NATIONAL ACADEMY *of* EDUCATION

EVALUATING AND IMPROVING
TEACHER PREPARATION PROGRAMS

Landscape of Teacher Preparation Program Evaluation Policies and Progress

STAFFORD L. HOOD, University of Illinois at Urbana-Champaign
MARY E. DILWORTH, Education Advisor
CONSTANCE A. LINDSAY, University of North Carolina at Chapel Hill

National Academy of Education

Evaluating and Improving Teacher Preparation Programs

Landscape of Teacher Preparation Program Evaluation Policies and Progress[1]

Stafford L. Hood, *University of Illinois at Urbana-Champaign*
Mary E. Dilworth, *Education Advisor*
Constance A. Lindsay, *University of North Carolina at Chapel Hill*

March 2022

CONTENTS

INTRODUCTION

The dialogue on what constitutes quality teacher preparation and how it should be assessed and evaluated is muddled. It begins neatly with a universal agreement and aspiration for quality teaching and enhanced PK-12 student achievement, then quickly scatters when there are attempts to define and weigh key components of academic excellence. No stakeholder group has the last say. The question "Who is *most* responsible?" for improving student achievement invariably prompts thoughtful discussion wherein each sector faults another and each may accept some ownership, but no one accepts full responsibility. We are well served to fix the problem, not the blame.

Recognizing that accountability is key, particularly to the nation's citizens, what emerges is a host of public summative measures intended to satisfy everyone with basic information and data points that have too often failed to substantively move the needle toward improved practice and outcomes. Moreover, the evaluation of teacher preparation programs (TPPs) does not occur in a vacuum isolated from the broader accountability movement in education, particularly the intensity of its focus to hold teachers, schools, and districts accountable. Questions about the effectiveness of teacher preparation, teacher classroom performance, and student achievement outcomes stem from a variety of sources that are inextricably linked to national, state, and local expectations, policies, and accountability systems. Those in the TPP sector of higher education are particularly nimble, being keenly aware of and ever ready with meaningful responses to probing questions with messaging intended to establish their credibility and generate support that allows for sufficient time and resources to facilitate designing and redesigning programs.

Our review reveals a multitude of issues that stifle useful evaluations, but we choose to focus on four key areas primed to leverage equitable TPP evaluation for future program improvement. First, the paper discusses the national and state policy authorities that establish large-scale TPP goals and incentives with the power to drive TPP designs and agenda. Second, the paper turns to prominent professional standards-setting organizations and other groups and individuals that are participants in the TPP evaluation sector with considerable influence on framing, creating metrics, and prioritizing what is deemed as fruitful areas for inquiry. Third, the paper discusses the impact of rapidly emerging models that require TPP evaluation criteria tailored for various approaches and standards to be useful to TPPs in their day-to-day work. Last, the paper discusses the critical need to re-examine all areas of TPP evaluation so that they capture and employ effective strategies addressing equity and social justice. The paper includes recommendations for improved alignment and consistency, timeliness and access, and equity, which may influence TPP evaluation in the future, as well as promising strategies for consideration.

As we provide an overview of the TPP evaluation landscape between 2013 (the year of the prior National Academy of Education [NAEd] report on the evaluation of TPP) and 2020, we also are cognizant of factors and conditions that may stall future progress. We contend that there are several areas of need in order to enhance formal and informal evaluations (i.e., better align public and private organizational policies and regulations, provide more and timely data accessible to TPPs and the public, and firmly establish an obligation to include matters of equity and social justice in all areas of the TPP evaluation sector).

Evaluations are tools intended to filter fact from fiction by providing what may be considered snapshots to inform decision-making. Unfortunately, this is not always the case but should be the intended goal. Specifically, evaluations can identify a course of action for TPPs to make progress toward achieving their visions for immediate and long-term goals. Still, nothing is static in the education evaluation space.[2]

Since 2013, numerous changes in the TPP context (e.g., proliferation of alternative routes to certification TPPs) have prompted different evaluation designs and methods. These new TPP formats have been significantly influenced by national, state, and local policymakers who are anxious about effectiveness, transparency, and speedy results. Furthermore, as one should reasonably expect, the strategies for the evaluative inquiry of TPPs must seriously consider the nation's uncompromising and partisan views on social justice that are influenced by the rapidly changing PK-12 student demographic, which requires a repertoire of culturally responsive knowledge and skills. In addition, there is widespread recognition that educators must consider students' social and emotional needs in order to advance their academic achievement.

The context of the "evaluand" (i.e., the object of the evaluation) is shaped by economic, political, historical, and cultural factors and dispositions of its primary stakeholders. This has, in some ways, been manifested in the continuing emphasis on accountability that relies on indicators as evidence of student achievement strategies preferred by executive branch initiatives, required by legislative mandates, framed by federal and state agencies, implemented by TPPs, and consumed by the public at large. For their part, states, TPPs, and accrediting agencies roll with the tide of innovation and reform in an effort to secure necessary resources to survive and thrive. All have played major roles in shaping the context of any evaluation lens that we might use to determine how well TPPs have succeeded in producing quality teachers. Understanding the context of a program is critically important to the validity of the evaluative findings and their usefulness for making formative and/or summative judgments, particularly if improvement is the priority.

A predecessor of this current paper, Feuer et al. (2013) focus on five categories in their review of the TPP evaluation landscape at that time: federal government, national accreditation, states, media/independent organizations, and TPPs. In this paper, we review the current TPP evaluation landscape with slightly different lenses by casting attention on influential public policies and organizations that inform and/or support TPP evaluation and prominent TPP formats, designs, methods, and assessments. Our lenses come from three perspectives: one author is a long-standing insider in the national teacher preparation and assessment policy arena, one is an academy-based evaluator and researcher with substantial experience in program evaluation and assessment (focusing on culture and cultural context), and one is an academy-based researcher who is well acquainted with emerging education issues in the economic and public policy domain. At the same time, this collaboration has established a certain

---

[2] There have been many events in the United States that shift and, in some cases, have delayed TPP evaluation policies and priorities. As the sector evolves rapidly, we acknowledge but have not addressed many of these changes. Certainly, the disruption of the COVID-19 global pandemic, nationwide civil unrest during the summer of 2020, and the appending incoherent delivery of PK-12 instruction, coupled with the yet to be imposed current U.S. presidential administration's agenda, will bring greater complexity to what had previously been a relatively predictable environmental context for evaluating TPPs.

level of symmetry among the co-authors and strengthened a more deliberate focus on issues of equity and access in the evaluation of TPPs.

FEDERAL, STATE, AND ORGANIZATION POLICIES

Federal-Level Policies and Regulations

It is clear that federal policymakers' primary interest in teacher preparation has not changed in decades—they seek quality and accountability. Similarly, the objectives in the evaluation of TPPs are determining quality, responding to accountability, and identifying areas for improvement. The importance, energy, and resources devoted to each is prompted by a variety of factors that pressure TPP institutions and organizations, namely the sentiment that the current structure or system for teacher preparation is expendable by failing to provide effective educators in a short enough time and at a reasonable cost. It has been persuasively argued that U.S. public investment in the PK-20 enterprise is insufficient to provide the necessary inputs for system improvement, yet those in business and industry expect that there should be a return on investment that is evidenced and documented by quantitative outcomes (Anderson, 2019; Moeller, 2020).

Arguably, one important factor that frames the current TPP evaluation environment has been presidential initiatives designed to encourage innovation, entrepreneurship, private investment, and control in public schools generally and public support for PK-12 charter schools specifically (Grossman & Loeb, 2016). The Higher Education Act (HEA), through Title II, authorizes programs designated for improving TPPs, but it has yet to be reauthorized. The annual HEA Title II report is a vehicle that was created to provide the transparency and public access heralded by the Bush and Obama administrations. These reports span more than 15 years, but their release is sporadic, data are inconsistent over time, and they are challenging for evaluations that require somewhat more precise metrics.

Building on the bipartisan No Child Left Behind Act, the Obama administration's stimulus package, American Recovery and Reinvestment Act of 2009, and the Race to the Top program established the need to better quantify TPP performance by calling for proficiency rankings and transparency. Cochran-Smith et al. (2017) assert that while the Bush administration leveraged education accountability standards generally, it was the Obama administration that raised the stakes for TPPs and teachers. It was

> exacerbated by the Obama Administration's Race to the Top policies and proposed federal requirements that states be required to rank teacher education institutions annually according to metrics established by the federal government, especially measurements of their graduates' impact on students' achievement. (p. 3)

The Obama administration's agenda was articulated by then Secretary of Education Arne Duncan in the U.S. Department of Education's (ED's) report *Our Future, Our Teachers* (U.S. Department of Education, 2011). The agenda firmly established proficiency rankings as desirable and transparency as a requirement. The preexisting annual HEA Title II report was one tool to provide public access. At one time, states were free to determine what data they provided to the federal government, which was reported to be more than 600 pieces of information (Cochran-Smith et al., 2018). Criticism from

the Obama administration and a U.S. Government Accountability Office (GAO) report (2015) indicated that few to none of the TPPs had been identified by states as having low-ranked preparation programs.

The Obama administration was unsuccessful in its attempts to strengthen the Title II legislative language through the reauthorization of HEA, settling in 2014 for a strategy of modifying the regulations that monitored its implementation. The proposed 2014 Title II regulations were reportedly opposed by both public and professional associations, with the American Association of Colleges for Teacher Education (AACTE) vocalizing opposition that it

> represented an unfunded mandate for schools, states, and higher education institutions; they impeded the recruitment of a diverse teacher workforce, particularly in high need areas; and they tied federal aid to preparation program evaluation based on expansion of an untested system. (Cochran-Smith et al., 2018, p. 28)

The regulations were ultimately approved in 2016, only to be repealed early in the Trump administration (Brown, 2017). The current HEA Title II Part A consists of a competitive grant program for a select group of TPPs and reporting requirements for accountability that are intended to track TPPs and improve program quality (Kuenzi, 2018).

