

# A New Agenda:

Research to Build a Better Teacher Preparation Program

Ashley LiBetti Mitchel and Melissa Steel King



**BELLWETHER**  
EDUCATION PARTNERS  
IDEAS | PEOPLE | RESULTS



## Table of Contents

Introduction	2
Research on Inputs Provides Little Guidance for Program Design	4
Outcomes-based Efforts Are on the Right Path, But Can't Yet Drive Program Improvement	9
Future Research Must Be Useful for Program Design	11
Recommendations	15
A New Agenda for Teacher Preparation Research	19
Endnotes	20
Acknowledgments	23

## Introduction

*Every year, new teachers collectively spend about \$4.8 billion on their training requirements—but it's unclear how well that money is spent.*

Every year, new teachers collectively spend about \$4.8 billion on their training requirements, nearly all of which goes to teacher preparation programs.<sup>1</sup> Unfortunately, it's unclear how well that money is spent. As a field, we've known for nearly half a century that teachers are the most important in-school factor affecting student achievement.<sup>2</sup> Yet despite decades of research and ongoing calls to improve the quality of teacher training, we still don't know how to sculpt an effective teacher preparation program.<sup>3</sup>

And it's not for lack of trying. For much of the history of teacher preparation, programs and policymakers assumed they could mandate the perfect cocktail of structural inputs, like the number of content-area courses, pedagogical courses, and student teaching hours, to guarantee a high-quality program. Over the years, policymakers and preparation programs designed training experiences using various combinations of these inputs. Teacher preparation research followed suit, examining the relationship between required inputs and completer performance to better understand what makes a program effective.

These efforts produced a good deal of research on teacher preparation inputs, but very few answers that could inform program design. That's because to date, most of this research has focused solely on broad, structural aspects of preparation programs. Research shows, for example, that content knowledge is important, but it's largely silent about course structure, types of content, and instructional methods—or any of the other

*Program design decisions are largely driven by what seems promising—rather than empirical research—because there is so little guiding evidence.*

specifics that determine program quality. Rather than drawing on empirical evidence to guide program design, the preparation field has relied heavily on theory or ideology. Even now, program design decisions are largely driven by what *seems* promising because there is so little guiding evidence.

This lack of evidence has led some in the field to advocate for a different approach to improving teacher preparation: outcomes-based accountability. Outcomes-based approaches attempt to obviate this traditionally intense focus on inputs. Instead of prescribing what a program should look like at the outset, outcomes-based systems link preparation programs to their completers' performance after they enter the classroom. In theory, collecting completer performance data allows policymakers to differentiate program quality and draw conclusions about appropriate interventions.

As with inputs, the outcomes-based approach presents its own challenges for improving design. Recent research has been unsuccessful in differentiating programs based on completer outcomes—the majority of programs produce completers who are comparatively similar, at least in terms of effectiveness in the classroom. If a state can't differentiate programs based on quality, then it also can't hold a program accountable or target interventions based on quality, either.

This history leads us to the current moment, where there's no clear right answer about how to improve teacher training; the two loudest arguments—focus on input-based standards vs. focus on completer outcomes—have very little evidence suggesting that either approach actually improves the quality of teacher preparation. Does that mean that neither approach is effective? Not necessarily. It's more likely that the current research doesn't ask the right questions, and so can't provide the right answers. State and federal policymakers, researchers, and programs all bear some responsibility for the current state of research.

*Rigorous research is at the heart of how to improve teacher preparation.*

Rigorous research is at the heart of how to improve teacher preparation. To create that body of research, the field needs **systems** that link completer performance data to preparation programs, make those data publicly accessible, and maintain individual privacy; **research methods** that use those data to produce actionable strategies and effective practices to improve program design; and **policies** that incentivize programs to evaluate the effectiveness of their model and adopt new, evidence-based practices. The end result should be a body of rigorous research that explores a multitude of possible improvement strategies, testing which components of program design are effective—specifically how effective, for whom, and under what circumstances. Until those pieces are in place, the quality of teacher preparation will remain stagnant.

## Research on Inputs Provides Little Guidance for Program Design

For much of the past 50 years, conversations about teacher preparation have focused on candidate and program inputs. Teacher preparation policy, like law and medicine,<sup>4</sup> required certain inputs as a way to ensure candidates and programs could meet a minimum bar of quality. These requirements emerged from the desire to professionalize teaching—which had long been viewed as semi-skilled female labor—as well as a desire to establish a common set of attitudes and dispositions among teachers. These requirements also reflected a different set of priorities: The goal of teacher preparation, for much of its history, was to provide candidates with a theoretical foundation for teaching and learning.<sup>5</sup>

With the advent of standards-based accountability in the 1990s, however, policymakers increasingly came to believe that teacher preparation programs should produce teachers who are effective in improving student achievement. This—along with the surge of alternative route programs that provided an accelerated path to teaching—stimulated new research seeking to understand the relationship between mandated teacher credentials and student learning outcomes. This research, which linked inputs to teacher performance to isolate the connection between the two, showed that the commonly mandated inputs are largely irrelevant to teaching quality. And because few, if any, of these inputs are backed by research, any shifts to or from specific inputs are primarily ideological, and anecdotal at best.

*Few of the inputs thought to improve teacher preparation are backed by research, so any shifts to or from specific inputs are primarily ideological, and anecdotal at best.*

## Candidate Inputs

Research on candidate inputs asks the question: Which candidate characteristics are associated with better-performing graduates, where “performance” is largely measured by student growth? Findings from these studies have been used to shape preparation program admissions processes, under the theory that programs can increase the quality of candidates by selecting for certain characteristics. In particular, this body of research homed in on three commonly studied candidate characteristics:

- SAT or ACT score
- High school or undergraduate GPA
- Selectivity of undergraduate institution attended by candidate

Research on these candidate inputs suggests that teachers who have higher SAT scores, have a higher undergraduate GPA, or attended a more selective institution<sup>6</sup> are more likely to become more effective teachers. This research has led to calls<sup>7</sup> for preparation programs to have higher admissions requirements.

