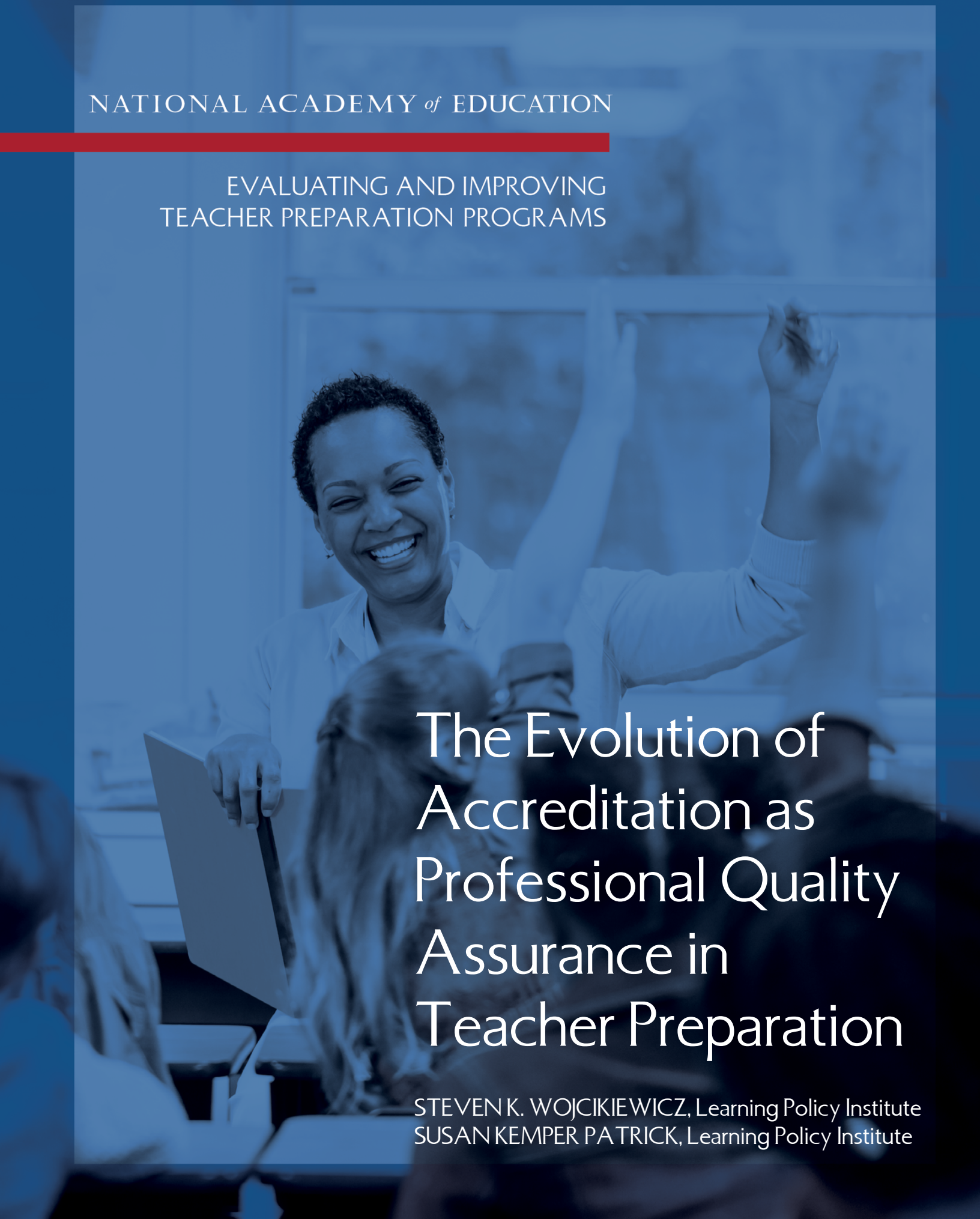
NATIONAL ACADEMY of EDUCATION

EVALUATING AND IMPROVING
TEACHER PREPARATION PROGRAMS



The Evolution of Accreditation as Professional Quality Assurance in Teacher Preparation

STEVEN K. WOJCIKIEWICZ, Learning Policy Institute
SUSAN KEMPER PATRICK, Learning Policy Institute

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

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

Digital Object Identifier: 10.31094/2021/3/8

Suggested citation: Wojcikiewicz, S. K., & Patrick, S. K. (2022). *The evolution of accreditation as professional quality assurance in teacher preparation*. National Academy of Education Committee on Evaluating and Improving Teacher Preparation Programs. National Academy of Education.

National Academy of Education

Evaluating and Improving Teacher Preparation Programs

The Evolution of Accreditation as Professional Quality Assurance in Teacher Preparation

Steven K. Wojcikiewicz, *Learning Policy Institute*
Susan Kemper Patrick, *Learning Policy Institute*

April 2022

CONTENTS

INTRODUCTION 2
A BRIEF HISTORY OF ACCREDITATION 4
The Beginning: Accreditation for Consistency and Quality in Teacher Preparation 4
A Nation at Risk: The Emerging Failure Narrative and Accountability Response 6
The 1990s: Reaching for Professionalization, Landing on Accountability 8
The Accountability Era, the Formation of CAEP, and the Push for Federal Regulation 10
A Shifting Landscape: CAEP and AAQEP Compete as Teacher Preparation Evolves 14
Recurring Themes in Teacher Preparation Accreditation 15
THE ROLE OF PERFORMANCE- OR OUTCOMES-BASED DATA IN THE ACCREDITATION PROCESS 16
Current Data Sources and Their Potential for Use in Accreditation 17
RECOMMENDATIONS FOR THE FUTURE OF ACCREDITATION 34
Limitations of Accreditation and Program Approval as Levers for Change 35
The Urgency of Action in Improving Accreditation 36
Recommendations to Refocus Accreditation and Data Use on Building the Profession and Improving Teacher Preparation 37
CONCLUSION 47
REFERENCES 48
AUTHOR BIOGRAPHIES 59