The 2016 Title II regulations had established a framework for evaluating TPPs that required states to extensively report data that included, for example, TPP graduates' passing rates on state certification assessments, graduation rates, enrollments, student demographics, and other related program data for the purpose of ranking their TPPs and identifying those deemed to be low performing or at risk based on their criteria (Hegji, 2018; U.S. Department of Education, 2016). The 2016 Title II regulations required the establishment of a "federally mandated, state enforced data system designed to measure teacher education quality by requiring significant and controversial new methods of scoring, ranking, and funding teacher preparation programs" (Cochran-Smith et al., 2018, p. 55). These regulations provided directives for how states should evaluate their TPPs and then rank them with federal funding being the reward or withheld to be punitive. Primarily, the federal directives to evaluate TPPs were intended to use "meaningful data" that are indicative of outcomes such as students' performance on measures of academic achievement (Cochran-Smith et al., 2018, p. 59).

Since 2017, other policy attempts of note relative to teacher preparation evaluation are reflected in the reauthorization of Every Student Succeeds Act (ESSA) of 2015. ESSA's *Title II: Preparing, Training, and Recruiting High Quality Teachers, Principals, and Other School Leaders Part A: Supporting Effective Instruction* included a provision for state education agencies to provide funding to TPPs with the requirement that they

> award a certificate of completion (or degree) to a teacher only after the teacher has demonstrated that he or she is an effective teacher, as determined by the state; and limiting admission to the academy to prospective candidates who demonstrate "strong potential to improve student achievement" (Section 2002(4)). (Skinner, 2019, p. 10)

In the absence of new mandates, TPPs continue to labor, building and submitting reports that comply with the preexisting requirements. The Trump administration

was relatively silent about the importance of teaching and teacher education reform. It messaged to the public that data collection and transparency are superfluous and did not issue a comprehensive report on Title II data since Trump's first summer in office (reflecting state TPP reports from 2012-2013). Since 2017, the political temperament toward TPPs can be characterized as one of benign neglect that has diminished interest in leveraging evaluation as a critical activity.

Federal Data Systems

The federal sector could leverage current investments more effectively for TPP evaluation. For instance, in the short term the federal sector could create a user-friendly system in which researchers can link data sets such as the Integrated Postsecondary Education Data System (IPEDS) and HEA Title II, and in the long term create a comprehensive data set that encompasses teacher preparation, accountability programs, and competitive grant programs that can be used to drive innovation.

IPEDS is one comprehensive federally sponsored program that is under-utilized. Housed in ED's National Center for Education Statistics, it serves as a primary source for postsecondary education data and includes a variety of user-friendly tools (e.g., data trends that often are not made public elsewhere while being widely used for research studies). Although IPEDS's data could be an important performance metric for TPP evaluation, it has a number of protocols that make it challenging for the average user to accurately disaggregate and analyze discipline-specific information such as teacher education (Dynarski et al., 2015). For instance, as an AACTE Issue Brief (King, 2020) report states,

> Institutions completing the IPEDS survey are instructed to include all degree programs offered, even if no degrees were awarded in that field in the subject year. As a result, these figures include institutions that reported having an education program but that awarded no degrees in the subject year. (p. 7)

Unfortunately, there is no federal data set on enrollment in education programs, so there is no systematic way to identify programs that award few degrees but have robust enrollment. Furthermore, federal data sets fall short when tracking the demographics of teacher candidates and programs with reports often relying on scores of other public and private data sets to fill information gaps. The definition of terms, selection of items, and schedules for data collection by ED make it challenging, if not impossible, for policymakers to identify and use certain data points with confidence. It is apparent that one coherent federal data system that reflects TPP candidates' demographic characteristics, completion, and placement would provide critically important information for state and local policy decisions and should be a federal priority investment.

State- and Local-Level Policies and Regulations

There has been a long-standing question in the U.S. educational policy arena about the extent to which the federal government should be involved in and influence state education policies as well as their implementation. As should be the case, state and federal education policies have a considerable impact on the evaluation of TPPs with the recognition that the responsibility for education constitutionally resides with the

states. At the same time, the federal government is often seen as encroaching on states' responsibility for education with its considerable influence and funding.

Congress authorizes, appropriates, and targets funding to states and postsecondary institutions for specific areas of operation, including educator preparation and professional development and student financial aid. The federal government develops guidelines and regulations aligned with these policies, which to a greater or lesser extent require performance assessment as an accountability measure in the evaluation of outcomes. Because there is no national evaluation system per se, the legislative branch of government directly and indirectly incentivizes a fragmented system of TPP evaluation.

The individual states provide the most likely examples of TPP evaluation systems, as they approve TPPs, determine how they are evaluated, and decide what assessments or other tools are used for these purposes. Each state and territory holds quality teaching and learning to be of utmost importance in its responsibility for education. In their efforts toward quality teacher preparation, they work in close collaboration with regional organizations such as the Southern Regional Education Board and national ones that include the Council of Chief State School Officers (CCSSO) and the National Association of State Directors of Teacher Education and Certification (NASDTEC) to promote key principles of practice while advocating for state and federal legislation that will support their agenda. In this effort there is a heavy reliance on the standards and expertise of specialty groups and organizations such as the National Association for the Education of Young Children, the Council for Exceptional Children, and the National Council of Teachers of English, to mention a few, that provide their review and fine-tuning of requirements in approving TPPs.

The indicators of TPP quality are elusive, although they are typically grouped around basic, well-established principles of instruction and student learning that include subject-matter knowledge and student engagement. However, these valued principles proliferate into a wide assortment of indicators depending on whose judgments and preferences are prioritized in setting the standards that form the basis for these judgments. State longitudinal data systems can and should play important roles to inform these judgments but there is considerable need for their refinement.

State and Local Longitudinal Data Systems

Di Carlo and Cervantes (2018) highlight concerns in consistency and access to state data that potentially can contribute to research and evaluation of TPPs particularly on matters of educator diversity. The limited racial and ethnic representation in the PK-12 educator workforce is widely recognized as a national issue, yet ED's Office for Civil Rights' biannual Civil Rights Data Collection does not require states to report this information. The authors effectively argue that a central, nationwide collection and promulgation of these data is the best way to ensure comprehensive availability to the public and can contribute to a more complete view of areas of need and resource to effectively fund programs and policies. Furthermore, the majority of states collect this information, but they are free to define demographic categories (e.g., include or omit "mixed race" that is a rapidly growing cohort in this nation's population). Lastly, the absence of a national and transparent data set can stifle TPP recruitment efforts for

candidates of color as well as interstate reciprocity for licensed educators. The challenge here is that the fractured nature of education governance does not ensure the consistency of data collection across states.

While there is recent work suggesting that matching teachers and students by race has a positive impact on PK-12 students of color, in particular (Cherng & Halpin, 2016; Egalite et al., 2015; Gershenson et al., 2018; Redding, 2019), capturing diversity data and program's diversity impact present formidable challenges. Fenwick's (2021) comprehensive review of TPP evaluation in the states highlights the wide range of authorities and directives that are intended to inform policy but at the same time distract TPPs from their essential teaching and learning missions.

Even though the states hold the primary authority for TPP approval and evaluation, the influence of local school districts cannot be overlooked—particularly large urban and suburban school districts—as they also engage in formal and informal evaluations of teacher preparation. Lastly, one frequently untapped data source is human resources data that can be found at the district or state level (Goings et al., 2021). The fact that these data are now able to link TPPs and student performance also comes as a result of efforts to develop summative evaluations for teachers.

Clearly, greater coordination of national, state, and local data collection efforts will yield TPP evaluations that are useful and meaningful to institutions and to the constituents that policymakers serve. At the same time prominent professional standards-setting organizations and others also significantly influence the frames, metrics, and priorities for TPP evaluation.

Organization and Other Policies and Influence

Accrediting Organizations

The most recognized players in the evaluation context are the two federally approved TPP accrediting groups (the Council for the Accreditation of Educator Preparation [CAEP] and the Association for Advancing Quality in Educator Preparation [AAQEP]) and other organizations that rate TPP performance. Program accreditation is often frustrating to institutions that are subject to their requirements. Yet, the enterprise continues to grow. Some TPPs do not see the necessity of national accreditation given the associated financial costs and labor-intensive exercises associated with it when state program approval suffices for teacher credentialing within the state. In recent years, there has been a reconfiguration of accrediting agencies, with new entities arguing that their approach is what the universe of TPP needs to move forward. Generally, each agrees that TPP quality is important and that TPPs should be engaging in ongoing improvement resulting in the enhanced academic and life success of U.S. students, but this sentiment does not distinguish one organization from another.

One key player is CAEP, which represents a "strategic union" between its predecessor accrediting agencies—the National Council for Accreditation of Teacher Education (NCATE) and the Teacher Education Accreditation Council (TEAC). It proclaims a new direction in the accreditation of TPPs that is more evidence based and congruent with the national trend of data-driven accountability, while also endorsing the revisions to the Title II regulations with its existing standards (Cochran-Smith et al., 2018). The

standards initially developed by CAEP were widely publicized to be congruent with the call for accountability that was strongly echoed by the Obama administration's programs and initiatives. Cochran-Smith et al. strongly assert:

> The CAEP standards seemed intended to appease both policy makers who worked from the neoliberal logic underlying the era of accountability and members of the profession who were resistant to the logic. (2018, p. 85)

However, Cochran-Smith et al.'s further assessment of CAEP was that in its "claims to be revolutionizing accreditation in terms of the content dimension of accountability, it was similar in many ways to accreditation through NCATE and TEAC at least on the surface" (2018, p. 85).

One newcomer in the TPP accreditation arena is AAQEP. Founded in 2017, AAQEP reports accrediting 25 TPPs and in 2021 received Council for Higher Education Accreditation recognition as an accrediting organization. Clearly, AAQEP is intended to provide an accreditation alternative to CAEP as the main accreditor of TPP—one that is more inclusive through strong collaborative partnerships with TPPs and intentional and direct involvement with PK-12 educators and administrators. Therefore, it is reasonable to suggest that CAEP left room for a new player to enter the game. In articulating its standards, the AAQEP website uses terms such as "culturally responsive practice" and "community/cultural context," conveying the message of inclusiveness that overlaps the TPP and the community that its graduates serve.