However, this research is of limited usefulness for improving program design. Although prior academic performance may predict eventual effectiveness, it is one factor among many. For example, a 2015 study of a large Southern school district found that undergraduate GPA and attendance at a competitive college account for less than 2 percent of the variation between teachers.<sup>8</sup> Focusing on candidate inputs obscures the fact that it is likely a combination of variables—the candidates’ natural aptitudes and traits as well as the quality of training the program provides—that make for an effective teacher. Studies that focus only on candidate-related variables, or that lump together graduates with similar characteristics but widely different preparation experiences, cannot provide a complete picture of what effective training looks like. And as a practical matter, although programs may choose to set minimum academic requirements, the sheer volume of teachers needed at any one time—about 3.5 million, more than any other job that requires a bachelor’s degree<sup>9</sup>—coupled with the fact that teacher salaries are lower than most other professions that require a BA, makes it impossible for programs to select *only* the highest-achieving candidates, particularly in less sought-after regions.

What’s more, this research doesn’t provide guidance about designing program content. Increased admission requirements only affect the type of candidate who enters the program, not the quality of training they receive. Even if research shows that graduates of more selective programs are more effective, that does not shed light on whether or how the program contributed meaningfully to the completers’ success. More information is needed about what strategies preparation programs should use to train a broad range of candidates into effective teachers.

## Program Inputs

Another strand of the research on inputs focused on the structural components of a program, such as:

- Number of required courses in program
- Number of required content courses in program
- Number of required pedagogical courses
- Number of student teaching hours

Research on program inputs provides only the barest outlines of form-focused guidance for program design. It's not clear, for example, how much coursework or what type a program should require. The quantity of coursework, regardless of subject area, doesn't seem to matter for teacher effectiveness, with the exception of math and science content for secondary math and science teachers. In a 2008 meta-analysis, the Education Commission of the States (ECS) found little evidence that pedagogical training contributes to teacher effectiveness.<sup>10</sup> In a study for the Institute of Education Sciences (IES), the research arm of the U.S. Department of Education, researchers found that teachers who took fewer courses during their preparation program performed as well as teachers with higher coursework requirements, regardless of coursework content or if they completed an alternative or traditional program.<sup>11</sup> Two other studies, one on performance of all students in San Diego<sup>12</sup> and another on math performance of 9th-grade students in Chicago,<sup>13</sup> looked at the effect of a teacher's content background, measured by the teacher's undergraduate major, on student performance. These studies found no relationship between content background and performance, even if the teacher majored in the tested subject.

There is evidence, however, that secondary math and science teachers should have deep content expertise in their subject area. The ECS meta-analysis found moderate evidence that content knowledge in math can be useful, and the Harris and Sass study found that content-focused training has a positive effect for middle and high school math teachers. In a 2006 meta-analysis, Floden and Meniketti reached a similar conclusion: Teachers with more math and science content knowledge—whether measured by major, minor, or additional courses—tend to produce higher student achievement for secondary students in those subjects.<sup>14</sup>

Taken together, preparation program leaders could glean from this research that the number of content and pedagogical courses only matters for certain teachers—but this information wouldn't help them design an effective course of study. Should programs get rid of all content requirements except for math and science teachers? Are there other subject area specialties—science, music, special education—where the number

of courses makes a difference? And, critically, if sheer *quantity* of coursework doesn't necessarily make for better teaching, is *quality* of coursework a key ingredient? What should candidates learn in their courses that will be most useful to them in the classroom? Several states already have specific coursework requirements—Texas, for example, requires elementary teachers to have a content major and certain general coursework—despite the fact that the evidence doesn't support firm conclusions about whether this matters for more effective teaching. As with candidate inputs, programs need actionable research on how to improve their curriculum and coursework requirements.

The research is similarly unclear about the amount of student teaching that a candidate should have. The common-sense answer is “more”—longer student teaching requirements will produce more effective teachers. But that's not necessarily the case. The same IES study mentioned above found that the number of student teaching hours had no discernible effect on teacher performance, whether measured by the number of hours daily, the length of experience in weeks, or the number of full-length school days that student teachers were expected to spend fully in charge of their classrooms.<sup>15</sup> A 2015 study used data from more than a thousand teacher candidates in a large urban district and found that the duration of student teaching has little effect on teacher outcomes.<sup>16</sup> As with the studies on coursework, these findings raise the question of whether quality of student teaching matters more than quantity—and if so, what key features of clinical training are linked with better future teaching practices?

The evidence on student teaching placement suggests that where candidates do their clinical training is important for future teaching performance, but it's not clear how programs should design their programs based on this research. Is it more important for candidates to do their student teaching in schools with a student population that is similar to the schools where they are likely to work after graduation, or in schools that offer stable, high-quality environments for teaching and learning? In a 2008 paper of New York City data, Boyd et al. found that teachers who student taught in schools that were demographically similar to their eventual job placement were more effective than teachers whose student teaching and job placements were mismatched.<sup>17</sup> Similarly, a 2016 study of Washington State found that teachers in the sample were more effective if they student taught in a school that was demographically similar.<sup>18</sup>

Other research suggests that a high-functioning student placement site is more important than demographics. A 2010 study, also using New York City data, found that candidates who student taught in schools with lower turnover (assuming turnover is a proxy for how well the school functions) were less likely to leave NYC schools in their first five years and had better student achievement gains after they became the teacher of record than teachers from other placements. In the author's words, “this study finds that learning to teach in easier-to-staff field placement schools has positive effects on teacher retention



and student achievement gains, even for teachers who work in the hardest-to-staff schools.”<sup>19</sup> The 2016 study in Washington similarly provided evidence that student teaching in low-turnover schools was connected to better retention rates. More research is needed to sort out the different benefits derived from student teaching in a high-performing school and from gaining experience in a school with a targeted student demographic.

