute of Research. https://www.air.org/sites/default/files/SIM-PD-Study-Summary-April-2020.pdf.
Gershenson, S., Hart, C., Hyman, J., Lindsay, C., & Papageorge, N. W. (2018). *The long-run impacts of same-race teachers* (Report No. w25254). National Bureau of Economic Research.
Giebelhaus, C. R., & Bowman, C. L. (2002). Teaching mentors: Is it worth the effort? *The Journal of Educational Research*, *95*(4), 246.
Gist, C. D., Bianco, M., & Lynn, M. (2019). Examining grow your own programs across the teacher development continuum: Mining research on teachers of color and nontraditional educator pipelines. *Journal of Teacher Education*, *70*(1), 13-25.

Gitomer, D. H., & Bell, C. A. (2013). Evaluating teaching and teachers. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, vol. 3: Testing and assessment in school psychology and education* (pp. 415-444). American Psychological Association.

Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record*, *109*(4), 877-896.

Goldhaber, D., Krieg, J. M., & Theobald, R. (2016). *Does the match matter? Exploring whether student teaching experiences affect teacher effectiveness and attrition* (CALDER Working Paper No. 149). National Center for Analysis of Longitudinal Data in Education Research.

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

Goldhaber, D., Krieg, J., & Theobald, R. (2020a). Effective like me? Does having a more productive mentor improve the productivity of mentees? *Labour Economics*, *63*, 101792.

Goldhaber, D., Krieg, J., Theobald, R., & Goggins, M. (2020b). *Front end to back end: Teacher preparation, workforce entry, and attrition* (CALDER Working Paper No. 246-1220). National Center for Analysis of Longitudinal Data in Education Research.

Goldhaber, D., Ronfeldt, M., Cowan, J., Gratz, T., Bardelli, E., Truwit, M., & Mullman, H. (2020c). *Room for improvement? Mentor teachers and the evolution of teacher preservice clinical evaluations* (CALDER Working Paper No. 239-0620). National Center for Analysis of Longitudinal Data in Education Research.

Greenberg, J., Pomerance, L., & Walsh, K. (2011). *Student teaching in the United States.* National Center on Teacher Quality.

Grissom, J., & Bartanen, B. (forthcoming). Investigating race and gender biases in high-stakes teacher observations.

Grissom, J. A., Kabourek, S. E., & Kramer, J. W. (2020). Exposure to same-race or same-ethnicity teachers and advanced math course-taking in high school: Evidence from a diverse urban district. *Teachers College Record*, *122*(7), 1-42.

Grissom, J. A., Kern, E. C., & Rodriguez, L. A. (2015). The "representative bureaucracy" in education: Educator workforce diversity, policy outputs, and outcomes for disadvantaged students. *Educational Researcher*, *44*(3), 185-192.

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

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, *111*(9), 2055-2100.

Grossman, P., & Loeb, S. (Eds.). (2021). *Alternative routes to teaching: Mapping the new landscape of teacher education*. Harvard Education Press.

Grossman, P., McDonald, M., Hammerness, K., & Ronfeldt, M. (2008). Constructing coherence: Structural predictors of perceptions of coherence in NYC teacher education programs. *Journal of Teacher Education*, *59*, 273-287.

Grossman, P., Ronfeldt, M., & Cohen, J. (2011). The power of setting: The role of field experience in learning to teach. In K. Harris, S. Graham, T. Urdan, A. Bus, S. Major, & H. L. Swanson (Eds.), *American Psychological Association (APA) educational psychology handbook, vol. 3: Applications to teaching and learning* (pp. 311-334). American Psychological Association.

Guha, R., Hyler, M. E., & Darling-Hammond, L. (2016). *The teacher residency.* Learning Policy Institute.

Haberman, M. (1995). Selecting "star" teachers for children and youth in urban poverty. *Phi Delta Kappan*, *76*(10), 777.

Hammerness, K. (2006). From coherence in theory to coherence in practice. *Teachers College Record*, *108*(7), 1241-1265.

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, *95*(7-8), 798-812.

Henke, R., Xianglei, C., Geis, S., & Knepper, P. (2000). Progress through the teacher pipeline: 1992-1993 college graduates and elementary/secondary school teaching as of 1997. National Center for Education Statistics.

Henry, G. T., Campbell, S. L., Thompson, C. L., Patriarca, L. A., Luterbach, K. J., Lys, D. B., & Covington, V. M. (2013). The predictive validity of measures of teacher candidate programs and performance: Toward an evidence-based approach to teacher preparation. *Journal of Teacher Education*, *64*(5), 439-453.

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, *48*, 794-831.

Holt, S., & Gershenson, S. (2015). *The impact of teacher demographic representation on student attendance and suspensions* (IZA Discussion Paper No. 9554). Institute for the Study of Labor.

Hughes, C., Bailey, C. M., Warren, P. Y., & Stewart, E. A. (2020). "Value in diversity": School racial and ethnic composition, teacher diversity, and school punishment. *Social Science Research*, *92*, 102481.