Cochran-Smith et al. suggest that AAQEP could have promise as it

> emphasizes diversity and equity in their procedures suggesting that standard solutions to local challenges will not suffice.… [There is] emphasis on teacher candidates' classroom performance rather than their impact on tested achievement of eventual students; and support of innovations and variations in keeping with diverse local contexts and communities. (2018, p. 179)

Both CAEP and AAQEP continue to tweak their messaging, but their ability to survive and thrive hinges to a great extent on state and local policymakers' understanding of cost and benefit value for the communities that they represent.

The National Council on Teacher Quality (NCTQ), created in the early 2000s as a private advocacy organization for improving the quality of teacher preparation, has the loudest voice within the education sector. While it is not an accrediting organization, it is closely affiliated with influential, conservative, and reform-minded groups and policymakers, such as the Thomas B. Fordham Institute, that have been critical of the teacher education establishment for many years. NCTQ's mark continues to be its highly publicized TPP rating and ranking system and subsequent reports, which are criticized by researchers and teacher educators based on their allegedly flawed methodology, minimal samples, and unsubstantiated conclusions. NCTQ initially focused primarily on input-based standards that included entry criteria, syllabi, and student teaching, as examples. The TPPs were rated on a five-point system for each of the standards, which then provided a composite score to determine program ranking (Cochran-Smith et al., 2018).

The 2015 NCTQ report *State of the States: Teaching Leading and Learning* was conspicuously released toward the end of the second Obama term and is perceived as an attempt to revise the TPP evaluation space through the proposed revisions of the Title II regulations. The NCTQ report responded positively to the more performance-based approach in evaluating teacher effectiveness, indicating that this was broadly evident in state policy.

NCTQ's January 2017 report *Running in Place: How New Teacher Evaluations Fail to Live Up to Promises* was not as favorable about the progress that had been made in the evaluation of TPPs since its 2015 report. This is not surprising because the revised Title II regulations of the Obama administration had only been approved in October 2016 after the failed approval of the revised 2014 regulations. Therefore, it is likely that the uncertainty of whether the 2016 revised regulations would be implemented by the next presidential administration probably resulted in a holding pattern for TPP evaluation. The NCTQ 2017 report noted that some progress had been made by the states to "significantly" use student academic growth in teacher evaluation, with 30 states making it a major priority and 10 states somewhat requiring it, but still another 10 states and the District of Columbia did not require any "objective" measure of student growth.

The report also argued that 18 of the state education agencies (SEAs) had lax regulations in the credentialing of teachers because the SEAs still provided some teachers with an "effective" summative rating even if the teachers received a "less than effective" score on their student learning evaluations. As expected, this report was not received well by the TPP community. It should also be apparent that there is not full participation by TPPs in the NCTQ process as the organization continues to generate controversial ranking reports and is considered to be an agitator by many in the TPP community with what can be characterized as limited evaluative inquiry of TPPs, based on its methodology and politicized positioning (Cochran-Smith et al., 2018).

Perhaps the most dominant shadow in this work is cast by AACTE. The association represents more than 700 colleges and universities in the teacher preparation enterprise, with its current "who we are" statement reporting that it is "dedicated to high-quality, evidence-based preparation that assures educators are ready to teach all learners." Collaborating with other national groups, AACTE generates research and policy briefs while serving as the primary advocate for TPPs in federal educational policy and in state educational policy through its affiliate groups. Yet, there have been tensions between AACTE and the TPP community particularly around connecting TPP quality to graduates' effectiveness, as indicated by the subsequent performance of their students on standardized achievement measures such as a value-added measure (VAM) approach. Cochran-Smith et al. (2017) suggest that a coalition of AACTE and other professional associations contributed to the demise of the proposed 2014 Title II revised regulations. AACTE maintains professional interest in TPP accreditation and evaluation, but no longer financially supports related activities as it did in prior years.

Nongovernmental Organizations

Aydarova (2020) effectively argues that absent policy limits, certain nongovernmental, intermediary organizations (IOs) constitute closely knit accountability regimes that "allow IO actors to amass material, informational, and relational resources to advance

their agendas despite seeming opposition to the measures they propose from the educational community" (p. 4). There are a number of organizations that have a legacy and thus prestige in the development of assessments that accumulate useful data for TPP evaluations. Key among them are the Educational Testing Service (ETS), Pearson, and research and development organizations such as the American Institutes for Research (AIR), Westat, RAND Corporation, and Mathematica. These organizations stand to advise the federal government, states, and districts and create assessment data systems on demand. They are often invisible knowledge brokers, but their work is often filtered by sponsors and access is restricted. Pertinently, there are a number of national non-profit organizations that over time have had a keen interest in how TPPs are evaluated. Aside from the mission of establishing a quality teaching force, their interests range from responding to the needs and safeguarding the viability of their member constituents to having some say in the financial resourcing of state and federal policies that may impact their work. They include but are not limited to CCSSO, NASDTEC, the American Federation of Teachers, and the National Education Association.

In addition to various organizations and professional associations, the involvement and influence of philanthropic entities cannot be overlooked. For instance, the Bill & Melinda Gates Foundation (the Gates Foundation) has made significant funding contributions at multiple levels since 2013, with $34.7 million going to fund five teacher preparation transformation centers to "develop, pilot and scale effective teacher preparation practices to help ensure that more teacher-candidates graduate ready to improve student outcomes in K-12 public schools" (Bill & Melinda Gates Foundation, 2015). The Gates Foundation announced that this was its "first investment as part of its teacher preparation strategy … focused on supporting programs that:

- Give candidates authentic opportunities to build and refine their skills;
- Commit to continuous improvement and accountability;
- Ensure that those who prepare new teachers are effective; and
- Are shaped by K-12 systems and the communities they serve" (Bill & Melinda Gates Foundation, 2015).

Yet, it is also important to note Will's 2018 article in *Education Week* titled *An Expensive Experiment: Gates Teacher Effectiveness Program Shows No Gains for Students*. The Gates Foundation had invested $212 million into the Memphis, Tennessee; Pittsburgh, Pennsylvania; and Hillsborough County, Florida, school districts as well as in a school consortium in California beginning in 2009-2010 with matching funds from the districts, which reportedly totaled $575 million for the initiative to design teacher evaluation systems that would include both observation rubrics and measures of "growth in student achievement." However, after 5 years, a study by RAND and AIR (funded by the Gates Foundation) reports no improvement in student outcomes. Will further noted that the study "found no evidence that low-income minority students had greater access to effective teachers than their white, more affluent peers, which had been another stated goal of the Gates Foundation" (2018, p. 9).

It is possible that the Measures of Effective Teaching (MET) Project, the Gates Foundation's investment in a 3-year study "on fair and reliable measures of effective teaching—improving student test scores" whose findings were reported in 2013

(Measures of Effective Teaching Project, 2013), was running in parallel with the afore-mentioned teacher evaluation project. Unquestionably, these investments by the Gates Foundation have made significant contributions to the evaluative inquiry of teacher preparation and teacher effectiveness. Grants to certain organizations do have the potential to leverage criteria on TPP evaluation components. For example, the William + Flora Hewlett Foundation's support for the National Commission on Social, Emotional, and Academic Development and its final report *From a Nation at Risk to a Nation at Hope* effectively advanced the need for social and emotional learning in more than 200 pieces of legislation (Shriver & Weissberg, 2020).

Foundations also have the wherewithal to test and substantiate certain research methods that find their way into evaluation. The concept of value added, for instance, rooted in the work of agricultural economist William Sanders, was effectively established as a key criterion in a number of TPP state and federal grant programs until its effectiveness was disavowed by researchers in the field (Amrein-Beardsley, 2008; McCaffrey et al., 2003). As Smith and Smith (2009) contend, many foundations carry a reputation of bipartisanship, have the opportunity to fund policy-changing strategies over a sustained period of time, and can serve as a countervailing force in society by representing views and providing financial support in areas that are different from those of the government. This situates them in a powerful place.

There are an increasing number of highly regarded professional educators and economists who have stepped out of the fray to establish organizations that allow them to promote new TPP evaluation methods that have utility. For example, Edward Crowe's Teacher Prep Inspection–US has adapted the British inspection method to the U.S. context, using inspection teams. It conducts on-site visits, interviews, reviews, examinations of data quality, and observations of teacher candidates. It has completed inspections of 180 TPPs in 21 states. Often, TPPs are invited to participate in these and similar initiatives, being typically identified by reputation and/or through professional acquaintances. Rarely is there an open call for programs to apply. The process tends to include the same TPPs (i.e., large research institutions) and omits many minority-serving and small private colleges. At the same time, there has been some progress made in the training and participation of evaluators of color who are increasingly involved in major evaluation projects (Collins & Hopson, 2014). However, their participation is not as evident in major TPP evaluations and particularly not as lead contractors for these evaluations.

Influential Reports

Notably a number of reports also influenced the TPP evaluation sector. For instance, one report, *Approaches to Evaluating Teacher Preparation Programs in Seven States* (Meyer et al., 2014), provides a glimpse of how TPPs in one region began to adjust their evaluation priorities in response to the Obama administration's 2011 publication *Our Future, Our Teachers* (U.S. Department of Education, 2011). Focusing on the seven states in the Regional Educational Laboratory (REL) Central region—Colorado, Kansas, Missouri, Nebraska, North Dakota, South Dakota, and Wyoming—the report suggests that the evaluation of TPPs mirrors findings in the 2013 NAEd report in that they are "primarily state program approval processes, which vary substantially" (Feuer et al., 2013, p. 2).

It was noted that TPPs in the REL Central region were increasingly emphasizing measures "that focus more closely on program outcomes for teacher candidates, practicing teachers, and their students" (Meyer et al., 2014, p. 18).

A 2015 report by GAO, *Teacher Preparation Programs: Education Should Ensure States Identify Low-Performing Programs and Improve Information-Sharing*, is also important in the context of TPP evaluation. This report was published shortly after the failure to approve the major revisions to HEA Title II in the 2014 regulations and reinforced that a major purpose for the Title II report was for states to identify TPPs that were *low performing.* However, the GAO findings were that the identification of these TPPs was minimally evident in the reporting by the states and viewed to be an inefficient or even a meaningless exercise. The report not only found that seven states had no process for identifying their low performing TPPs but also that ED officials had not adequately verified the processes used by states to identify low-performing TPPs. The report further strengthens the argument that more useful data needs to be collected from states in their annual Title II reports that would contribute to assessing TPP quality. Both the inadequate identification of low-performing and at-risk TPPs by the states and the less than useful data submitted by states in their annual Title II reports were major aspects of the revised regulations of 2016 that were approved for the short term. It is also important to note that a review of the 2017-2018 reported data (a report has yet to be published) on the Title II website (https://www.ed.gov) indicates that 162 TPPs were identified as at risk or low performing, a 260 percent increase compared to 2014.