Finally, unlike most other inputs, the evidence on student teaching structure provides promising guidance for program improvement. Evidence suggests that programs that give candidates the opportunity to engage in actual teaching practice, such as planning a guided reading lesson, listening to a child read aloud for the purpose of assessment, and reviewing their future curriculum, produce substantially more effective first-year teachers in both ELA and math.<sup>20</sup> And programs that provide more oversight<sup>21</sup> of student teaching, including multiple observations by program staff and careful selection of cooperating teachers, also produce better teachers.<sup>22</sup>

Unfortunately, few programs have actually acted on this research—perhaps reflecting the general lack of emphasis on using actionable research to improve program design. In a 2014 analysis of over 1,600 teacher preparation programs, the National Council on Teacher Quality (NCTQ) found that only a third of programs followed the research’s recommendations for the number and quality of student teacher observations, and only 2 percent of programs played an active role in selecting cooperating teachers.<sup>23</sup>

## Outcomes-based Efforts Are on the Right Path, But Can't Yet Drive Program Improvement

The lack of evidence for input requirements led the field to try to improve in other ways. Instead of predicting program effectiveness based on inputs, the idea was to improve quality by assessing programs based on completer performance outcomes. Outcomes-based improvement efforts have been the focus of the past decade.

On the policy side, states have led these efforts, primarily through outcomes-based accountability systems. All states require that preparation programs go through a cyclical approval process in order to continue preparing teachers. Several states, like Louisiana, New Jersey, and Colorado, now link preparation programs with completer performance as part of those systems, and a few states use completer outcomes to make approval decisions. Florida, for example, differentiates programs into performance levels based on their completer outcomes, then makes approval decisions based on outcome data and an on-site visit. Federal policy recently moved towards outcomes: The recent federal regulations for Title II of the Higher Education Act require all states to link preparation programs to completer outcomes and publicize the results, and there are consequences for programs that are consistently low-performing.<sup>24</sup>

With the expanding availability of completer performance and program data, outcomes-focused teacher preparation research has increased. Researchers use these data, where available, to link program and completer performance and assess program effectiveness without the formal consequences of a state accountability system.

*In order for either policy or research efforts to improve teacher preparation, two conditions need to be true: Researchers need to be able to discern meaningful and consistent differences in program performance, and accountability must be paired with targeted, evidence-based interventions to improve program quality. At this point, neither criterion is met.*

But in order for either policy or research efforts to improve teacher preparation, two conditions need to be true: Researchers need to be able to discern meaningful and consistent differences in program performance, and accountability must be paired with targeted, evidence-based interventions to improve program quality. At this point, neither criterion is met.

First, outside of the 2000 Louisiana study, the research to date has not found authentic differences in program quality based on the student learning gains produced by program completers. Two large-scale studies in Missouri<sup>25</sup> and Texas<sup>26</sup> linked preparation programs to student achievement and found no meaningful differences between the programs. Two other studies, looking at data from North Carolina<sup>27</sup> and Washington State,<sup>28</sup> found some variation between programs based on completers' student outcomes, but the authors ultimately found that the variation within programs was larger than the variation between programs. In other words, all the programs produced teachers with a range of effectiveness; there were no clear "good" or "bad" programs. This result is not entirely unexpected—after decades of mandatory inputs and no formal incentive to innovate, many programs follow the same general outlines, so it's understandable that similar programs would produce similar quality completers. But if there are no measurable differences in program quality based on student learning gains, then the entire outcomes-based accountability theory of action falls apart.

Second, even if it were possible to identify high and low performers, that would not be enough to drive program improvement. Low-performing programs need targeted interventions, specific to their context, to design a more effective program and implement more effective practices. Neither outcomes-based research nor state approval systems, at least in their current states, are set up to provide the information necessary to design such interventions. The majority of these systems focus on determining only if a program "works," not on how well, under what circumstances, or for whom. For example, of the states that use outcome data to differentiate programs by performance, most separate programs into only a handful of performance tiers or levels, each encompassing a range of programs that differ in their challenges, assets, resources, and population. It is impossible to use these rating systems to prescribe corrective action tailored to different programs' needs, as the existing body of outcomes-based preparation program research does not provide guidance on promising strategies. The research solely seeks to determine what works—not how to make teacher preparation better.

## Future Research Must Be Useful for Program Design

*As a starting point, econometric research provided valuable insight. Overall, however, this body of research has not produced actionable information for program improvement.*

Recent policies to improve teacher preparation have been driven by econometric research. This type of research links administrative program data to quantifiable outcome measures to specifically and exclusively assess if teacher preparation “works.”

As a starting point, econometric research provided valuable insight. It helped to focus the field on demonstrating the effectiveness of teacher preparation practices rather than designing programs based on theory. And the econometric research led to several widely accepted conclusions: Certain common program inputs are not associated with complete effectiveness, and most programs currently prepare teachers of similar effectiveness.

Overall, however, this body of research has not produced actionable information for programs. There are four key problems with the econometric research to date: quality, specificity, timeliness, and applicability.