Ingersoll, R. M., & May, H. (2012). The magnitude, destinations, and determinants of mathematics and science teacher turnover. *Educational Evaluation and Policy Analysis*, *34*(4), 435-464.

Ingersoll, R., Merrill, L., & May, H. (2014). *What are the effects of teacher education and preparation on beginning teacher attrition?* (CPRE Research Report 78). Consortium for Policy Research in Education.

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

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

Joyce, B. R., & Showers, B. (1981). Transfer of training: The contributions of "coaching." *Journal of Education*, *163*(2), 163-172.

Kang, H., & Windschitl, M. (2018). How does practice-based teacher preparation influence novices' first-year instruction? *Teachers College Record*, *120*(8).

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, *88*(4), 547-588.

Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, 36(4), 476-500.

Krieg, J., Goldhaber, D., & Theobald, R. (2020). *Disconnected development? The importance of specific human capital in the transition from student teaching to the classroom* (CALDER Working Paper No. 236-0520). National Center for Analysis of Longitudinal Data in Education Research.

Labaree, D. F. (1997). Public goods, private goods: The American struggle over educational goals. *American Educational Research Journal*, *34*(1), 39-81.

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, *32*(3), 465-491.

Latham, N., Mertens, S. B., & Hamann, K. (2015). A comparison of teacher preparation models and implications for teacher attrition: Evidence from a 14-year longitudinal study. *School-University Partnerships*, *8*(2), 79-89.

Lee, C. (2007). *Culture, literacy, and learning: Taking bloom in the midst of the whirlwind*. Teachers College Press.

Lightfoot, S. L. (1986). On goodness in schools: Themes of empowerment. *Peabody Journal of Education*, *63*(3), 9-28.

Little, J. W. (1982). Norms of collegiality and experimentation: Workplace conditions of school success. *American Educational Research Journal*, 19(3), 325-340.

Loeb, S., Kalogrides, D., & Beteille, T. (2012). Effective schools: Teacher hiring, assignment, development, and retention. *Education Finance and Policy*, 7(3), 269-304.

Matsko, K. K., Ronfeldt, M., & Greene Nolan, H. (2021). How different are they? Comparing preparation offered by traditional, alternative, and residency pathways. *Journal of Teacher Education*. https://doi.org/10.1177/00224871211015976

Matsko, K. K., Ronfeldt, M., Green Nolan, H., Klugman, J., Reininger, M., & Brockman, S. L. (2020). Cooperating teacher as model and coach: What leads to student teachers' perceptions of preparedness? *Journal of Teacher Education*, *71*(1), 41-62.

McIntyre, D. J., Byrd, D. M., & Foxx, S. M. (1996). Field and laboratory experiences. In J. Sikula (Ed.), *Handbook of research on teacher education*. Simon & Schuster Macmillan.

McQueen, K. (2018). Promoting instructional improvement: Promising evidence of coaching that benefits teachers' practice [Unpublished doctoral dissertation]. University of Michigan. https://deepblue.lib.umich.edu/handle/2027.42/143904.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, *13*(2), 125-145.

Monk, D. H., & King, J. A. (1994). Multilevel teacher resource effects in pupil performance in secondary mathematics and science: The case of teacher subject matter preparation. In G. M. Fink (Ed.), *Choices and consequences: Contemporary policy issues in education* (pp. 29-58). Cornell University Press.

Moss, P. A., & Haertel, E. H. (2016). Engaging methodological pluralism. In D. Gitomer & C. Bell (Eds.), *Handbook of research on teaching* (pp. 127-247). American Educational Research Association.

National Academies of Sciences, Engineering, and Medicine. (2020). *Changing expectations for the K-12 teacher workforce: policies, preservice education, professional development, and the workplace.* The National Academies Press.

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

Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy Analysis Archives*, *18*, 23-35.

Papay, J. P., Taylor, E. S., Tyler, J. H., & Laski, M. E. (2020). Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data. *American Economic Journal: Economic Policy*, *12*(1), 359-88.

Preston, C. (2017). University-based teacher preparation and middle grades teacher effectiveness. *Journal of Teacher Education*, *68*(1), 102-116.

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

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

Ronfeldt, M., Bardelli, E., Mullman, H., Truwit, M., Schaaf, K., & Baker, J. (2020a). Improving student teachers' readiness to teach through recruitment of instructionally effective and experienced cooperating teachers: A randomized experiment. *Educational Evaluation and Policy Analysis*, *42*(4), 551-575.

Ronfeldt, M., Brockman, S. L., & Campbell, S. L. (2018a). Does cooperating teachers' instructional effectiveness improve preservice teachers' future performance? *Educational Researcher*, *47*(7), 405-418.

Ronfeldt, M., Farmer, S., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal*, *52*(3), 475-514.

Ronfeldt, M., Goldhaber, D., Cowan, J., Bardelli, E., Johnson, J., & Tien, C. D. (2018b). Identifying promising clinical placements using administrative data: Preliminary results from ISTI placement initiative pilot (CALDER Working Paper No. 189). National Center for Analysis of Longitudinal Data in Education Research.