EVALUATION MEASURES AND METRICS

Although TPP formats are vastly different, there are critical components that virtually all program models purport to include, such as some measure of basic subject-matter knowledge and clinical field experiences. While accrediting organizations remain a predominant model for program evaluation, the proliferation of TPP designs has called forth additional factors for consideration.

Proliferation of Program Designs

The phrase "traditional teacher education" is a misnomer. Since the mid-1980s, the initial and continuing professional development of teachers has shifted from being firmly situated in college and university-based programs to a host of new venues designed to swiftly fill state and local needs in certain disciplines (e.g., science, technology, engineering, and mathematics; special education) and rectify the broadening racial, ethnic, and linguistic gap between PK-12 students and quality educators that work to teach them (Dilworth & Coleman, 2014; McFarland et al., 2018; U.S. Department of Education, 2016b). Once challenged by postsecondary institutions as competitors in the sector, many schools, colleges, and departments of education now host and/or collaborate with them. Today, the roughly 30 percent of TPPs that are classified as *alternative route* are hosted by local public school districts; public and for-profit charter schools; state, regional, and local education agencies; community college systems; foundations; and nonprofit programs (Fenwick, 2021; U.S. Department of Education, 2016b; Wilson & Kelly, 2021). These programs vary significantly in design and delivery and operate

under various state authorities; thus, "in practice, all states are not requiring that all providers and programs meet the same standards" (Fenwick, 2021, p. 19). Debatably, there are no apparent efforts to craft measures that recognize distinctions between and among program types and at the same time signal program quality.

As TPP formats proliferate, so too grows the need for useful and reliable evaluation frameworks (Bartell et al., 2018). In a comprehensive review of alternative models of teacher education programs, Cochran-Smith and Villegas (2016) find that studies address one or more of the following questions:

- Is this particular teacher preparation program successfully doing what it claims to be doing (or wants to be doing)?
- What is the evidence for this (and how could it be demonstrated to outsiders)?
- How can program faculty and administrators use this evidence or the explanatory frameworks developed in conjunction with it in order to improve the program and/or to contribute to the broader knowledge base about teacher preparation? (p. 463)

These are important questions, but to what extent do they prompt the development of new qualitative and quantitative measures, as well as evaluative insights, that are of the most interest to the communities they serve (Wells & Roda, 2016)?

There are a multitude of intersecting entities that direct and inform TPP evaluation. Key among them are state governing boards and authorities and program accreditation and licensing organizations. Fenwick (2021), in the comprehensive report *A Tale of Two Cities: State Evaluation Systems of Teacher Evaluation Programs*, provides a useful comparison of "typical" traditional and alternative route provider and program approval processes and standards (e.g., admissions, institutional mission, quality of instruction) (see Table 1). The comparison suggests that the evaluative evidence provided to decision-makers for determining TPP quality varies by program type with traditional programs carrying a heavier burden of proof than others.

Teacher residency programs are a case in point. This popular TPP model is highly regarded as it offers a universally supported preparation component of clinical experience and at the same time employs individuals as they prepare, which makes the programs more attractive to individuals of color than traditional programs (Cochran-Smith & Villegas, 2016; Dilworth & Coleman, 2014; Guha et al., 2016; Papay et al., 2012; Rice & Brent, 2002) and are often framed within a "third space" (Beck, 2016), in another word, hybrid spaces that provide for an authentic teaching and learning environment between campus based and school-based work (Zeichner, 2010).

One element for comparison is a TPP's effectiveness in preparing new teachers who are employable and stay in the field. Generally, here traditional programs offer pass rates on licensure exams and/or hiring and retention data while alternative route programs offer an assessment and evaluation of candidates for certification and TPP improvement. Acceptance to teacher residency programs typically require formal agreements to work in cooperating PK-12 schools while in training and upon program completion commit to work in these districts. TPP reports to authorizing agencies may be useful documentation but of minimal use to evaluation. The length of time teachers

TABLE 1 Comparison of Typical Traditional and Alternative Route Provider and Program Approval Processes and Standards

Traditional	Alternative (Not IHE-Based)
Admissions criteria • GPA of incoming class • Average licensure/entrance exam scores	Admission and recruitment criteria • Bachelor's degree from an accredited institution • Average licensure/entrance exam scores • Target cohort size and a plan for recruiting candidates
Institutional mission, vision, goals, conceptual framework • Narrative evidence of alignment of unit conceptual framework with institutional mission, vision, and goals	Ownership, governance, and physical location/address Budget and revenue sources
Quality and substance of instruction • Coursework and syllabi aligned with CAEP/state standards with special emphasis on diversity, equity, and inclusion and assessment/data driven instructional decision making • Planned program of study with required course content and hours • Student and program rubrics, assessments, and data aligned with standards	Coursework • Description of instructional modules (typically online modules) aligned with targeted categories of certificates • Description of how students are evaluated
Quality of student teaching experience • Fieldwork policies, including requisite hours in handbook • Qualifications of fieldwork supervisor and mentor teacher • Record of regularly scheduled observations of student teaching by university supervisor	Clinical training • Evidence of support during training, clinical teaching, internship, and practicum • Description of support and communication between students, cooperating teachers, and the alternative certification program • Description of conditions under which clinical teaching may be implemented
Faculty qualifications and orientation • Percentage of faculty with advanced degrees and PK-12 teaching experience • Percentage of full-time, part-time, and adjunct faculty • Profile of clinical and internship partner schools • University orientation for university supervisor, adjunct faculty, and cooperating teachers	Selection criteria for supervisors and cooperating teachers • Selection criteria for clinical supervisors • Selection criteria for cooperating teachers • Code of professional conduct of staff and students
Effectiveness in preparing new teachers who are employable and stay in the field • Pass rates on licensure exams • Hiring and retention data	Assessment and evaluation of candidates for certification and TPP improvement

TABLE 1 Continued

Traditional	Alternative (Not IHE-Based)
Success in preparing high quality teachers • Teacher performance assessments adminis-tered near end of program • Ratings of graduates by principals/employers • Program completers' self-assessment of knowl-edge, skills, and dispositions • Impact on PK-12 learning outcomes	Certification procedures
Quality assurances	Complaint procedures
Typically 5- to 7-year cycle	Typical 3-year cycle, can range up to 7 years

NOTE: IHE = institution of higher education.
SOURCE: Fenwick, 2021.

from the respective program types remain in the field may provide critically important and useful information to consider as well. Therefore, it is reasonable to explore the identification of criteria that may better inform the evaluation of emerging alternative models and the measures and metrics to be used. Yet, we must also address the limitation of current measures and metrics used to evaluate TPPs, particularly the misalignment of the data that are available.

Evaluation Measures, Metrics, and Data Misalignment

The aforementioned VAMs have represented a field shift in the conception of teacher and school quality. These measures of teacher performance undergird a larger movement in education that seeks to rank schools using data generated from test scores and provide transparent metrics for multiple sets of stakeholders. Teacher quality has come to mean a teacher's ability to grow student learning over time as measured by these models. As data proliferate, all elements of the education system have been influenced by this concept of teacher quality.

TPPs are not exempt from the movement to provide performance metrics indicative of their production of quality teachers entering the teaching profession. A significant element of the Race to the Top legislation required that states produce report cards for each TPP (Crowe, 2011). These report cards were to use data about programs and their graduates that would ideally link their performance to the academic performance of the students in the schools where they are initially placed. There was also a desire that state TPPs should be rated and ranked based on these metrics. Here, the Obama administration sought to induce improvement in the quality of these programs by making these report cards public and using summative measures as indicators of quality for the consumers (e.g., districts, principals, parents) of TPP graduates. Many states implemented these systems and continue to use some form of public reporting for their TPPs.

It is not surprising that these efforts were not without controversy within the TPP community. In particular, many programs felt strongly about the inappropriateness of using value-added estimates from their candidates' students to judge their programs. Indeed, one might imagine a scenario where certain metrics have unintended

consequences that harm programs and do not induce improvement, particularly if the program places its teacher candidates at high-needs and hard to staff schools (Cochran-Smith et al., 2016).

The shift in how TPP quality would be evaluated had begun prior to the 2013 NAEd report, from input indicators to outputs and outcomes based on some form of a performance metric. While the more input-focused metrics for TPPs (admission criteria, curriculum, faculty, etc.) continue to be argued as important, it is clear that the outcome and performance types of metrics of TPP effectiveness, such as graduates' successful performance on state teacher certification tests, VAMs, and student growth, are more highly valued through the persistent lens of accountability. At the same time, there is some appreciation for the value of TPP graduates and principals' surveys as important indicators of consumer satisfaction. There does appear to be some consensus that there should be evidence of a teacher's contribution to their students' learning but there is no consensus about what that evidence should look like and who determines what evidence is acceptable to show this impact. All of the accrediting entities agree that teacher and student performance are indicative of teacher effectiveness and TPP quality.

Accountability Measures

An understanding of methods and assessment with regard to TPP evaluation should have a primary focus on understanding the pressures of accountability now facing TPPs. As mentioned in the previous sections, the major push for accountability measures largely comes from ED reporting requirements as formerly espoused in the Title II regulations, CAEP standards, and the development and widespread adaptation of portfolio-based assessments (Cochran-Smith et al., 2016). What these calls for accountability have in common is a focus on public summative measures (i.e., measures that seek to distill performance into a summative rating that captures program performance). This focus on a single, summative rating represents a true shift in how these programs are evaluated and is consistent with trends in education accountability systems. Prior to this current focus, the field relied on state approval of programs, pass rates on licensure exams, and whether programs and schools met accreditation requirements (Donovan et al., 2014).