First, there is wide variability in the quality of existing studies. Several meta-analyses of research on preparation program effectiveness have found issues with the quality of research design in the majority of these studies, including problems with small sample size, lack of control or comparison group, and subjective outcome measures.<sup>29</sup> Second, econometric research is a blunt instrument. It attempts to assess what works, which is not specific enough to provide the type of diagnostic information needed to improve program quality. And third, even if the previous two conditions were met, econometric research is slow. It often takes years to move from research question to conclusion.

*Future teacher preparation research should be actionable in real time, use multiple measures of teacher performance, and produce strategies that can inform program design and practice.*

Finally, econometric research is not conducted with program design in mind. As a result, translating results into action is ultimately a guessing game. By using only this type of research, we miss out on insights that can improve the quality of teacher preparation.

Future teacher preparation research must fix the issues with econometric research, and move past the exclusive reliance on this type of research. Specifically, future research should embrace a **more nuanced approach** to better understand how and why a program or practice is effective. Future research should also be **actionable in real time**, use **multiple measures of teacher performance**, and produce strategies that **can inform program design and practice**.

Future teacher preparation research should be **more nuanced**, asking: What program components, if any, produce effective teachers, for what populations of candidates and students, in what types of schools, under what circumstances? It would take seriously the question of what happens to candidates in teacher preparation and how that relates to differences in practice as teachers. It would also recognize that what works for some candidates preparing to teach in certain schools might be different from what works for other types of candidates or in a different set of schools.

Future research should also include studies that are **actionable in real time** by expanding the use of rapid-cycle evaluations. Rapid-cycle evaluations are a product of the business world that the education sector has adopted.<sup>30</sup> Rapid-cycle evaluations are a form of quick-turnaround, smaller scale research that use existing administrative data to test and adjust specific program practices or strategies. Rapid-cycle evaluations can assess effectiveness in less than a year and begin to provide feedback in weeks or months—which is a very different timeline than the years that it takes to conduct a randomized controlled trial. The idea behind rapid-cycle evaluations is not to definitively decide if a practice is effective, but to quickly determine how well a practice is working for a specific set of candidates, students, and circumstances.

In the context of teacher preparation, a variety of actors could conduct rapid-cycle evaluations. Programs could partner with researchers to analyze program data and collect teacher performance data. Or programs could partner with districts, school operators, and/or states to access existing teacher performance data, then work with researchers or in-house to analyze the data. Ultimately, rapid-cycle research could produce a cache of effective practices, giving programs information on program performance and strategies for improvement on a faster timeline and targeted to the program's specific needs.

Rapid-cycle evaluations shouldn't only be used to assess existing practices. As programs build their research capacity, they should test out innovative program practices—specifically innovative practices building on related bodies of research and evidence from other rapid-cycle evaluations. Frontiers of Innovation, for example, is a research and development platform out of the Harvard Center on the Developing Child. Through

Frontiers of Innovation, a community of organizations develop intervention strategies based on the science of child development; use common measures to share findings; analyze the data to understand what works, for whom, in what context, and why; and test and iterate the interventions in a rapid cycle of learning and adaptation.<sup>31</sup> Similarly, rapid-cycle evaluations should look to other bodies of evidence to develop innovative practices. Leveraging research in brain development and child and adult learning, for example, could ultimately inform our understanding of both what good teaching looks like and how to help prospective teachers learn how to do it.

Rapid-cycle evaluations can complement the current, commonly used research methods, forming a suite of rigorous research methods that produce strategies that can **inform program design and practice**. To do that, we need to reframe the goal of teacher preparation research: to inform program design. Some of the best insights from econometric research have come from authors examining specific program components, like Boyd et al. did in their 2008 paper.<sup>32</sup> The authors provide much-needed guidance about effective student teaching training, by offering evidence that more vigilant program oversight of a student teaching experience may produce more effective teachers. This is only one study, looking at a relatively narrow question, but it is much more actionable than most research. To move the field forward, we need hundreds of more studies like this, asking similarly specific questions about a range of program components, activities, and practices.

Finally, future research must look at a **broader range** of teacher and student outcome measures. Recent research on the impact of teacher preparation programs focuses primarily on student growth, as measured using state test scores. Student achievement is a crucial indicator, and one researchers should continue to rely on, but research also shows that measures such as standardized classroom observations and student surveys can effectively identify teachers who produce better student learning gains.<sup>33</sup> Using a range of teacher performance measures can better inform program design by revealing program differentiation that was masked when looking at student growth alone. A study earlier this year, for example, linked observation scores of Tennessee teachers to their preparation programs and found authentic differences in the observation ratings of preparation programs. This study also found that the observation ratings correlated with teacher impact on student achievement.<sup>34</sup> This study also suggests that observation scores may reveal effective teacher behaviors, and provide guidance on how to train candidates in those behaviors.

There are also other valuable outcome measures that are linked to students' academic success, including program-level metrics, such as completer retention and job placement, and student-level outcomes, like chronic absenteeism, incarceration rates, and teen pregnancy rates. In fact, these data already exist in several states, including Florida, Rhode Island, and Colorado, which currently link preparation programs to performance measures such as job placement, retention, and teacher evaluation results. The point of this research isn't to hold programs accountable for these outcomes, but rather to see what effect, if any, teacher preparation programs have on a broad range of outcomes.

Crucially, alternative measures should always be accompanied by ongoing research assessing the relationship between the tool and student learning, even if the alternative performance data are not used for accountability. In Head Start, for example, the Classroom Assessment Scoring System (CLASS) is an observation-based tool used as one measure of Head Start program quality. Before CLASS was adopted, research showed that a teacher's CLASS score is predictive of her effect on student outcomes.<sup>35</sup> Since then, however, other research<sup>36</sup> has shown that teachers can improve their CLASS scores without any appreciable improvement in student achievement.

