

Working Paper

Review of Research on Clinical Practice in Teacher Preparation: What Do We Know and How Can We Advance Knowledge?

H. Alix Gallagher
Naa Ammah-Tagoe

April 25, 2017

This paper was prepared by SRI International with funds provided by the Charles and Lynn Schusterman Family Foundation. However, the contents do not necessarily represent the viewpoints of the Charles and Lynn Schusterman Family Foundation.

Introduction

There is widespread understanding of the need to improve teacher preparation because many novices enter the profession lacking the knowledge and skills to teach to new, challenging standards such as the Common Core State Standards and Next Generation Science Standards. The purpose of this review was to synthesize research-based knowledge about effective preparation, with a specific focus on clinical practice.

Given this purpose, we made two decisions that shaped our inclusion of studies. The first was that studies needed to be of high quality. For empirical studies, this meant the researchers used a design and measures that supported the inferences made in the text and reported on the results transparently. The second was that any outcome measures in empirical studies had to be closely related to the goal of having a quality teacher workforce; we defined these as measures of teaching effectiveness (e.g., observational measures and/or value-added measures) or of teacher retention. Thus, we excluded studies that examined such outcomes as recent graduates' reported self-efficacy, which may matter in its own right but is relatively removed from the outcome of effectively teaching to Common Core Standards.¹ Finally, we note that this was not an exhaustive review. We attempted to describe key studies in enough depth to be informative about the strengths and weaknesses of the current knowledge base, recognizing that one trade-off in a short review was comprehensiveness.

The landscape of teacher preparation research

It is important to situate the discussion of what we know about effective clinical teacher preparation in the state of research on the topic. To find rigorous studies, we had to cast our net beyond studies solely of clinical practice (our initial charge) to those that addressed effective teacher preparation more generally. Even with that expansion, we found few large-scale and systematic reviews of research on teacher preparation. Those reviews found that the research base is surprisingly thin and that claims about practices that are “research based” are sometimes supported by insufficient evidence.

For example, in 2002 the Office of Educational Research and Improvement and the U.S. Department of Education asked experts to review high-quality research to answer five basic questions on effective teacher preparation.² The experts found only 57 peer-reviewed studies that addressed any of the five research questions and met their standard for “disciplined inquiry”—a pretty broad category that did not limit studies to a specific research methodology but rather to the quality with which the researchers carried out and reported on their methods. While the experts made some broad statements about teacher

¹ This meant excluding a few well-known studies, such as one by Darling-Hammond, Chung, and Frelow (2002). That study relied on a survey with a 33% response rate to support the claim that teachers who had experienced traditional preparation felt more equipped to enter teaching than those who entered through alternative routes. Further, while the abstract and text both say that nearly 3,000 teachers responded to the survey, the information that respondents represented about one third of the teachers surveyed is relegated to an endnote.

² One of the questions was, “What are the effects of student teaching?” Relevant to the topic of our review, the experts reported that most studies of student teaching addressed changes in attitude rather than acquisition of knowledge and skills, but that “focused and well-structured” clinical experiences are more effective (Wilson, Floden, & Ferrini-Mundi, 2002, p. 195).

preparation (e.g., teachers' content knowledge matters), their overall conclusion was that the research base on teacher preparation was weak (Wilson, Floden, & Ferrini-Mundy, 2002).

The National Research Council (NRC) (2010) undertook a similarly broad review of research on preparing effective reading, mathematics, and science teachers. The NRC found that there “is currently little definitive evidence that particular approaches to teacher preparation yield teachers whose students are more successful than others” (p. 174). The NRC noted that there were many descriptive studies of teacher preparation but that studies that attempted to link program characteristics with measures of teacher effectiveness “have generally been insensitive to the details of teacher preparation that are most likely to result in differences in quality” (p. 178).³

Furthermore, Zeichner and Conklin (2005) posited that studies of teacher preparation that focused largely on program structures, such as hours of clinical practice or number of credits, rather than on specific design features and substantive program elements have little potential to explain much of the variation in what graduates know and can do. For example, knowing how many hours a candidate works with a cooperating teacher may be less useful to understanding a program's impact than illuminating what the interactions between the candidate and cooperating teacher are like.

Taking these three reviews as a jumping off point for our own review, we consider Zeichner and Conklin's assertion in light of more recent research. Our goals are to share what we think research points to in terms of high-quality clinical preparation and to recommend a more productive path for future teacher preparation research.