Predictive Effectiveness Measures

The increased use of student and teacher data in evaluating TPP performance is a result of many states now having longitudinal data systems and other infrastructure that make it possible to link teacher preparation candidates directly to the performance of their students. In particular, the use of student value-added metrics as measures of TPPs is a natural outgrowth of their use in teacher evaluation systems. However, as much as value-added models have proven controversial in the PK-12 space, they are also contested in the teacher preparation space. Additionally, their use as evaluative measures for programs has not been empirically borne out in the data. For example, Goldhaber (2019) uses administrative data from the state of Washington to show that there are minor differences in value added among graduates of preparation programs. He notes that there are few studies that capture the actual features of preparation pro-

grams and workforce outcomes. Similarly, Lincove et al. (2014) find that statistically robust, value-added metrics can be estimated, but they are sensitive to the selection of teachers into programs and jobs, decisions about accountability criteria, and the selection of control variables.

In addition to value-added metrics, some scholars have investigated how other elements of state-level teacher evaluation systems might be used to judge TPP effectiveness. Some studies show that few individual program requirements are positively associated with achievement gains (Preston, 2017). Rating instruments each measure a single underlying construct rather than multiple constructs (Henry et al., 2013). Bastian et al. (2018) analyze the relationship with the evaluation rating of program graduates and find that there were significant differences by TPPs, but that it was critical to control for school context. They argue that evaluation ratings provide evidence on the performance of TPPs that is distinct from value added. Using data from the North Carolina Educator Effectiveness System, they uncovered large variation among and within programs and found that the ratings on the observation rubrics based on North Carolina teacher and administrator standards are good predictors of performance because they capture elements of the preparation program in practice.

A report of the National Academies of Sciences, Engineering, and Medicine (2020) concludes:

> The research base on preservice teacher preparation supplies little evidence about its impact on teacher candidates and their performance once they are in the classroom. Preservice programs in many states assess the performance of teacher candidates for purposes of licensure, but few states have developed data systems that link information about individual teachers' preservice experiences with other data about those teachers or their performance. Overall, it is difficult to assess the causal impact of teacher preparation programs. (p. 6)

Another promising program feature is observation ratings. Using a sample of 44 providers offering 184 programs across Tennessee, Ronfeldt and Campbell (2016) find that observational ratings such as those from the state teaching evaluation rubric are associated with student achievement gains.

Portfolio assessment (e.g., edTPA and PPAT) is a highly subscribed tool to gauge a beginning teacher's readiness to practice. As of 2018, 45 states had adopted some form of portfolio assessment (Whittaker et al., 2018). These assessments serve a dual purpose: to measure candidate performance and to evaluate program performance. These assessments come with recommended cut scores that are aligned with a state's professional standards and are subject to local needs and political intent. TPPs can use evidence from portfolio assessments for continuous improvement when the scores exhibit construct validity, reliability, and have predictive power (Admiraal et al., 2011). The scores from the exams can also be used by programs for continuous improvement via comparisons to other programs in their home state (Bastian et al., 2016). Bastian et al. (2018) demonstrate that the edTPA in particular can be a useful way to understand profiles of instructional practices by TPPs. They also find statistically significant relationships between the edTPA and the Education Value-Added Assessment System, meaning that the edTPA can be a useful predictor of eventual teacher performance. Though the edTPA is most widely used, there are a variety of portfolio assessments available to the field,

including the PPAT developed by ETS and loosely aligned with the Interstate Teacher Assessment and Support Consortium standards, the Texas-sponsored and ETS-developed Pre-Admission Content Test, the California Teaching Performance Assessment hosted by the California Commission on Teacher Credentialing, the Resident Educator Summative Assessment hosted by the Ohio Department of Education, and the recently defunct Washington State Professional Educator Standards Board portfolio. Critics of the portfolio assessment purport that it is an additional tool in a movement to privatize public education because it is often used as a high stakes accountability assessment that can place significant burdens on the candidates (Whittaker et al., 2018). The edTPA is grounded in the more senior portfolio assessment (i.e., the well-regarded National Board for Professional Teaching Standards assessment). Similar to the National Teacher Examination assessment of the 1970s and the Praxis® examinations of the 1990s, the edTPA has been highly scrutinized for a host of issues including its relevance to current teaching and learning theories, psychometric measures, and impact on underrepresented racial, ethnic, and linguistically diverse groups (Gitomer et al., 2019). More recent criticisms of the edTPA focus on challenges around norming and validity, and the lack of sustained oversight by technical committee members (Gitomer et al., 2021). Although there is a fair amount of controversy surrounding the merits of the assessment (Gitomer et al., 2019; Goldhaber et al., 2017; Peck et al., 2014; Tuck & Gorlewski, 2016), it is still well situated in the initial teacher performance domain. It is apparent that there continues to be considerable debate regarding the measures and metrics used to provide meaningful information in the evaluation of TPPs.

EQUITY AND SOCIAL JUSTICES

Targeting groups' (stakeholders') positionality relative to school reform and social justice is particularly important. Underlying this movement toward public summative measures as evaluators of program success is a critical discussion of what should be used to evaluate teachers. Cochran-Smith et al. (2016) describe this as a tension between "thin equity" and "thick equity," where the former focuses solely on in-school conditions as drivers of educational disparities and the latter focuses on both in-school and out-of-school factors. The public generally and racially and ethnically marginalized communities specifically are increasingly weary of evaluation findings that state and restate the existence of a PK-12 achievement gap between and among White students and others. They have come to understand that well-prepared teachers and more teachers of color in particular are key drivers of better student performance. Yet, rarely is this quantitative or qualitative information explicit in proposed or existing legislation or acted on (Dilworth, in press).

Evaluation is typically recognized as a tool for TPP accountability and program improvement but fail to appreciate its possibilities as a vehicle to advance institutional equity and/or the nation's social justice agenda (Hood et al., 2015a; House, 2019, 2020). The extent to which TPPs prepare educators who successfully support PK-12 academic achievement, particularly for racially, ethnically, and linguistically diverse underserved students, is arguably an important metric that should influence the allocation of financial and other resources. Therefore, it seems reasonable to recognize and review TPPs with an evaluative lens that meets quality practice and productivity thresholds. It is

apparent that minority serving institutions (MSIs) should be included in this group. As Petchauer and Mawhinney (2017) posit "policy demands facing teacher education at this contemporary moment also make this the right time to see MSIs as a collective unit in teacher education" (p. 6).

Contributions to Providing More Teachers of Color

MSIs are a subset of the postsecondary sector and are distinguished by their missions, goals, and affiliation. Notably, historically Black colleges and universities (HBCUs) and American Indian Tribally Controlled Colleges and Universities have historical roots that bind them in significant ways. Together with Asian American and Native American and Pacific Islander–serving institutions and Hispanic-serving institutions, these institutions generate a significant number of educators generally and teachers of color specifically (Dilworth, 2012; Dilworth & Brown, 2008; Gasman et al., 2016; Lindsay & Lee, 2018).

The need and merits of a diverse teaching force is well documented most recently by Cherng and Halpin (2016); Gershenson et al. (2018, 2021); and Gist (2017). There is a critical need to increase the number of Black, Indigenous, and people of color as the racial, ethnic, and linguistic diversity of the nation's PK-12 student population has grown exponentially. The societal expectation is that all TPPs should recruit and prepare educators from various cultures and that school districts should do a better job of retaining them in PK-12 classrooms. At the same time, it is evident that this responsibility has not been fully shared by TPPs as MSIs continue their long-standing tradition to be more responsive in meeting this need than others.

The reasons for the under-representation of educators of color are complex, varied, and have changed somewhat over time, including inadequate financial support to pursue teaching, poorly constructed career ladders, and a limited number of individuals pursuing teaching degrees who came from distressed urban and rural areas, completed college, and returned to their home communities. Furthermore, a focus on accountability measures that include challenging teacher assessment licensure examinations and the dominance of a postbaccalaureate licensure format that adds the cost of a fifth year of study are deterrents (Carter & Goodwin, 1994; Carver-Thomas, 2018; Dilworth & Coleman, 2014; King, 1993).

One factor that has influenced the number of potential PK-12 educators generally and those of color specifically is an increased interest and participation in alternative routes to licensure. These programs are hosted by IHEs and states, school districts, and nonprofit organizations and typically provide individuals with the option to be trained and work and to be simultaneously compensated. The merits of this pipeline are that individuals enter PK-12 classrooms quickly and qualify for school positions. The shortcoming is that those who are trained through these alternative routes tend to retreat from the classroom sooner than those prepared in traditional college- and university-based TPPs (Espinoza et al., 2018). One can reasonably assume that enrollment trends favoring alternative route programs will continue to rise in MSIs, boosting efforts to diversify the teaching force. King and Mahaffie (2016) document the contribution of HBCUs, noting that 16 percent of Black or African American individuals who enrolled in IHE-based TPPs matriculated in HBCUs, and alternative, IHE-based programs had

a higher percentage of students enrolled in HBCUs (4 percent) than that of traditional IHEs (2 percent).

Secretary of Education Arne Duncan's 2013 annual report (U.S. Department of Education, 2013) to Congress on teacher quality notes that 69 percent of TPPs are classified as traditional, 21 percent are alternative route TPPs based at IHEs, and 10 percent are alternative route TPPs not based at IHEs. Approximately 37 percent of enrollees in IHE-based alternative programs are of color and 53.7 percent are of color in non-IHE based alternative programs.

In their review of effective teacher diversity state initiatives, Dilworth and Coleman (2014) suggest that there is merit in embracing alternative route teaching and learning formats, but at the same time there is a need to establish clear and universal standards and guidelines. Given the successes of MSIs in generating a diverse corps of educators in any format, evaluation criteria that reflect their work and are grounded in the culturally responsive program principles should be developed and utilized.

A number of studies and reports have sparked interest in factors that broaden thinking, theory, and practice in educational evaluation to address issues of access, equity, inclusion, and social justice. For example, Hood et al. (2015a, 2015b) argue for leveraging the importance and critical need to view evaluation through a culturally responsive lens; the National Academies of Sciences, Engineering, and Medicine's *Monitoring Educational Equity* (2019) promotes the quantification of equity indicators for large-scale data collection; and Wimberly's 2015 volume *LGBTQ Issues in Education: Advancing a Research Agenda* includes the use of large-scale data sets in examining LGBTQ education. In addition, there are a number of recent, highly publicized works that have expanded the discussion of access, equity, inclusion, and social justice in TPPs and TPP evaluation, including *Who Believes in Me?: The Effect of Student–Teacher Demographic Match on Teacher Expectations* (Gershenson et al., 2016); *The Importance of Minority Teachers: Student Perceptions of Minority Versus White Teachers* (Cherng & Halpin, 2016); and *The Long-Run Impacts of Same-Race Teachers* (Gershenson et al., 2018). Lastly, Dilworth (2018) promotes the idea that there is merit in considering the intersectionality of teachers' race, ethnicity, and age as a factor in program assessment and evaluation.