## Recommendations

To authentically improve teacher preparation, programs must have a cache of specific, actionable, evidence-based, and effective practices, and the flexibility, incentive, and knowledge to replace strategies that aren't working. To that end, we make several interrelated recommendations that are the shared responsibility of researchers, preparation programs, and state and federal policymakers.

- 1 Researchers should rethink the focus and design of future studies.
- 2 Teacher preparation programs should be a source of new research and effective strategies.
- 3 State and federal policymakers should foster innovation and collaboration.

### Researchers Should Rethink the Focus and Design of Future Studies

Researchers taking on this work in the future should move away from the current norm of teacher preparation research. From the perspective of improving teacher preparation, assessing which elements of a program are most—and least—helpful to novice teachers is more productive than asking, “Does the program work?” Instead of looking at high-level structural elements, like program and candidate inputs, researchers should go deeper into the specific components of preparation programs using formative research methods. Conclusions from rapid-cycle evaluations may not be as generalizable as a randomized controlled trial, but if the research meets rigorous, accepted standards for rapid-cycle evaluations, then this research can form a foundation for others to replicate and to build more rigorous studies upon.



Using these methods, researchers should specifically design studies where the independent variables are programs' specific practices and activities and, to the extent possible, include measures of the quality of those practices. As dependent variables, researchers should look into multiple measures of teacher performance: student achievement growth, observations of teacher practice, and principal and student surveys.

Putting these variables together, we can imagine studies that test out how different program components affect different outcome measures, giving us information about how and how well specific strategies in teacher preparation influence how teachers teach. Many programs, for example, require math and English content-specific courses. But which courses and types of course assignments are most closely linked with effective teaching strategies and higher student outcomes, in different subjects and grade levels? What elements do strong courses share? Does purposeful alignment between content courses and clinical experience produce more effective teaching? By asking—and beginning to answer—new questions, researchers can develop a body of effective strategies targeted to specific needs.

Researchers can only pursue this line of research, however, if states and programs make completer performance and program data accessible. Without those data, researchers are limited in the evidence they can produce—so as much as it is the responsibility of researchers to revise study designs, it's equally the responsibility of programs and states to give researchers the data to do so. States and programs are also responsible for the fact that existing research has been limited to blunt, high-level questions that cannot inform program design. Research partnerships, discussed in detail below, are one way states, programs, and researchers can improve the quality of teacher preparation evidence.

## **Preparation Programs Should Be a Source of New Research and Effective Strategies**

Teacher preparation programs are uniquely situated to identify, pilot, replicate, and evaluate promising training practices and strategies. Indeed, if teacher training is to improve, it is crucial that programs take an evidence-based approach to program design, and advocate for the support and resources necessary to test the impact of program practices on first-year teaching performance.

A critical first step in developing an evidence-based program design is defining the program's theory of action for training high-performing new teachers. Using this as a compass, programs can evaluate the existing teacher preparation research and determine which findings are most relevant for their population and context. Based on that evidence, they should determine: What are the “high-impact”<sup>37</sup> teacher practices—those that lead to higher student achievement—that we believe are important to inculcate in aspiring teachers? What program courses, activities, or content can we provide to help teacher candidates develop those high-impact practices? These questions should lead to an articulated program design and improvement strategy that programs can build on and tweak as new evidence arises.

*If teacher training is to improve, it is crucial that programs take an evidence-based approach to program design.*

Programs can then test their theory of action with rapid-cycle evaluations. As programs identify strengths and weaknesses, they should pilot and analyze the effectiveness of new practices. Research initiatives that identify effective—and ineffective—program practices should share those results with other programs.

Taking on this work requires programs to have several elements in place:

- All teacher preparation programs should regularly collect data on multiple measures of completer performance, including observation data for all teachers and student achievement data, where available. Programs should identify additional measures of classroom teaching performance that can be consistently collected in completers' first years of teaching.
- Preparation programs should form partnerships with school operators, districts, and states to share, analyze, and interpret these data. State and federal policies, as noted below, can support this effort.
- Program leadership should develop an intentional process—including the necessary data systems, partnerships with researchers, and internal structures—for conducting rapid-cycle formative research that meets accepted standards of rigor for empirical research. It is crucial that programs rely on external, unbiased researchers to conduct evaluations, and that programs are responsive, in real time, to the evaluation results. Programs should work with researchers to evaluate existing program components, pilot innovative practices, and try to identify high-impact teacher practices and the program activities that lead to those practices.
- Programs that identify relationships between specific practices and teacher performance—whether positive or negative—should publicize those results.

### **State and Federal Policymakers Should Foster Innovation and Collaboration**

To date, the state role in improving teacher preparation has been limited to compliance and accountability: States set minimum program requirements up front—some of which are excessively burdensome and constrain innovation—and then assess program completers on the back end. Similarly, the federal role has traditionally been limited to disbursing funding and collecting select program data. In a more effective system, state and federal policymakers would improve teacher preparation by creating the space and flexibility for programs to collaborate and innovate.