Structure of this review

The results of this review are presented in five main sections, each of which homes in more on how to craft a research agenda to yield policy-relevant answers about effective teacher preparation. First, we review large-scale correlational studies, which provided the best evidence to date on which program features are related to better outcomes for teacher preparation programs. We then look at one promising model of teacher preparation: teacher residencies. We focused on this model both because of available studies and because there is an assumption that it is a better approach to teacher preparation, an assertion that we wanted to examine. Third, we examine efficacy studies of specific programs, with an eye toward detecting similarities in findings that suggest broader patterns in effective teacher preparation. When that approach yields little, we turn to theoretical and small-scale studies that build theory. Finally, we step back, look at the collection of research reviewed, and formulate an approach to narrowing the knowledge gap about teacher preparation over the longer term.

³ More recently, Mitchell & King (2016) released a critique of the state of teacher preparation. We have not included it in the main discussion because the reporting about its methodology was insufficient. However, its core claim does not contradict what more rigorous—although older—reviews found about the paucity of generalizable claims about effective teacher preparation, suggesting that the research base has not expanded dramatically in the past few years.

Large-scale correlational studies

Several studies have used large-scale data sets to correlate teacher characteristics and preparation (teacher test scores, majors, degrees, preparation route, etc.) with measures of teacher effectiveness. Cumulatively, these studies support the idea that teacher preparation programs and institutions can make a difference and that those differences are meaningful. In general, these studies support the conclusion that teacher preparation matters by showing that significant variation exists in teachers' effectiveness, most often measured by VAM (value-added measures), based on their personal characteristics or credentialing program.

We did not find any recent large-scale correlational studies that tested whether teacher preparation itself matters (i.e., comparing the teaching effectiveness of those who had preparation and those who had no preparation). This is probably because state policies ensure that an overwhelming proportion of public school teachers either have completed or are enrolled in a structured preparation program (including alternative certification programs). One exception is the study in which Goldhaber and Brewer (2000) compared the 12th-grade math and science performance of students whose teachers had probationary, emergency, private school, or no certification with the performance of students whose teachers had standard certification in their subject area. They found benefits of standard certification relative to private school or no certification in mathematics but no relationship between certification status and science performance and no difference between standard certification and emergency credentials in mathematics. Further, these certification categories are so broad as to provide an example of our overall point that existing data sets provide little useful description about the nature of teacher preparation programs.

Researchers have found relatively few significant relationships between measured characteristics and student achievement, however, suggesting the need for studies that sufficiently tie the substantive features of programs to outcomes. Some studies suggest that teachers' own academic backgrounds and content knowledge are related to student achievement. Wayne and Youngs (2003) found that for mathematics, high school students learned more from teachers who had certifications, degrees, and coursework related to mathematics. Findings in other subjects were not positive and conclusive. Wayne and Youngs also found a positive relationship between the ratings of teachers' undergraduate institutions or teachers' test scores and student outcomes but noted a need for research into the specific institutional characteristics and tested skills and knowledge that correlate with higher student achievement.

In more recent studies we see a constellation of significant findings, but they do not combine to form a rich and empirically defensible theory about which aspects of teacher preparation are related to effective teaching or student achievement. For example, Boyd et al. (2009) found relationships between teachers' VAM scores and the extent to which programs grounded student experience in practice (measured through proxies of whether or not the program required a capstone project, the extent of oversight of field experiences, and studying program curricula). Goldhaber, Liddle, and Theobald (2013) found variation attributable to credentialing institutions in the impact of graduates on student achievement in reading, although not in math. Their analysis accounted for variation in institutional

selectivity but did not distinguish other characteristics of institutions. Goldhaber and colleagues estimated that the difference in effectiveness of teachers credentialed by various institutions was greater than the difference in effectiveness between a first-year teacher and a veteran with 5 or more years of experience. More recently, Ronfeldt and Campbell (2016) found that the difference between graduates from the top and bottom quartiles of Tennessee's preparation institutions and programs was the equivalent of an additional year of initial teaching experience on observational measures of teaching practice. In contrast, using a large and diverse data set from Texas to evaluate statistical methods for estimating differences in teacher quality for teacher preparation programs, von Hippel et al. (2016) found that even the best method, based on a value-added model, is rarely able to distinguish which programs are better or worse than average. Finally, Goldhaber, Krieg, and Theobald (2016) found that teachers are more effective, as measured by VAM, when the student demographics of their student teaching placement match those of their current school.