Efforts to provide the public with summative measures and reliance on publicly generated databases too often omit important qualitative data that can provide contemporary and culturally responsive lenses. These data are rarely valued in the state and federal policymaking domain. As Toldson (2019) states:

> Today, researchers routinely separate numbers from people. We use deficit statistics, test scores, achievement gaps, graduation rates, and school ratings, without a humanistic interpretation. We also create false dichotomies between qualitative and quantitative research. (p. 3)

Some advocacy and special interest organizations, such as Excelencia in Education, the Urban Institute, and the Albert Shanker Institute, and publications—notably *Diverse Issues in Higher Education*—with and without private support, fill a void by accepting the task of extrapolating quantitative data from large databases and analyzing the information for consumption and consideration in policy initiatives that target education issues of race, ethnicity, language, exceptionality, and inclusion. They do so

in user-friendly technology formats, but also provide technical reports to inform those in the research and evaluation sector.

It is not necessary to create new models on how to include members of the community in the evaluation of TPPs, as extensive examples can be found in the literature on evaluation theory and practice. There are encouraging examples in health, social work, Indigenous evaluation, and some sectors of education in which community stakeholders are more substantively included in the evaluation process (i.e., design, implementation, and interpretation of results) but are not clearly apparent in the evaluation of TPPs nationwide.

The substantive inclusion of community stakeholders in the program evaluation process is most closely aligned with multicultural validity (Kirkhart, 1995), deliberative democratic evaluation (House & Howe, 1999), culturally responsive evaluation (Frierson et al., 2010; Hood et al., 2015b), and the Indigenous evaluation framework (LaFrance & Nichols, 2008). This call for the inclusion of community stakeholders has also been accompanied by the long-standing one to increase the number of evaluators of color and those with "shared lived experiences" when conducting evaluations in culturally diverse communities to strengthen evaluative validity (Collins & Hopson, 2014; Hood, 2001; Hood et al., 2005; Reid et al., 2020). House and Howe (2000) provide examples of what the deliberative democratic evaluation approach looks like in practice with Cochran-Smith et al. (2017), offering this approach for consideration to address democratic accountability in teacher education. Frazier-Anderson et al. (2011) provide the African American Culturally Responsive Evaluation System for Academic Settings, applying the Culturally Responsive Evaluation's lens for the inclusion of community stakeholders throughout the evaluation process. Numerous chapters in Hood et al. (2015a) provide examples as to how community stakeholders have been included in program evaluation in culturally diverse settings. However, the most robust examples are evaluations conducted in Indigenous communities, primarily by Indigenous evaluators (Cram et al., 2014; LaFrance et al., 2012).

CONCLUSION

For a variety of reasons, evaluations intended to inform policymakers and the public on TPP performance typically do not meet their goals. Public and private initiatives that are designed to promote quality teacher preparation, improve PK-12 instruction, and enhance student learning are advanced, absent thoughtful consideration of evaluation findings. It is counterproductive for TPP institutions and organizations to respond to various accountability directives without the time and opportunity to understand their meaning and to make reasonable adjustments in operations before moving to one politically fueled concept after another.

There are examples of TPP evaluations having an impact on federal or state policies intended to improve TPPs and TPP procedures (Bastian et al., 2016; Sykes & Dibner, 2009). Yet, since 2013, we find that there is limited information suggesting that these initiatives have met their program improvement goals. It can be argued that the Trump administration's immediate repeal of the Obama administration's 2016 revisions to the Title II regulations may have created a vacuum, resulting in a pause in the attention

to the evaluation of TPPs. At the same time, repealing these regulations seems to have signaled that those priorities for TPP accountability were no longer important and were being left to be addressed by the states. One could surmise that this vacuum hindered innovation and change at the state and institutional level.

Research has indicated that there is as much variation in teacher outcomes within TPPs as there is among programs (Goldhaber et al., 2013). The fundamental purpose of TPP evaluation should be to provide valid and useful information to make evaluative judgment about TPP performance and program improvement. As we have described, the countervailing notions and movements that happen in education policy often work at cross-purposes against these goals for TPPs. Good, sound evaluations offer a clear path to program improvement if the system allows. What appears to be lacking are clear, consistent, and transparent goals defined by all stakeholders (i.e., state and national policymakers, program accrediting agencies, organizations, and the public). At the same time it is clearly apparent that there should be a central, nationwide collection of useful data to improve the evaluative inquiry of TPPs that includes a current and accurate compilation of state data. The availability of more comprehensive data to the public can contribute to a more complete view of areas of need and resources to effectively fund programs and policies. Key to a more fruitful investment of time, money, and resources is to retreat from public summative measures by establishing data systems that accommodate quantitative and qualitative indicators that explicitly target community needs—candidate outcomes and TPP improvement—and incorporate equity indicators that are often overlooked.

We believe that this paper provides a reasonably clear snapshot of the TPP evaluation landscape's complexity that exists within the context of federal and state education policy environment, varying TPP models, standards-setting accreditation groups, and influential organizations and individuals. We offer our observations with examples of how each of these entities influence the development and operation of data systems that too often generate information with limited utility. In addition, we promote a message to all that there is a critical need to re-examine all areas of TPP evaluation in order to capture and employ effective strategies that address equity and social justice.

Certainly, there is more ground to be covered as researchers and practitioners continue to interrogate, articulate, explore, and refine the TPP evaluation landscape. We believe one place to start is with a clear and deliberate understanding that TPP evaluation is an essential tool for meaningful program improvement that is the primary responsibility of TPP providers. Of course, this evaluation of TPP quality and utility for program improvement must rely on sound evaluation measures and metrics that do not reify quantitative information as the only real truth or minimize the importance of TPPs' social responsibility. We expect that more than a few will disagree with our call to substantively increase the participation of highly trained and experienced evaluators from marginalized communities in the TPP evaluation landscape. We believe such participation is not only important in bringing in diverse and culturally relevant knowledge and experiences into the evaluation process but also, more importantly, can contribute to the validity of the findings from these evaluations. Particularly, when these TPPs are major providers of teachers in these communities. The challenge before us shall not be an easy one to undertake. Nor should it be.

With these concluding reflections in mind, we offer the following recommendations as a place to start the next phase of this important discourse to improve and evolve the evaluation of TPPs.

Recommendations

Data Alignment, Consistency, Timeliness, and Access

Public- and private-sector agencies and influencers should work to establish a coherent, TPP data collection system. This system should:

- Establish and adhere to data collection schedules that are calibrated with similar information-gathering efforts and initiatives
- Define terminology and metrics that are current and accommodate the needs and capacity of states, local school districts, and the communities they serve
- Expand the capacity for decision-making on the ground (e.g., tailor rankings and report cards for consumer knowledge and use)
- Align TPP state program approval and professional accreditation data collection and reporting processes into more rapid cycles that allow for ongoing continuous improvement and the formative evaluation of TPPs
- Make readily available assistance on methods for appropriately interpreting quantitative and qualitative data to TPPs, states, and school districts

Equity and Social Justice

Publicly supported TPP data collection activities should:

- Encourage the involvement of researchers from all TPP levels and types (e.g., liberal arts, teacher residency in evaluation initiatives)
- Identify and incentivize TPPs in MSIs that are successful in producing teachers of color
- Prioritize the participation of evaluators from marginalized communities who have substantive evaluation training and experience
- Encourage and support nongovernmental organizations' data review and analysis, particularly those whose missions focus on traditionally disenfranchised teacher candidates and communities
- Explicitly prioritize diversifying the PK-12 teaching force as one of the most important goals and establish substantive criteria as a requirement in competitions for research, practice, and evaluation grants and contracts

---

REFERENCES

Admiraal, W., Hoeksma, M., van de Kamp, M-T., & van Duin, G. (2011). Assessment of teacher competence using video portfolios: Reliability, construct validity, and consequential validity. *Teaching and Teacher Education: An International Journal of Research and Studies*, 27(6), 1019-1028.

Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, 37(2), 65-75.

Anderson, L. (2019). Private interests in a public profession: Teacher education and racial capitalism. *Teachers College Record*, 121(6), 1-38.

Aydarova, E. (2020). Shadow elite of teacher education reforms: Intermediary organizations' construction of accountability regimes. *Educational Policy*. OnlineFirst. https://doi.org/10.1177/0895904820951121.

Bartell, T., Floden, R., & Richmond, G. (2018). What data and measures should inform teacher preparation? Reclaiming accountability. *Journal of Teacher Education*, 69(5), 426-428. https://doi.org/10.1177/0022487118797326.

Bastian, K. C., Lys, D., & Pan, Y. (2018). A framework for improvement: Analyzing performance-assessment scores for evidence-based teacher preparation program reforms. *Journal of Teacher Education*, 69(5), 448-462.

Bastian, K. C., Henry, G. T., Pan, Y., & Lys, D. (2016). Teacher candidate performance assessments: Local scoring and implications for teacher preparation program improvement. *Teaching and Teacher Education*, 59, 1-12.

Beck, J. S. (2016). The complexities of a third-space partnership in an urban teacher residency. *Teacher Education Quarterly*, 43(1), 51-70.

Bill & Melinda Gates Foundation. (2015, November 18). *Gates Foundation awards over $34 million in grants to help improve teacher preparation programs* [Press release]. https://www.gatesfoundation.org/Media-Center/Press-Releases/2015/11/Teacher-Prep-Grants.

Brown, E. (2017, March 8). *Senate overturns Obama-era regulations on teacher preparation*. *The Washington Post*. https://www.washingtonpost.com/local/education/senate-overturns-obama-era-regulations-on-teacher-preparation/2017/03/08/b8cf127a-041c-11e7-b9fa-ed727b644a0b_story.html.