Ronfeldt, M., Matsko, K. K., Greene Nolan, H., & Reininger, M. (2020b). Three different measures of graduates' instructional readiness and the features of preservice preparation that predict them, *Journal of Teacher Education*, *72*(1), 56-71.

Ronfeldt, M., & Reininger, M. (2012). More or better student teaching? *Teaching and Teacher Education*, *28*(8), 1091-1106.

Ronfeldt, M., Reininger, M., & Kwok, A. (2013). Recruitment or preparation? Investigating the effects of teacher characteristics and student teaching. *Journal of Teacher Education*, *64*(4), 319-337.

Ronfeldt, M., Schwartz, N., & Jacob, B. (2014). Does pre-service preparation matter? Examining an old question in new ways. *Teachers College Record*, *116*(10), 1-46.

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

Rozelle, J. J., & Wilson, S. M. (2012). Opening the black box of field experiences: How cooperating teachers' beliefs and practices shape student teachers' beliefs and practices. *Teaching and Teacher Education*, *28*(8), 1196-1205.

Sleeter, C. E. (2001). Preparing teachers for culturally diverse schools research and the overwhelming presence of Whiteness. *Journal of Teacher Education*, *52*(2), 94-106.

Stallings, J. A., & Kowalski, T. (1990). Research on professional development schools. In R. Houston, W. R. Haberman, & J. Sikiula (Eds.), *Handbook of research on teacher education* (pp. 251-263). Macmillan.

Steinberg, M. P., & Garrett, R. (2016). Classroom composition and measured teacher performance: What do teacher observation scores really measure? *Educational Evaluation and Policy Analysis*, *38*(2), 293-317.

Whitehurst, G. J. R., Chingos, M. M., & Lindquist, K. M. (2014). *Evaluating teachers with classroom observations*. The Brookings Institution.

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.

Youngs, P., Molloy Elreda, L., Anagnostopoulos, D., Cohen, J., Drake, C., & Konstantopoulos, S. (under review). The development of ambitious instruction: How beginning elementary teachers' personal characteristics and preparation experiences are associated with their mathematics and reading practices.

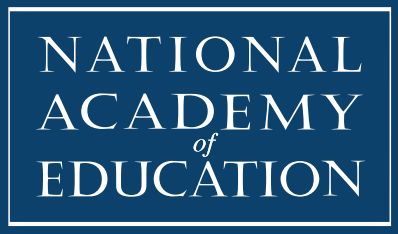
Youngs, P., & Qian, H. (2013). The influence of university courses and field experiences on Chinese elementary candidates' mathematical knowledge for teaching. *Journal of Teacher Education*, *64*(3), 244-261.

Zeichner, K. (2012). The turn once again toward practice-based teacher education. *Journal of Teacher Education*, *63*(5), 376-382.

Zeichner, K., & Gore, M. (1990). Teacher socialization. In R. W. Huston (Ed.), *Handbook of research on teacher education* (pp. 329-348). Macmillan.

AUTHOR BIOGRAPHY

Matthew Ronfeldt is an associate professor of educational studies at the University of Michigan. Ronfeldt seeks to understand how to improve teaching quality, particularly in schools and districts that serve marginalized students. His research sits at the intersection of educational practice and policy and focuses on teacher preparation, the development and support of teaching quality, teacher recruitment and retention (particularly teachers of color), teacher induction, and the assessment of teaching and preparation programs. Ronfeldt is primarily interested in whether and how pre-service teacher education and school/district factors, especially working conditions, are related to the employment, instructional quality, and retention of teachers. His current work mostly employs large-scale, quantitative methodologies, though he also engages in qualitative and mixed-methods research. Ronfeldt has served as a principal investigator (PI) or co-PI on a number of projects using large-scale, quantitative methods to assess and improve teacher education across four states. These projects have received funding from the Bill & Melinda Gates Foundation, the Institute of Education Sciences, The Joyce Foundation, and the Spencer Foundation. Ronfeldt received (with Shanyce Campbell) the 2019 Palmer O. Johnson Award for the outstanding article appearing in an AERA-sponsored journal. He also received the 2016 American Association of Colleges for Teacher Education Outstanding *Journal of Teacher Education* Article Award. Ronfeldt earned his Ph.D. from Stanford University, where he concentrated on teacher education. After receiving his doctoral degree, he completed two more years at Stanford as an Institute of Education Sciences postdoctoral fellow in the Institute for Research on Educational Policy and Practice (now the Stanford Center for Education Policy Analysis [CEPA]), focusing on large-scale quantitative research. Ronfeldt previously taught middle school mathematics and science for 7 years.



The **National Academy of Education (NAEd)** advances high-quality research to improve education policy and practice. Founded in 1965, the NAEd consists of U.S. members and international associates who are elected on the basis of scholarship related to education. The Academy undertakes research studies to address pressing educational issues and administers professional development fellowship programs to enhance the preparation of the next generation of education scholars.

naeducation.org