These correlational studies can apportion variance, that is, they can obtain estimates of the variance in teaching quality attributable to differences in preparation programs as opposed to other sources of variation. Yet the variables in these data sets explain relatively little about what features of teacher preparation programs are related to the variation in estimated program quality. This is possibly because of the bluntness of the measures of teacher characteristics or preparation. For example, in studies comparing the teaching effectiveness of those who completed traditional versus alternative certification programs, there is both substantial *variation within* the categories "traditional" and "alternative" in the nature of the programs and substantial *overlap across* categories in terms of candidates' experiences and requirements for credentials (NRC, 2010; Zeichner & Conklin, 2005). Similarly, whether or not a teacher has a master's degree in mathematics is potentially less related to teaching effectiveness than the content of the courses taken and whether those courses prepare the teacher to unpack rich mathematical knowledge for students. Without substantially more nuanced measures of teachers and their programs—including more information on program content and opportunities for candidates to learn specific pedagogical skills—these types of studies will probably not yield substantially better information in the future for two related reasons. First, the measures of programs that typically exist in extant data sets are too blunt to capture potentially important differences in candidates' opportunities to learn. Second, the measure of outcomes by which to judge program effectiveness in extant data sets is often student achievement scores. These are distal from teacher preparation programs and so are less sensitive than direct measures of teaching effectiveness would be to variations in graduates' quality. One way to address at least the problem of blunt measures of programs is to study specific program models with designs that attend to nuances in program features.

Studies of a specific preparation model: Residencies

While large-scale correlational studies often include some comparison across models of traditional and alternative teacher preparation, researchers are also studying increasingly prevalent models such as teacher residency programs. Teacher residency programs (TRP) hold the promise of attracting, preparing, inducting, and retaining more effective teachers because they require closer collaboration between universities and clinical placement sites

than is typical in traditional preparation models. We are starting to gain a good understanding of what TRP look like, but we know less about how they work and their ultimate impact on teaching effectiveness.

Most of the recent empirical studies describing the structure, features, and stakeholders of TRP have drawn conclusions from sample sizes too small to offer generalizability. One Institute of Education Sciences-funded study by Silva et al. (2014) does provide a comprehensive empirical description of key features of TRP. The first broad descriptive study of TRP, it provides implementation findings on 30 programs that received Teacher Quality Partnership residency grants awarded in fall 2009 and spring 2010. It also includes in-depth information on residents, mentors, TRP novice teachers, and non-TRP novice teachers for a subset of the 12 largest and most experienced programs. The study did not include an analysis of student achievement results, but both the initial report and the follow-up brief (Silva et al., 2015) present findings on retention of TRP teachers compared with non-TRP teachers. Regarding clinical experiences, Silva et al. (2014) found that Teacher Quality Partnership residency programs provided candidates fieldwork with experienced and trained mentors, integrated coursework, and substantial and increasing instructional responsibility over the course of the residency year, in keeping with major findings of qualitative studies of residencies. In addition to describing the residency lengths, structures, and placements, the authors examined reported instructional responsibilities and focus areas of interactions between residents and their mentors, other teachers, and fellow residents. Combined with expert opinions on what the field thinks matters in clinical practice, these descriptive analyses could suggest areas for future research on clinical experiences, as discussed at the end of this review.

Although such descriptive detailing of the residency landscape is important, the empirical literature does little to explain what makes TRP effective or even if they are more effective in preparing teachers than traditional models. The qualitative literature that could provide a better explanation of how TRP impact teaching effectiveness is dominated by small-scale studies outside our inclusion criteria. However, some high-quality case studies have been conducted by or in conjunction with teacher residency program organizations. Among these, the National Center for Teacher Residencies (NCTR) recently reported on the landscape, structures, supporting conditions, and challenges TRP face as the teacher preparation field shifts toward an emphasis on clinical experiences. TRP across the country are rethinking the nature of clinical training by positioning teacher candidates as co-teachers, emphasizing candidate performance and accountability through competency-based assessments and the use of district or state-aligned evaluation tools, increasing mentor selectivity, concentrating on mentor development, and devising new, clinically based roles to accommodate programmatic changes (NCTR, 2015). These qualitative studies offer an initial description of how residencies prepare teachers.

On the question of impact, a small body of empirical research has addressed TRP's contributions to teacher retention, with findings in two areas: whether graduates are retained and where and whom they teach.⁴ In a comprehensive review, Zumwalt & Craig

⁴ Teacher preparation programs are arguably expected not only to prepare effective teachers, but also to prepare teachers who will remain in the profession to continue raising student achievement. TRP often explicitly aim to improve teacher retention in their communities.

(2005) found that prior literature offered mixed results on whether graduates of alternative certification programs, including residencies, remained in their initial district or the profession for longer than traditionally prepared teachers. More recently, Papay et al. (2012), studying a specific TRP, found that Boston Teacher Residency graduates were more likely than similarly experienced peers to stay in the district in their fifth year and that attrition did not rise notably after the end of the 3-year commitment. Similarly, in the follow-up brief to their 2014 study of residencies, Silva et al. (2015) analyzed retention among third- and fourth-year teachers and found that graduates of Teacher Quality Partnership TRP were more likely to remain in the same district than non-TRP teachers with similar initial teaching placements, while retention in the same school was similar. Their results suggested that the difference in district retention emerged between TRP and non-TRP teachers' second and third years of teaching. Although these two studies did not explain the mechanism, they suggest that residencies can contribute to producing teachers who will stay in their initial placement district longer than teachers trained in other programs.