Carver-Thomas, D. (2018). *Diversifying the teaching profession: How to recruit and retain teachers of color*. Learning Policy Institute.

Carter, R. T., & Goodwin, A. L. (1994). Chapter 7: Racial identity and education. *Review of Research in Education*, 20(1), 291-336.

Cherng, H-Y. S., & Halpin, P. F. (2016). The importance of minority teachers: Student perceptions of minority versus White teachers. *Educational Researcher*, 45(7), 407-420.

Cochran-Smith, M., Baker, M., Burton, S., Chang, W-C., Cummings Carney, M., Fernández, M. B., Stringer Keefe, E., Miller, A. F., & Sánchez, J. G. (2017). The accountability era in US teacher education: Looking back, looking forward. *European Journal of Teacher Education*, 40(5), 572-588.

Cochran-Smith, M., Carney, M., Keefe, E., Burton, S., Chang, W., Fernandez, M. B., Miller, A. F., Sanchez, J. G., & Baker, M. (2018). *Reclaiming accountability in teacher education*. Teachers College Press.

Cochran-Smith, M., Stern, R., Sánchez, J. G., Miller, A. F., Stringer Keefe, E., Fernández, M. B., Chang, W-C., Cummings Carney, M., Burton, S., & Baker, M. (2016). *Holding teacher preparation accountable: A review of claims and evidence*. National Education Policy Center.

Cochran-Smith, M., & Villegas, A. M. (2016). Research on teacher preparation: Charting the landscape of a sprawling field. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed.) [eBook]. American Educational Research Association.

Collins, P. M., & Hopson, R. (Eds.). (2014). *Building a new generation of culturally responsive evaluators through AEA's graduate education diversity internship program: new directions for evaluation, Number 143*. John Wiley & Sons.

Cram, F., Kennedy, V., Paipa, K., Pipi, K., & Wehipeihana, N. (2014). Being culturally responsive through Kaupapa Māori evaluation. In S. Hood, R. Hopson, & H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. 289-311). Information Age Publishing.

Crowe, E. (2011). *Getting better at teacher preparation and state accountability: Strategies, innovations, and challenges under the federal Race to the Top program*. Center for American Progress. https://cdn.americanprogress.org/wp-content/uploads/issues/2012/01/pdf/teacher_preparation.pdf?_ga=2.193825517.1948718756.1607544112-746937442.1607544112.

Di Carlo, M., & Cervantes, K. (2018, September). *The collection and availability of teacher diversity data: A state-by-state survey* [Research brief]. Albert Shanker Institute. https://www.shankerinstitute.org/sites/default/files/teacherracedataFINAL.pdf?_ga=2.232721567.1848603042.1607545253-1978149100.

Dilworth, M. E. (2012). Historically Black colleges and universities in teacher education reform. *The Journal of Negro Education*, *81*(2), 121-135.

Dilworth, M. E. (Ed.). (2018). *Millennial teachers of color*. Harvard Education Press.

Dilworth, M. E. (In press). The absence and probability of effective public policies for teacher diversity. In C. Gist & T. Bristol (Eds.), *Handbook of research on teachers of color*. American Educational Research Association.

Dilworth, M. E., & Brown, A. L. (2008). Teachers of color: Quality and effective teachers one way or another. *Handbook of Research on Teacher Education*, 424-467.

Dilworth, M. E., & Coleman, M. J. (2014). *Time for a change: Diversity in teaching revisited*. National Education Association.

Donovan, C. B., Ashdown, J. E., & Mungai, A. M. (2014). A new approach to educator preparation evaluation: Evidence for continuous improvement? *Journal of Curriculum and Instruction*, *8*(1), 86-110.

Dynarski, S. M., Hemelt, S. W., & Hyman, J. M. (2015). The missing manual: Using National Student Clearinghouse data to track postsecondary outcomes. *Educational Evaluation and Policy Analysis*, *37*(1 Suppl), 53S-79S.

Egalite, A. J., Kisida, B., & Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review*, *45*, 44-52.

Espinoza, D., Saunders, R., Kini, T., & Darling-Hammond, L. (2018). *Taking the long view: State efforts to solve teacher shortages by strengthening the profession*. Learning Policy Institute.

Fenwick, L. (2021). *A tale of two cities: State evaluation systems of teacher preparation programs.* American Association of Colleges of Teacher Education. https://3e0hjncy0c1gzjht1dopq44b-wpengine.netdna-ssl.com/wp-content/uploads/2021/10/AACTE_Final.pdf.

Feuer, M. J., Floden, R. E., Chudowsky, N., & Ahn, J. (2013). *Evaluation of teacher preparation programs: Purposes, methods, and policy options*. National Academy of Education.

Frazier-Anderson, P., Hood, S., & Hopson, R. K. (2011). Preliminary considerations of an African American culturally responsive evaluation system. In S. D. Lapan, M. T. Quartoli, & F. J. Riemer (Eds.), *Qualitative research: An introduction to methods and designs* (pp. 347-372). Jossey Bass.

Frierson, H., Hood, S., Hughes, G., & Thomas, V. (2010). A guide to conducting culturally responsive evaluation. In J. Frechtling (Ed.), *The 2010 user-friendly handbook for project evaluation* (Report No. REC 99-12175, pp. 75-96). Division of Research and Learning in Formal and Informal Settings, Directorate for Education and Human Resources, National Science Foundation.

Gasman, M., Castro Samayoa, A., & Ginsberg, A. (2016). *A rich source for teachers of color and learning: Minority serving institutions*. Penn Center for Minority Serving Institutions.

Gershenson, S., Hansen, M., & Lindsay, C. A. (2021). *Teacher diversity and student success: Why racial representation matters in the classroom*. Harvard Education Press.

Gershenson, S., Hart, C., Hyman, J., Lindsay, C., & Papageorge, N. W. (2018). *The long-run impacts of same-race teachers* (Report No. w25254). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w25254/w25254.pdf.

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, *52*, 209-224.

Gist, C. D. (2017). Voices of aspiring teachers of color: Unraveling the double bind in teacher education. *Urban Education*, *52*(8), 927-956.

Gitomer, D. H., Martínez, J. F., & Battey, D. (2021). Who's assessing the assessment? The cautionary tale of the edTPA. *Phi Delta Kappan*, *102*(6), 38-43.

Gitomer, D. H., Martínez, J. F., Battey, D., & Hyland, N. E. (2019). Assessing the assessment: Evidence of reliability and validity in the edTPA. *American Educational Research Journal*, *58*(1), 3-31.

Goings, R. B., Walker, L. J., & Wade, K. L. (2021). The influence of intuition on human resource officers' perspectives on hiring teachers of color. *Journal of School Leadership*, *31*(3), 189-208.

Goldhaber, D. (2019). Evidence-based teacher preparation: Policy context and what we know. *Journal of Teacher Education*, *70*(2), 90-101.

Goldhaber, D., Cowan, J., & Theobald, R. (2017). Evaluating prospective teachers: Testing the predictive validity of the edTPA. *Journal of Teacher Education*, *68*(4), 377-393.

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., & Loeb, S. (2016). Improving the teacher workforce. In M. Hansen & J. Valant (Eds.), *Memos to the president on the future of U.S. education policy*. The Brookings Institution.

Guha, R., Hyler, M. E., & Darling-Hammond, L. (2016). *The teacher residency: An innovative model for preparing teachers*. Learning Policy Institute.

Hegji, A. (2018). *The Higher Education Act (HEA): A primer* (Report No. 7-5700 R43351). Congressional Research Service. https://fas.org/sgp/crs/misc/R43351.pdf.

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

Hood, S. (2001). Nobody knows my name: In praise of African American evaluators who were responsive. In J. Greene & T. Abma (Eds.), *Responsive evaluation: Roots and wings* (pp. 31-43). New Directions for Evaluation, no. 92: Winter 2001. Jossey-Bass.

Hood, S., Hopson, R. K., & Frierson, H. T. (Eds.) (2005). *The role of culture and cultural context: A mandate for inclusion, the discovery of truth and understanding in evaluative theory and practice*. Information Age Publishing.

Hood, S., Hopson, R., & Frierson, H. (Eds.). (2015a). *Continuing the journey to reposition culture and cultural context in evaluation theory and practice*. Information Age Publishing.

Hood, S., Hopson, R. K., & Kirkhart, K. E. (2015b). Culturally responsive evaluation. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (pp. 281-317). Wiley.

House, E. R. (2019). Evaluation with a focus on justice. *New Directions for Evaluation*, *2012*(163), 61-72.

House, E. R. (2020). Evaluating in a fragmented society. *Journal of MultiDisciplinary Evaluation*, *16*(36), 26-36.

House, E., & Howe, K. R. (1999). *Values in evaluation and social research*. Sage Publications.

House, E. R., & Howe, K. R. (2000). Deliberative democratic evaluation. *New Directions for Evaluation*, *85*, 3-12.

King, J. (2020, October 26). *Institutions offering degrees in education: 2009-10 to 2018-19* [Issue brief]. American Association of Colleges for Teacher Education.

King, J., & Mahaffie, L. (2016). *Preparing and credentialing the nation's teachers: The secretary's 10th report on teacher quality*. Office of Postsecondary Education, U.S. Department of Education. https://files.eric.ed.gov/fulltext/ED576185.pdf.

King, S. H. (1993). The limited presence of African-American teachers. *Review of Educational Research*, *63*(2), 115-149.

Kirkhart, K. E. (1995). 1994 conference theme: Evaluation and social justice seeking multicultural validity: A postcard from the road. *Evaluation Practice*, *16*(1), 1-12.

Kuenzi, J. J. (2018). *Teacher preparation policies and issues in the Higher Education Act* (CRS Report R45407, Version 3). Congressional Research Service. https://fas.org/sgp/crs/misc/R45407.pdf.

LaFrance, J., & Nichols, R. (2008). Reframing evaluation: Defining an Indigenous evaluation framework. *The Canadian Journal of Program Evaluation*, *23*(2), 13.

LaFrance, J., Nichols, R., & Kirkhart, K. E. (2012). Culture writes the script: On the centrality of context in indigenous evaluation. *New Directions for Evaluation*, *2012*(135), 59-74.