States and federal policymakers can foster innovation in several key ways:

- States and federal policymakers should encourage research partnerships between preparation programs and academic researchers. Through these research partnerships, researchers can access program data, and programs can access evaluation expertise to make program changes in real time. State and federal policymakers can foster this type of collaboration by dedicating existing research funds to these types of partnerships. Regional Educational Laboratories, for example, are federally funded research centers based at institutions of higher education, agencies, or partnerships between those entities that must use their funding to conduct and disseminate applied research and evaluations.<sup>38</sup>
- States should revise their preparation program approval processes to give high-performing programs the flexibility and incentive to take calculated risks with program practice. States could grant flexibility to programs that wanted to try something innovative and outside the parameters of current state requirements, provided the program showed the data that supported the innovation and had the strategy, structure, and capacity to test it.
- States that collect teacher performance data as part of the teacher evaluation or program approval system should make data available to researchers and programs. Given the issues to date with holding programs accountable based on their completer outcomes, it is particularly important that policymakers make data accessible and useful to programs and research teams, and set clear, high expectations about program improvement.
- Federal policymakers should use existing funding to incentivize preparation programs to form research partnerships and take on innovative program practices. A pot of funding from the Higher Education Act, for example, can be set aside for programs that take on these types of research partnerships.
- Policymakers and philanthropists should support replication and further testing of promising practices. If a program identifies a positive strategy, the state should convene and evaluate a voluntary pilot for other programs in the state. If the strategy is effective in the pilot, the state should consider adding the strategy as a potential corrective action in its program approval process. Federal policymakers and philanthropists can also identify promising strategies and support funding to pressure test those strategies.
- Federal and state policymakers should work with researchers, preparation programs, and districts to provide confidential access to the data necessary to track performance of program completers who train in one state and teach in another.

## A New Agenda for Teacher Preparation Research

If the field is ever going to systematically improve the effectiveness of new teachers, we first need to improve the quality of teacher preparation. The only way we can improve teacher preparation, whether alternative or traditional, is by improving preparation program design; to do that, we need research on what specific program design components work, how well, for whom, and under what circumstances. The existing teacher preparation research is not set up to provide that type of information. In this paper, we propose a new agenda for teacher preparation research that will.

As a field, we still don't know how to create an effective teacher preparation program, and yet states, programs, and analysts continue to place their faith—and the fate of nearly 200,000 teachers annually—in strategies that lack evidence. Instead of blindly swinging from one popular reform to the next, we propose that programs, states, and researchers form partnerships to systematically assess promising practices and strategically adjust program design based on what is best for that program. In other words, this agenda should be read as an argument for pluralism. Earlier this year, we proposed one option for improving teacher preparation policy;<sup>39</sup> other analysts have their own ideas.<sup>40</sup> These proposals each have their own merits, but the fact is that we do not have the research necessary to assess which proposal, or combination of proposals, would be most promising for improving the effectiveness of first-year teachers. Let's stop guessing, and start trying to find answers.

## Endnotes

- 1 Chad Aldeman and Ashley LiBetti Mitchel, "No Guarantees: Is it Possible to Ensure Teachers are Ready on Day One?" Bellwether Education Partners, February 2016.
- 2 James S. Coleman et al., "Equality of Educational Opportunity," National Center for Education Statistics, 1966.
- 3 Donald Boyd et al., "Teacher Preparation and Student Achievement," Working Paper No. 14314, National Bureau of Economic Research, <http://www.nber.org/papers/w14314.pdf>; Thomas Good et al., "How Well Do 1st Year Teachers Teach? Does Type of Preparation Make a Difference?" *Journal of Teacher Education* 57 (September/October 2006), No. 4: 410-430.
- 4 Diane Ravitch, "A Brief History of Teacher Professionalism" (presented at the White House Conference on Preparing Tomorrow's Teachers, March 5, 2002), <http://www2.ed.gov/admins/tchrqual/learn/preparingteachersconference/ravitch.html>
- 5 See, for example: Jacqueline A. Hughes, "Bridging the Theory-Practice Divide: A Creative Approach to Effective Teacher Preparation," *Journal of Scholarship and Teaching* 6 (August 2006), No. 1: 110-117; Zongyi Deng, "The Role of Theory in Teacher Preparation: An Analysis of the Concept of Theory Application," *Asia-Pacific Journal of Teacher Education* 32 (July 2004), No. 2: 143-157; Sharon Feiman-Nemser, "Teacher Preparation: Structural and Conceptual Alternatives," in W.R. Houston (Ed.), *Handbook of Research on Teacher Education* (New York: Macmillan, 1990).
- 6 See, for example: Matthew M. Chingos and Paul E. Peterson, "It's Easier to Pick a Good Teacher than to Train One: Familiar and New Results on the Correlates of Teacher Effectiveness" (paper prepared for a symposium, Harvard Kennedy School, 2010), [http://www.hks.harvard.edu/pepg/PDF/Papers/2010-22\\_PEPG\\_Chingos\\_Peterson.pdf](http://www.hks.harvard.edu/pepg/PDF/Papers/2010-22_PEPG_Chingos_Peterson.pdf); Dan Goldhaber, "Everybody's Doing It, But What Does Teacher Testing Tell Us About Teacher Effectiveness?," *Journal of Human Resources* 42 (2007), No. 4: 765-794; Jill Constantine et al., "An Evaluation of Teachers Trained Through Different Routes to Certification: Final Report" (NCEE 2009-4043), National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2009; Donald Boyd et al., "The Narrowing Gap in New York City Teacher Qualifications and Its Implications for Student Achievement in High-Poverty Schools," *Journal of Policy Analysis and Management* 27 (2008), No. 4: 793-818.
- 7 See, for example, the TeachStrong coalition's proposals, <https://www.americanprogress.org/press/release/2016/06/23/140331/release-teachstrong-coalition-calls-for-bold-ideas-to-strengthen-teacher-preparation/>
- 8 Jennifer L. Steele et al., "The Distribution and Mobility of Effective Teachers: Evidence from a Large, Urban School District," *Economics of Education Review* 48 (October 2015): 86-101.
- 9 U.S. Department of Education, National Center for Education Statistics, "Teacher Attrition and Mobility: Results From the 2012-13 Teacher Follow-up Survey" (NCES 2014-077), <http://nces.ed.gov/fastfacts/display.asp?id=28>
- 10 Michael Allen, "Eight Questions on Teacher Preparation: What Does the Research Say?," Education Commission of the States, 2003, <http://eric.ed.gov/?id=ED479051>
- 11 Jill Constantine et al., "An Evaluation of Teachers Trained Through Different Routes to Certification: Final Report" (NCEE 2009-4043), National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2009.
- 12 Julian R. Betts, Andrew C. Zau, and Lorien A. Rice, *Determinants of Student Achievement: New Evidence from San Diego* (San Francisco: Public Policy Institute of California, 2003).
- 13 Daniel Aaronson, Lisa Barrow, and William Sander, "Teachers and Student Achievement in the Chicago Public High Schools," *Journal of Labor Economics* 25 (2007), No. 1: 95-135.
- 14 Robert Floden and Marco Meniketti, "Research on the Effects of Coursework in the Arts and Sciences and in the Foundations of Education," in *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education* (Mahwah, NJ: American Educational Research Association, 2006), 261-308.
- 15 Jill Constantine et al., "An Evaluation of Teachers Trained Through Different Routes to Certification: Final Report" (NCEE 2009-4043), National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2009. Matthew Ronfeldt and Michelle Reininger, "More or Better Student Teaching?," *Teaching and Teacher Education* 28 (2012): 1091-1106.