Given interest in placing effective teachers in schools with traditionally underserved or underperforming populations, residencies often attend to whom their graduates teach. Hence, a second question of interest is whether residency graduates choose to continue teaching in schools with the highest needs. Evidence indicates that the answer depends on the measure of needs. Silva et al. (2015) found that in Teacher Quality Partnership residency partner districts, TRP graduates who moved to different schools in the same district tended to join ones where similar proportions of students were from low-income families. In line with other teacher retention literature, Silva et al. (2015) also found that teachers prepared by residencies changed to schools with a lower proportion of black students and higher achievement. These findings suggest the need for further research to confirm which populations benefit from residency teachers retained in the field and what factors influence these teachers' decisions to change schools.

Collectively, the studies described in this section provide an initial understanding of how teacher residency programs fulfill their promise to attract, prepare, induct, and retain teachers. Yet there is still a need for mixed-methods research to describe the current features of residency programs and then quantify and explain how these programs impact teaching effectiveness.

Studies of specific programs

Another way to accumulate knowledge about effective teacher preparation and clinical practice is by synthesizing well-designed studies of specific programs to look for patterns in results. One challenge in interpreting these sets of studies, as with studies on the residency model overall, is that their design does not enable researchers to distinguish the impact of any program features (i.e., how the program selects versus trains teacher candidates) on outcomes. The inability to distinguish the impact of selection from the design of the training embedded in the program is important to highlight for two reasons. First, selection and training can both be resource-intensive aspects of programs, so providing guidance on how to allocate resources would be beneficial. Additionally, the inability to isolate the effects of selection could indicate that whatever is effective in any given program does not generalize beyond the specific population the program recruits. As

a result, the program might not offer larger lessons on how to prepare the teacher workforce writ large. That said, given research described above that shows correlations between candidate characteristics and teaching effectiveness, recruitment and selection into programs are arguably important program characteristics (and for programs like Teach For America, a primary strategy) for ensuring highly effective teachers enter the workforce. Additionally, the fact that instructional aspects of the program cannot be disentangled from each other makes it more challenging to draw broader lessons on what makes programs effective. As a result, while these studies provide important information about the effectiveness of teacher preparation programs, they do not provide insights into exactly what it is about the programs that makes them more or less effective.

Similar to the correlational studies of teaching effectiveness discussed above, a set of recent program evaluations found few significant relationships between measured characteristics and student achievement. For example, in the most extensive examination of a residency program and teaching effectiveness, Papay et al. (2012) found that Boston Teacher Residency graduates teaching math outperformed veteran teachers by their fourth or fifth year of teaching. However, they found no evidence that Boston Teacher Residency graduates were more effective in their first year of teaching. Focusing on secondary math, Clark et al. (2013) found in an experimental study that TNTP Teaching Fellows were no more or less effective than other teachers teaching the same math courses in the same schools.

Evaluations of Teach For America have shown mixed although generally positive results. Here, we describe two of the more recent studies.⁵ In the parallel experimental study to the Teaching Fellows evaluation, Clark et al. (2013) found Teach For America corps members were more effective than the comparison teachers in teaching secondary math in the same schools. Building on this work in a 2015 evaluation of Teach For America's Investing in Innovation scale-up, the authors found that on average Teach For America teachers were as effective as comparison teachers in both reading and math and that lower elementary Teach For America teachers had a significant positive effect on reading (Clark et al., 2015). However, they found no significant impacts for Teach For America secondary math teachers, which was inconsistent with their previous findings.

Most large-scale studies that estimate the effectiveness of teacher preparation programs use VAM as the outcome measure. Ronfeldt and Campbell (2016) made an important contribution to the empirical literature by using direct measures of graduates' instructional quality. Evaluating preparation programs using graduates' observational ratings, Ronfeldt and Campbell found that observational ratings can measure differences in program performance and support robust classification of programs into quartiles. Further, they found that programs' rankings based on observational ratings are positively and significantly related to rankings based on graduates' value-added scores. The results were based on four observational rubrics used by Tennessee school districts: the Tennessee Educator Acceleration Model (TEAM, an adaptation of the National Institute for Excellence and Teaching's TAP rubric), Project COACH, Teacher Effectiveness Measure (TEM), and Teacher Instructional Growth for Effectiveness and Results (TIGER).

⁵ There is a much larger set of studies of Teach For America, but we have left them out for the sake of parsimony as adding them does not change our conclusions or recommendations.

Ronfeldt and Campbell proposed observational ratings as a complement to VAM in measuring the performance of teacher preparation. There are several advantages to this approach including availability of observational data and direct measurement of instruction. Perhaps most important, although Ronfeldt and Campbell did not conduct such analyses in this initial paper, the approach of using observational measurements provides the opportunity to gather detailed information that could be more readily tied to substantive program features and therefore could prove useful for program improvement. For example, some observational measures include ratings of specific teaching practices (e.g., questioning), which could enable programs to assess the strengths and weaknesses of their candidates and graduates on specific skills.