Lincove, J. A., Osborne, C., Dillon, A., & Mills, N. (2014). The politics and statistics of value-added modeling for accountability of teacher preparation programs. *Journal of Teacher Education*, *65*(1), 24-38.

Lindsay, C. A., & Lee, V. J. (2018, September 5). *Which colleges are helping create a diverse teacher workforce?* Urban Institute. https://www.urban.org/features/which-colleges-are-helping-create-diverse-teacher-workforce.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Monograph. RAND Corporation.

McFarland, J., Hussar, B., Wang, X., Zhang, J., Wang, K., Rathbun, A., Barmer, A., Forrest Cataldi, E., & Mann, F. B. (2018). *The condition of education 2018* (Report No. NCES 2018-144). National Center for Education Statistics, U.S. Department of Education. https://nces.ed.gov/pubs2018/2018144.pdf.

Measures of Effective Teaching Project. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study*. Bill & Melinda Gates Foundation. http://www.metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf.

Meyer, S. J., Brodersen, R. M., & Linick, M. A. (2014). *Approaches to evaluating teacher preparation programs in seven states* (Report No. REL 2015-044). Regional Educational Laboratory Central, U.S. Department of Education. https://ies.ed.gov/ncee/edlabs/regions/central/pdf/REL_2015044.pdf.

Moeller, K. (2020). Accounting for the corporate: An analytic framework for understanding corporations in education. *Educational Researcher*, *49*(4), 232-240. https://doi.org/10.3102/0013189X20909831.

National Academies of Sciences, Engineering, and Medicine. (2019). *Monitoring educational equity*. The National Academies Press.

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

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.

Peck, C. A., Singer-Gabella, M., Sloan, T., & Lin, S. (2014). Driving blind: Why we need standardized performance assessment in teacher education. *Journal of Curriculum and Instruction*, *8*(1), 8-30.

Petchauer, E., & Mawhinney, L. (2017). *Teacher education across minority-serving institutions*. Rutgers University Press.

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

Redding, C. (2019). A teacher like me: A review of the effect of student–teacher racial/ethnic matching on teacher perceptions of students and student academic and behavioral outcomes. *Review of Educational Research*, *89*(4), 499-535.

Reid, A. M., Boyce, A. S., Adetogun, A., Moller, J. R., & Avent, C. (2020). If not us, then who? Evaluators of color and social change. *New Directions for Evaluation*, *2020*(166), 23-36.

Rice, J. K., & Brent, B. O. (2002). An alternative avenue to teacher certification: A cost analysis of the pathways to teaching careers program. *Journal of Education Finance*, *27*(4), 1029-1048.

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

Ronfeldt, M., & Campbell, S. L. (2016). Evaluating teacher preparation using graduates' observational ratings. *Educational Evaluation and Policy Analysis*, *38*(4), 603-625.

Shriver, T. P., & Weissberg, R. P. (2020). A response to constructive criticism of social and emotional learning. *Phi Delta Kappan*, *101*(7), 52-57. https://doi.org/10.1177/0031721720917543.

Skinner, R. R. (2019, October 17*). The Elementary and Secondary Education Act (ESEA), as amended by the Every Student Succeeds Act (ESSA): A primer* (Report No. CRS R45977, version 2). Congressional Research Service. https://crsreports.congress.gov/product/pdf/R/R45977/2.

Smith, M. S., & Smith, M. L. (2009). Research in the policy process. In G. Sykes, B. Schneider, & D. N. Plank (Eds.), *Handbook of education policy research* (pp. 372-398). Routledge.

Sykes, G., & Dibner, K. (2009). *Fifty years of federal teacher policy: An appraisal*. Center on Education Policy.

Toldson, I. A. (2019). *No BS (Bad stats): Black people need people who believe in Black people enough not to believe every bad thing they hear about Black people*. Brill Sense. https://brill.com/view/title/54716?language=en.

Tuck, E., & Gorlewski, J. (2016). Racist ordering, settler colonialism, and edTPA: A participatory policy analysis. *Educational Policy*, *30*(1), 197-217.

U.S. Department of Education. (2011). *Our future, our teachers: The Obama administration's plan for teacher education reform and improvement*. https://www.ed.gov/sites/default/files/our-future-our-teachers.pdf.

U.S. Department of Education. (2013). *Preparing and credentialing the nation's teachers: The secretary's ninth report on teacher quality*. Office of Postsecondary Education.
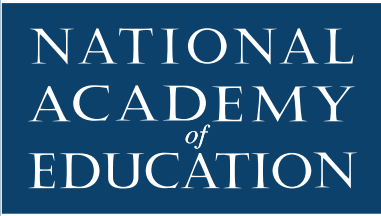
U.S. Department of Education. (2016a, October 31). Teacher preparation issues. 34 CFR Parts 612 and 686 [Docket ID ED-2014-OPE-0057] RIN 1840-AD07. Office of Postsecondary Education. Final regulations. *Federal Register*, *81*(210), 75494-75622.

U.S. Department of Education. (2016b). *The state of racial diversity in the educator workforce*. Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. http://www2.ed.gov/rschstat/eval/highered/racial-diversity/state-racial-diversityworkforce.pdf.

U.S. Government Accountability Office. (2015, July). *Teacher preparation programs: Education should ensure states identify low performing, programs and improve information sharing* (Report No. GAO-15-598). House of Representatives, Committee on Education and the Workforce, Subcommittee on Health, Employment, Labor, and Pensions. https://www.gao.gov/assets/680/671603.pdf.

Wells, A. S., & Roda, A. (2016). The impact of political context on the questions asked and answered: The evolution of education research on racial inequality. *Review of Research in Education*, *40*(1), 62-93.

Whittaker, A., Pecheone, R. L., & Stansbury, K. (2018). Fulfilling our educative mission: A response to edTPA critique. *Education Policy Analysis Archives*, *26*(30), 1-20.

Will, M. (2018, June 21). "An expensive experiment": Gates teacher-effectiveness program shows no gains for students. *Education Week*. https://www.edweek.org/teaching-learning/an-expensive-experiment-gates-teacher-effectiveness-program-shows-no-gains-for-students/2018/06.

Wilson, S., & Kelly, S. L. (2022). *Landscape of teacher preparation programs and teacher candidates*. National Academy of Education Committee on Evaluating and Improving Teacher Preparation Programs. National Academy of Education.

Wimberly, G. L. (2015). Use of large-scale data sets and LGBTQ education. In G. L. Wimberly (Ed.), *LGBTQ issues in education: Advancing a research agenda*. American Educational Research Association.

Zeichner, K. (2010). Rethinking the connections between campus courses and field experiences in college- and university-based teacher education. *Journal of Teacher Education*, *61*(1-2), 89-99. https://doi.org/10.1177/0022487109347671.

AUTHOR BIOGRAPHIES

Stafford L. Hood is the founding director of the Center for Culturally Responsive Evaluation and Assessment (CREA) and the Sheila M. Miller Professor of Education/ Curriculum & Instruction emeritus in the College of Education at the University of Illinois at Urbana-Champaign (UIUC). Hood's research and scholarly activities have focused primarily on the role of culture/cultural context in program evaluation and educational assessment and the contributions of African American evaluators during the pre-*Brown v. Board of Education* (1930-1954) period. For the past two decades, he collaboratively established CREA as an international and interdisciplinary community of researchers, scholars, and practitioners advocating the use of a culturally responsive lens in systematic inquiry across evaluation, assessment, policy analysis, applied research, and action research. Hood is a fellow of the American Educational Research Association (2016), a recipient of the American Evaluation Association's 2015 Paul F. Lazarsfeld Evaluation Theory Award, conferred an honorary appointment as an adjunct professor at Dublin City University (School of Education Studies) in Dublin, Ireland, in 2014, and is a fellow of the American Council on Education (2001-2002). His membership on many advisory boards and committees includes the Educational Testing Service's Visiting Panel for Research, the National Board for Professional Teaching Standards' Assessment Certification Advisory Panel, and the American Indian Higher Education Consortium's Building an Indigenous Framework for STEM Evaluation. He earned a B.A. in political science, an M.A. in counseling from the University of Wisconsin–Whitewater, and a Ph.D. in education (emphases program evaluation, administration, and policy analysis) from UIUC.

Mary E. Dilworth is a senior education policy and research advisor to nonprofit education organizations and institutions and the chair of the District of Columbia Higher Education Licensure Commission. Her work is keenly focused on matters of teacher quality and preparation, particularly as they intersect with race and ethnicity. Dilworth has a host of professional experiences that inform her work, including vice president for research and higher education at the National Board for Professional Teaching Standards and senior vice-president of the American Association of Colleges for Teacher Education (AACTE). She is a frequent contributor to national and state forums (e.g., the National Academies of Sciences, Engineering and Medicine and the Council of Chief State School Officers). She has written, edited, and contributed to scores of scholarly books, articles, policy, and research reports. She is the author of a chapter on the presence and absence of policies to diversify the teaching force for the upcoming *Handbook of Research on Teachers of Color* (Bristol & Gist) and the editor of *Millennial Teachers of Color* (Harvard Education Press), which is a recipient of the AACTE Outstanding Book of the Year. Dilworth holds and has held a number of elected and appointed positions on boards and commissions, including the American Educational Research Association, the Educational Testing Service, the National Education Association, the American Federation of Teachers, and the Ford Foundation. She earned a B.A. and an M.A. from Howard University and a doctorate from The Catholic University of America, each in the field of education.

Constance A. Lindsay is an assistant professor at the University of North Carolina at Chapel Hill. Lindsay earned a doctorate in human development and social policy from Northwestern University, where she was an Institute of Education Sciences' predoctoral fellow. Since leaving Northwestern, Lindsay has worked in education policy in various contexts, applying her research training in traditional studies and in creating and evaluating new systems and policies regarding teachers. Lindsay's areas of expertise include teacher quality and diversity, analyzing and closing racial achievement gaps, and adolescent development. Her work has been published in such journals as *Educational Evaluation and Policy Analysis* and *Social Science Research*. Lindsay received a bachelor's degree in economics from Duke University and an M.P.P. from Georgetown University. Before her doctoral study at Northwestern, she was a presidential management fellow at the U.S. Department of Education.

NATIONAL ACADEMY *of* EDUCATION

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

naeducation.org