- 16 Matthew Ronfeldt and Michelle Reininger, "More or Better Student Teaching?," *Teaching and Teacher Education* 28 (2012): 1091-1106.
- 17 Donald Boyd et al., "Teacher Preparation and Student Achievement" (Working Paper No. 14314, National Bureau of Economic Research, 2008).
- 18 Dan Goldhaber, John M. Krieg, and Ronny Theobald, "Does the Match Matter? Exploring Whether Student Teaching Experiences Affect Teacher Effectiveness and Attrition" (CALDER Working Paper No. 149), January 2016.
- 19 Matthew Ronfeldt, "Where Should Student Teachers Learn to Teach?" (faculty paper, Stanford University, 2010), [http://web.stanford.edu/group/cepa/FACULTYPUB/FP\\_ROMFELDT\\_2010.pdf](http://web.stanford.edu/group/cepa/FACULTYPUB/FP_ROMFELDT_2010.pdf)
- 20 Donald Boyd et al., "Teacher Preparation and Student Achievement" (Working Paper No. 14314, National Bureau of Economic Research, 2008).
- 21 The authors combine three sub-measures to form the "oversight of student teaching" variable: If the program 1) requires cooperating teachers to have a minimum number of years of teaching experience, 2) picks the cooperating teacher rather than having the pre-k through 12 school select them or the teacher volunteer, and 3) if a program supervisor observes each candidate at least five times during the student teaching.
- 22 Donald Boyd et al., "Teacher Preparation and Student Achievement" (Working Paper No. 14314, National Bureau of Economic Research, 2008).
- 23 Julie Greenberg, Kate Walsh, and Arthur McKee, "2014 Teacher Prep Review: A Review of the Nation's Teacher Preparation Programs," National Council on Teacher Quality, revised February 2015.
- 24 Department of Education, *Education Department Releases Final Teacher Preparation Regulations*, 2016. Retrieved from <http://www.ed.gov/news/press-releases/education-department-releases-final-teacher-preparation-regulations>.
- 25 Cory Koedel et al., "Teacher Preparation Programs and Teacher Quality: Are There Real Differences Across Programs?" (CALDER Working Paper No. 79), July 2012.
- 26 Paul T. Von Hippel et al., "Teacher Quality Differences Between Teacher Preparation Programs: How Big? How Reliable? Which Programs are Different?," Social Science Research Network (October 7, 2014).
- 27 Charles T. Clotfelter, Helen F. Ladd, and Jacob L. Vigdor, "How and Why Do Teacher Credentials Matter for Student Achievement?" (CALDER Working Paper No. 2, The Urban Institute, 2007).
- 28 Dan Goldhaber and Stephanie Liddle, "The Gateway to the Profession: Assessing Teacher Preparation Programs Based on Student Achievement," Center for Education Data & Research, accessed June 20, 2016, [http://www.cedr.us/papers/working/CEDR%20WP%202011-2%20Teacher%20Training%20\(9-26\).pdf](http://www.cedr.us/papers/working/CEDR%20WP%202011-2%20Teacher%20Training%20(9-26).pdf)
- 29 National Council on Teacher Quality, "Research Inventories on Features of Teacher Preparation Relevant to NCTQ Standards: Rationale and Methods," [http://www.nctq.org/dmsView/Intro\\_Research\\_Inventories](http://www.nctq.org/dmsView/Intro_Research_Inventories)
- 30 See, for example, the U.S. Department of Education's District Reform Support Network's recent tools on rapid-cycle evaluations for educators: <https://rttd.grads360.org/services/PDCService.svc/GetPDCDocumentFile?fileId=23420>
- 31 Center on the Developing Child, Harvard University, "Frontiers of Innovation," <http://developingchild.harvard.edu/innovation-application/frontiers-of-innovation/>
- 32 Donald Boyd et al., "Teacher Preparation and Student Achievement" (Working Paper No. 14314, National Bureau of Economic Research, 2008), <http://www.nber.org/papers/w14314.pdf>
- 33 Bill & Melinda Gates Foundation, "Ensuring Fair and Reliable Measures of Effective Teaching," January 2013, <http://www.edweek.org/media/17teach-met1.pdf>
- 34 Matthew Ronfeldt and Shanyce L. Campbell, "Evaluating Teacher Preparation Using Graduates' Observational Ratings," *Educational Evaluation and Policy Analysis*, 2016.
- 35 For an overview of CLASS, see: National Center on Quality Teaching and Learning, "Understanding and Using CLASS for Program Improvement," August 2012, <http://eclkc.ohs.acf.hhs.gov/hslc/tta-system/teaching/docs/class-brief.pdf>
- 36 Christina Weiland et al., "Associations Between Classroom Quality and Children's Vocabulary and Executive Function Skills in an Urban Public Prekindergarten Program," *Early Childhood Research Quarterly* 28 (2013), No. 2: 199-209.