Looking across this set of studies on specific programs does little to advance the broader knowledge base about why programs are more or less effective because the evaluations were not designed to disentangle the effects of program features, and the studies that relied on large-scale data sets typically had blunt measures of teacher preparation. On the other hand, many of the small-scale descriptive studies that dominate the literature are problematic for our purposes because they may focus on outcomes that are too distantly connected to student achievement (e.g., candidate self-confidence at the conclusion of their preparation program) and/or because the sample size was too small to support generalizable statements. However, well-conducted small-scale studies can build theory in ways that could yield testable hypotheses and understandings supported by qualitative evidence about teacher preparation. We turn next to a few small-scale studies and texts by experts in the field as sources of understanding what might be effective in teacher preparation. These studies help us conceptualize what teachers need to know to teach effectively and suggest that clinical practice is likely key for developing effective educators.

Toward a theory of how to structure effective clinical practice in teacher preparation

As described, the research base on the effectiveness of teacher preparation programs and approaches to clinical practices gives little guidance on how best to design teacher preparation. However, there is a trove of high-quality small empirical studies and theoretical papers that collectively support a few important assertions, which have implications for ultimately building strong evidence about effective ways to design and implement teacher preparation programs. Given the design of these studies, they do not support generalizable claims about effective approaches teacher preparation or clinical practice. However, they are very suitable for theory development.

Teaching is a cognitively complex and interpersonal profession. Because of the interpersonal demands, clinical practice is critical to learning to teach.

Shulman (1987) wrote a foundational paper classifying different types of knowledge necessary for effective teaching. Refined over time (e.g., Ball, Thames, & Phelps, 2008), this work made an enormous contribution by building a trajectory of work that has helped explicate the complexity of expert teaching practice. Critically, this line of research makes clear that expert teaching requires not only deep content knowledge and specialized knowledge to develop students' knowledge of content, but also skills to manage content within the interpersonal relationships in a classroom (Lampert, 2010). Because

orchestrating the interactions among individuals in a classroom is core to teaching and is also an applied skill, it follows logically that those learning to teach need to practice these skills in addition to acquiring knowledge and theoretical understanding about content and instruction. This represents an important evolution of Shulman’s framework in that the field of teacher education currently embraces the idea that pedagogical skills and routines—as opposed to just “knowledge”—are most central in what candidates need to learn because they embody the actual “practice” of teaching (Ball & Forzani, 2009; Lampert, 2010; Lampert, Beasley, Ghouseini, Kaemi, & Franke, 2010). The fact that teaching requires the deployment of complex skills as well as knowledge in a complex interpersonal environment puts designing opportunities for clinical practice at the center of thinking about teacher education.

Traditional approaches to clinical practice lack the consistency and scaffolds necessary for teacher preparation to consistently yield well-prepared novices.

In traditional teacher preparation programs, teacher candidates take courses to build their knowledge and skills, and their clinical experiences occur largely during student teaching. However, teacher preparation programs typically have relatively few mechanisms (e.g., extensive influence over mentor selection and training) to ensure consistency in the student teaching experience (Hatch & Grossman, 2009). While student teaching, candidates interact most frequently with their cooperating teachers, whom they report have a large impact on their learning experiences. Candidates also interact with supervisors, typically connected to the university, who evaluate their progress. A small-scale study of nine student teachers found enormous variations in their experiences based on a range of factors outside the control of the teacher preparation program (Valencia, Martin, Place, & Grossman, 2009). The authors noted, “[W]hat the [teacher candidates] were able to practice and, consequently, what they learned varied dramatically depending on the instructional organization, classroom norms, curriculum materials, and cooperating teacher’s stance unique to each classroom and site” (p. 313). The authors also found that very little of the feedback teacher candidates received during their clinical experience was on how to teach their content—arguably a critical aspect of effective teaching—with classroom management instead being the main focus. The traditional structures of clinical practice in teacher preparation yield idiosyncratic opportunities for candidates to learn key skills for teaching content.

Grossman et al. (2009) took a different deep dive into clinical practice in teacher preparation comparing how novices are trained in three professions—the clergy, clinical psychology, and teaching. All three require not just content knowledge and technical skills, but also the ability to handle the multiple and complex demands of applying the knowledge and skills in the context of relationships. Using interviews, observations, and focus groups in eight preparation programs across the three professions, the authors developed a framework to describe three types of learning experiences in clinical preparation:

- Representations—how practice is represented, and what different representations make visible about practice.
- Decomposition—how practice is broken down so it can be taught to novices.

- Approximations of practice—opportunities to engage in simulations of practice that are less or more authentic simulations (i.e., more or less scaffolded) of full practice.

This framework describes varying levels of scaffolding for teaching candidates core pedagogical knowledge and skills. Grossman et al. noted that the teacher preparation programs they studied offered candidates fewer opportunities to engage in approximations of practice on the most complex relational aspects of teaching (e.g., those that require responding to or interacting with students) than the programs in other fields offered their candidates.

We found that in comparison with our other two professions, teacher education provides multiple approximations of various aspects of preactive practice in teaching; novices are asked to engage in simulated lesson planning, unit planning, even planning for classroom management. However, they encountered many fewer opportunities in the context of coursework to engage in approximations of interactive practice— such as how to respond to a student’s question or orchestrate a discussion—than did novices in clinical psychology. Because many of the most difficult aspects of teaching lie in these interactive dimensions of practice, novice teachers may be losing valuable opportunities to hone their skills in these areas. (pp. 2094–2095)

In addition to the challenges of providing students the opportunity to learn how to handle the interpersonal components of teaching, Borko and Livingston (1989) focused on the cognitive aspects of learning to teach. They studied three pairs of student teachers assigned to expert cooperating teachers and examined the differences in their experiences of instruction. They noted that the traditional student teaching experience was cognitively overwhelming for student teachers because they lacked the cognitive frameworks for quickly and successfully making sense of the vast quantities of information they needed to process to plan and deliver instruction. Borko and Livingston came to the conclusion that novices are so overwhelmed by the cognitive demands that they are not fully able to learn the routines and strategies of their expert mentors and may not even be able to learn effectively from their own experiences. Implicitly, this argument suggests that student teachers’ learning is varied not only for the reasons Valencia Martin, Place, and Grossman (2009) articulated, but also because of the unevenness in what individuals will successfully process from the varied opportunities to learn in student teaching. Borko and Livingston suggested that student teachers’ clinical experiences be more structured, with a more gradual release of responsibility, so that student teachers can “develop and elaborate knowledge structures for teaching and pedagogical reasoning skills” (p. 39).

Taken together, these studies suggest that traditional clinical experiences offer idiosyncratic learning opportunities for teacher candidates and that those opportunities insufficiently address the intersection between content and the pedagogical skills necessary to support candidate learning. However, the authors are hopeful that clinical experiences could be structured differently to achieve better results by strategically scaffolding candidate learning.

Different pedagogies for teacher preparation have different limitations and affordances. These different approaches could be strategically sequenced to provide better scaffolding for learning to teach.

McDonald, Kazemi, and Kavanagh (2013) presented a cycle by which candidates could be exposed to new instructional strategies, then gain increasingly more independent practice with them—thus scaffolding clinical practice—and finally reflect on their experience. They identified a four-stage cycle, which aligns closely with Grossman and colleagues' framework (2009), and particular pedagogies of teacher preparation that fit best in each stage:

1. **Introducing and Learning about the Activity**—This stage provides representations of practice chosen to best illustrate the particular strategy (e.g., modeling, examining video exemplars, examining written cases).⁶
2. **Preparing for and Rehearsing the Activity**—In this stage, candidates experience approximations of practice, i.e., carefully scaffolded experiences that provide opportunities to try out parts of the activity without the full cognitive load present when teaching a class (e.g., collaborative planning, microteaching, rehearsal).
3. **Enacting the Activity with Students**—This stage gives candidates an opportunity to enact practice with students with more or less support (e.g., co-teaching, live coaching).
4. **Analyzing Enactment and Moving Forward**—In this final stage of the cycle, candidates reflect to support learning from their experience enacting (e.g., analysis of video or transcripts of practice enactments of instruction, reflection writing).

This framework is helpful in thinking about which types of activities provide the most appropriate clinical practice at various points in a candidate's processes of learning new pedagogical skills.

Studies have also investigated particular approaches to clinical practice that could be situated within McDonald, Kazemi, and Kavanagh's (2013) cycle. For example, Hatch and Grossman (2009) discussed the affordance of video in making visible specific aspects of teaching and enabling candidates and teacher educators to collectively and repeatedly view and analyze practice. Video also has the affordance of giving the teacher preparation program, as opposed to the cooperating teacher, control over the instruction viewed and the frequency and nature of the discussion of practice that follows. Video could be used in either the first or fourth stage of McDonald and colleagues' (2013) cycle.

Lampert et al. (2010) discussed how they designed "instructional activities," which are participation structures that specify how students and teachers interact with each other on particular content. They argued that instructional activities can be used to help candidates learn to implement high-leverage instructional practices by reducing the cognitive load on novices in orchestrating the interpersonal components of instruction. Instructional activities could be used in stages 2 and/or 3 of McDonald, Kazemi, and Kavanagh's (2013) cycle. Foreshadowing the conclusion of this review, we think that McDonald, Kazemi, and Kavanagh's framework (or some similar framework) could play a key role in identifying meaningful aspects of clinical practice in teacher preparation programs.