- 37 Some programs have already begun designing teacher training around “high-impact” skills. The University of Michigan’s TeachingWorks organization, for example, has identified specific “high-leverage” practices, or a set of core skills that beginning teachers need to be successful; similarly, TNTP has identified a set of essential “launch” skills for novice teachers that are used to train aspiring teachers in the TNTP Teaching Fellows preparation program.
- 38 U.S. Congress, House of Representatives, “H.R. 4366: Strengthening Education Through Research Act,” 113th Cong., 2nd sess., April 2, 2014, [http://edworkforce.house.gov/uploadedfiles/billtext\\_hr4366.pdf](http://edworkforce.house.gov/uploadedfiles/billtext_hr4366.pdf)
- 39 Chad Aldeman and Ashley LiBetti Mitchel, “No Guarantees: Is it Possible to Ensure Teachers are Ready on Day One?” Bellwether Education Partners, February 2016.
- 40 See, for example, the teacher preparation recommendations proposed by ThirdWay, the National Center on Teacher Quality, TeachStrong, the American Federation of Teachers, and a coalition of nine teacher preparation programs (listed in order): <http://www.thirdway.org/report/creating-a-consistent-and-rigorous-teacher-licensure-process>; [http://www.nctq.org/dmsStage/Teacher\\_Prep\\_Review\\_2014\\_Report](http://www.nctq.org/dmsStage/Teacher_Prep_Review_2014_Report); <https://www.americanprogress.org/press/release/2015/11/10/125052/>; <http://www.aft.org/press-release/aft-report-teacher-preparation-recommends-bar-exam-and-demonstration>; <http://static1.squarespace.com/static/54dc2642e4b0469314195dca/t/5761ee149f7456a0f0a4e03f/>.

## Acknowledgments

The authors would like to thank our critical friends who reviewed this paper: It is miles better because of your thoughtful feedback. In particular, we'd like to acknowledge the contributions of James Cibulka, Dan Goldhaber, Cara Jackson, Sara Mead, Ana Menezes, Hannah Putnam, Benjamin Riley, and Andrew Rotherham. We would also like to thank the Bill & Melinda Gates Foundation for providing funding for this paper, giving us the opportunity to think through issues that are dear to our hearts. As always, the conclusions and recommendations of this paper are those of the authors alone.

For questions or further discussion of this paper, please email [ashley.mitchel@bellwethereducation.org](mailto:ashley.mitchel@bellwethereducation.org).



## About the Authors



### Ashley LiBetti Mitchel

Ashley LiBetti Mitchel is a senior analyst with Bellwether Education Partners' Policy and Thought Leadership team. Ashley has deep expertise in teacher preparation and has done extensive research, analysis, and writing on a range of teacher quality topics. Her work has been featured in *U.S. News and World Report*, *Huffington Post*, *The Washington Post*, *The 74*, *Hechinger Report*, and *RealClearEducation*, among others. Prior to joining Bellwether, Ashley worked for the U.S. Department of Education, Baltimore Education Research Consortium, Baltimore City Public Schools, Urban Teachers, and City Year Los Angeles. Ashley has a bachelor of arts with honors from McGill University and a master's of public policy from Johns Hopkins University.



### Melissa Steel King

Melissa Steel King is an associate partner at Bellwether Education Partners, where her work includes research, analysis, and writing on topics such as teacher preparation, alternative routes to teacher certification, and building teachers' capacity to implement personalized learning. Prior to joining Bellwether, Melissa gained deep expertise in education program evaluation as a research scientist at SUNY Albany's Center for Human Service Research and at Scholastic Inc's Research and Validation Department. She has worked as a teacher trainer for a school-based ethical awareness program, and as a kindergarten and first grade teacher in New York City and Boston. Melissa earned an M.A. in Elementary Education from Teachers College, Columbia University and an Ed.D. in Human Development and Psychology from the Harvard Graduate School of Education.

## About Bellwether Education Partners



Bellwether Education Partners is a nonprofit dedicated to helping education organizations in the public, private, and nonprofit sectors become more effective in their work and achieve dramatic results, especially for high-need students. To do so, we provide a unique combination of exceptional thinking, talent, and hands-on strategic support.

© 2016 Bellwether Education Partners



This report carries a Creative Commons license, which permits noncommercial re-use of content when proper attribution is provided. This means you are free to copy, display and distribute this work, or include content from this report in derivative works, under the following conditions:



**Attribution.** You must clearly attribute the work to Bellwether Education Partners, and provide a link back to the publication at <http://bellwethereducation.org/>.



**Noncommercial.** You may not use this work for commercial purposes without explicit prior permission from Bellwether Education Partners.



**Share Alike.** If you alter, transform, or build upon this work, you may distribute the resulting work only under a license identical to this one.

For the full legal code of this Creative Commons license, please visit [www.creativecommons.org](http://www.creativecommons.org). If you have any questions about citing or reusing Bellwether Education Partners content, please contact us.