⁶ One key strategy for helping novices or candidates in this part of the cycle might be decomposition, as described by Grossman et al. (2009).

Finally, the field would benefit from a consensus on which knowledge and skills novice teachers should have.

Cutting across much of this research is the theme that teacher preparation tends to be idiosyncratic. This is partially because of a program design where key clinical experiences are largely delegated to a vast array of cooperating teachers and also because of a lack of consensus on a concise list of high-leverage skills that all novices need to acquire. Given the extent of knowledge and skills expert teaching requires (Ball & Forzani, 2009; Lampert et al., 2010; Shulman, 1987) and the fact that it is unreasonable to expect novices to acquire them all in the relatively short time devoted to their preparation (Borko & Livingston, 1989), it may be understandable that the field has not coalesced around baseline skills required for entry. Ball and Forzani (2009) have emphasized the importance for teacher preparation of coming to agreement on “high leverage” practices, but this has not been widely heeded. Possible reasons include state policies, some of which require preparation programs to cover a much broader array of teaching skills (for example, see the California Teacher Performance Expectations <http://www.ctc.ca.gov/educator-prep/standards/adopted-TPEs-2013.pdf>.) For the purposes of designing a research study, differences in focal knowledge skills across programs is another key source of variation in preparation experiences to consider, along with variation in how clinical practice is structured to scaffold candidate learning.

What do we know, and where do we go from here?

Most of the research on teacher preparation provides little guidance on how to design preparation programs—including clinical experiences—to best prepare teachers to teach effectively to the Common Core State Standards and other rigorous learning goals for all students. Metaphorically, it is like the night sky, with many points of light disconnected from each other. We know that teacher preparation matters, but we cannot attribute much of the variation we see in teaching quality to the characteristics of candidates or programs. We also see inconsistency in the relationship between specific well-regarded programs and the full range of outcomes they would like to affect.

The greatest strength of the research base is in very thoughtful pieces by experts whose theoretical and empirical work, as well as practical experiences in teaching and preparing teachers, forms the basis for important hypotheses about what effective teacher preparation might look like. It is clear that teaching is an applied field, with specific and teachable knowledge and skills; this in turn makes it safe to assume that clinical preparation is key. Given the importance of content and pedagogical knowledge for successful teaching, however, it is also highly likely that improving the coherence between coursework and clinical experiences would be beneficial. It also should be apparent and seems critical that programs and ultimately the field as a whole decide what knowledge and skills are necessary for entry into the profession and develop a rich practice in consistently helping teaching candidates acquire those skills. From expert knowledge of teaching, we gain insights into some strategies for accomplishing this.

The next step is to turn these insights into hypotheses and then test them systematically. The goal needs to be to do more than just put forth a few more high-quality

studies; it needs to be to connect studies into a coherent and tested framework about effective teacher preparation.⁷

A starting point for developing and testing a framework for teacher preparation is to answer questions about (a) what we believe novice teachers need to know and know how to do, (b) how we think candidates learn those knowledge and skills, and (c) how to measure program outcomes. It is critical that we include not only the process of teacher learning (i.e., structures)—which has dominated research to date—but also the substance of teacher preparation (i.e., knowledge and skills) because the field currently lacks consensus on both the *what* and the *how* of teacher preparation. This vision of a research approach shares similarities with a call for a framework for studying teacher professional development, where the field has reached at least an initial consensus on key structures for supporting teacher learning (Desimone, 2009). Desimone also notes the importance of a consensus framework for developing an appropriate timeline for the yields of high-quality teacher preparation. For example, research on the Boston Teacher Residency suggests that the benefits of its TRP for student learning are most apparent after a few years, perhaps when initial variation in classroom management skills between graduates from its and other preparation programs wash out and the benefits of the TRP approach to teaching pedagogical knowledge and skills become evident. Setting a timeline of impact suggests the need to measure both more proximal instructional quality outcomes along with the more distal student outcomes that are the ultimate goal of teacher preparation. With a consensus framework, a range of designs could be used to either attempt a large-scale study or to ensure that a series of smaller, program-level studies could be combined to test and refine the theory about effective teacher preparation. A comprehensive consensus framework would ensure some degree of consistency in key measures of programs and outcomes across studies to facilitate the development of a knowledge base in the field—which to date has been elusive.

The conclusion of this review is that the basis for a consensus framework may exist even though it has been insufficiently developed and tested to date. Building such a framework would better position upcoming research to build a knowledge base to effectively answer questions about how to design effective teacher preparation programs.

⁷ One other gap in the research base identified by the National Research Council (2010) is the lack of a consistent national data set describing teacher preparation. This is problematic in multiple ways. Among the information we searched for was a basic description of the nature of clinical practice experienced by teacher candidates. We found an older ERIC synthesis on early field experiences, which included some studies of apparently lesser methodological quality, but we found nothing that would enable a researcher to correlate meaningfully detailed descriptions of teachers' preparation experiences with extant measures of outcomes (Huling, 1998).

References

- Ball, D. L., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education, 60*(5), 497–511.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389–407.
- Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal, 26*(4), 473–498.
- 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.
- Clark, M. A., Chiang, H. S., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013). *The effectiveness of secondary math teachers from Teach For America and the Teaching Fellows programs*. Princeton, NJ: Mathematica Policy Research.
- Clark, M. A., Isenberg, E., Liu, A. Y., Makowsky, L., & Zukiewicz, M. (2015). *Impacts of the Teach for America Investing in Innovation scale-up*. Princeton, NJ: Mathematica Policy Research.
- Commission on Teacher Credentialing. 2013. *California teacher performance expectations*. Available at <http://www.ctc.ca.gov/educator-prep/standards/adopted-TPEs-2013.pdf>
- Desimone, L. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–189.
- Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis, 22*(2), 129–145.
- Goldhaber, D., Krieg, J., & Theobald, R. (2016). *Does the match matter? Exploring whether student teaching experiences affect teacher effectiveness and attrition*. CALDER Working Paper 149. Washington, DC: AIR (American Institutes for Research).
- Goldhaber, D., Liddle, S., & Theobald, R. (2013). The gateway to the profession: Assessing teacher preparation programs based on student achievement. *Economics of Education Review, 34*, 29–44.
- 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.
- Hatch, T., & Grossman, P. (2009). Learning to look beyond the boundaries of representation. *Journal of Teacher Education, 60*(1), 70–85.
- Huling, L. (1998). *Early field experiences in teacher education*. ERIC Digest. Washington, DC: ERIC Clearinghouse on Teaching and Teacher Education.
- Lampert, M. (2010). Learning teaching in, from, and for practice: What do we mean? *Journal of Teacher Education, 61*(1-2), 21–34.
- Lampert, M., Beasley, H., Ghouseini, H., Kazemi, E., & Franke, M. (2010). Using designed instructional activities to enable novices to manage ambitious mathematics teaching. *Instructional Explanations in the Disciplines* (pp. 129–141). Springer US.

- McDonald, M., Kazemi, E., & Kavanagh, S. S. (2013). Core practices and pedagogies of teacher education a call for a common language and collective activity. *Journal of Teacher Education, 64*(5), 378–386.
- Mitchell, A. L., & King, M. S. (2016). *A new agenda: Research to build a better teacher preparation program*. Bellwether Education Partners. Available at <http://bellwethereducation.org/>
- National Center for Teacher Residencies. (2015). *Clinically oriented teacher preparation*. Chicago, IL: Author.
- National Research Council. (2010). Division of Behavioral and Social Sciences and Education, Committee on the Study of Teacher Preparation Programs in the United States, Center for Education. *Preparing teachers: Building evidence for sound policy*. Washington, DC: National Academies Press.
- Papay, J. P., West, M. R., Fullerton, J. B., & Kane, T. J. (2012). Does an urban teacher residency increase student achievement? Early evidence from Boston. *Educational Evaluation and Policy Analysis, 34*(4), 413–434.
- Ronfeldt, M., & Campbell, S. L. (2016). Evaluating teacher preparation using graduates' observational ratings. *Educational Evaluation and Policy Analysis, 38*(4), 603–625.
- Silva, T., McKie, A., & Gleason, P. (2014). *Teaching residency programs: A multisite look at a new model to prepare teachers for high-need schools*. NCEE 2015-4002. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Silva, T., McKie, A., & Gleason, P. (2015). *New findings on the retention of novice teachers from teaching residency programs*. NCEE Evaluation Brief. NCEE 2015-4015. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–23.
- Valencia, S. W., Martin, S. D., Place, N. A., & Grossman, P. (2009). Complex interactions in student teaching lost opportunities for learning. *Journal of Teacher Education, 60*(3), 304–322.
- Von Hippel, P. T., Bellows, L., Osborne, C., Lincove, J. A., & Mills, N. (2016). Teacher quality differences between teacher preparation programs: How big? How reliable? Which programs are different? *Economics of Education Review, 53*, 31–45.
- Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*(1), 89–122.
- 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.
- Zeichner, K. M., & Conklin, H.G. (2005). Teacher education programs. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education* (pp. 645–736). Routledge.
- Zumwalt, K., & Craig, E. (2005). Teachers' characteristics: Research on the indicators of quality. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on on research and teacher education* (pp. 645–736). Routledge.