INTRODUCTION

In the field of education, voluntary national professional accreditation and mandatory state program approval are the primary means of assessing teacher preparation program (TPP) quality. Accreditation for U.S. educational institutions is carried out by nongovernmental professional organizations recognized and empowered by the federal government (Council for Higher Education Accreditation, 2006; Hail et al., 2019). Accreditors provide quality assurance for the public and a quality control mechanism by which professions can define and maintain standards (Hail et al., 2019; National Research Council, 2010). Historically, regional accreditors provided institution-level accreditation, while national accreditation aimed at profession- or subject-specific programs (U.S. Department of Education, 2021). A 2020 regulatory change allowing regional accreditors to operate nationwide eliminated geographic distinctions but retained separate institution- versus program-level processes (Middle States Commission on Higher Education, 2021; Reed, 2020; U.S. Department of Education, 2020).

In teacher preparation, national professional accreditation is a program-level, voluntary process (Council of Chief State School Officers, 2017). Before the creation of the first national TPP accrediting body in the 1950s, states and regional accreditors, along with the American Association of Colleges for Teacher Education (AACTE), formed a patchwork system of quality control (Vegari & Hess, 2002). Today, national accreditation is provided by two bodies: the Council for the Accreditation of Educator Preparation (CAEP) and the Association for Advancing Quality in Educator Preparation (AAQEP). It is possible, as Zeichner (2016) has pointed out, for independent TPPs to operate under regional accreditation (see, e.g., Alder Graduate School of Education, 2021), while some such programs have attained both regional and national accreditation (see, e.g., Relay Graduate School of Education, 2021), but this is unusual. Hereafter, we use the term "accreditation" to refer to national professional accreditation.

In contrast to accreditation, program approval is a state-run process by which TPPs are given permission to operate and to recommend program completers for state teacher licensure. Still, these processes do overlap in multiple ways. A few states require national professional accreditation as a condition for approval, while other states allow accreditation to substitute for state program approval (e.g., New York, North Carolina). In most cases, TPPs must be approved by state education agencies or standards boards but can also choose to pursue national accreditation. Even so, the overlap persists, as state-specific program approval standards may be based on, or even the same as, national accreditation standards. Indeed, some states refer to their program approval processes as state accreditation (Feuer et al., 2013).

Related to the overlap between accreditation and program approval for TPPs is the overlap between licensure and certification for teachers. Across professions, licensure refers to state authorization for basic professional practice while certification is a profession-granted acknowledgment of knowledge and skills (American Board of Medical Specialties, 2021; Lilly, 1992). In teaching, licensure and certification are often used interchangeably, blurring the distinction between state authorization and professional recognition. Furthermore, there are limited options for professional certification available to teachers aside from certification of accomplished practice through the National Board for Professional Teaching Standards (NBPTS) and a handful of licensure endorsements. In nursing, to give one contrasting example, certification is available for more

than 100 specialties (Kaiser Permanente Nursing, 2018). The lack of specialization and role differentiation for teachers affects TPP accreditation, focusing the debates about what teachers should know and be able to do at the level of professional entry. It also ensures that TPP accreditation standards must address a vast spectrum of knowledge and skills needed by educators, mostly concentrated in the single role of "teacher."

For all of the overlap that exists in terminology and process, program approval does stand apart from accreditation in several ways. First, there is significant variety in how program approval is authorized and carried out across states. Program requirements may be found in statute, administrative code, or agency rules. The entities responsible for program approval include chief education officers, state education agencies, boards of education, and standards boards (Council of Chief State School Officers, 2011; DeMonte, 2013). Program approval also plays a major role in state licensure because states grant TPPs the authority to recommend candidates for licensure. This gives TPPs a significant gatekeeping function in the field, or it would, were it not for the profusion of alternative routes allowing candidates to start teaching without completing a TPP (Gay & Daniel, 1972; Hess, 2006; Imig & Imig, 2008). In some states, alternative routes supply more than 50% of new teachers (National Education Association, 2021).

The role TPPs play in assessing candidate quality and recommending candidates for licensure is an example of an important feature of accreditation in education: it does not fulfill the same roles as accreditation in other fields. In medicine and nursing, and to a lesser extent in the legal profession, graduation from an *accredited* program is a vital step in professional entry and advancement. It is, with some exceptions, a requirement for sitting for a state or national exam, applying for state licensure, gaining access to employment, and participating in advanced training and certification (American Bar Association, n.d.; American Board of Medical Specialties, 2021; University of Nebraska Medical Center, 2021). In education, graduation from an accredited program carries no such weight. It is not even a requirement for National Board Certification, much less for employment. Among TPPs, the accreditation rate hovers just around 50%, unevenly split between the two national accrediting bodies. Meanwhile, numerous state-created alternative licensure routes allow entry into teaching without preservice preparation. While some such alternative programs are run by accredited TPPs, some are not. Still others are run by entities not categorized as TPPs, rendering accreditation irrelevant for new teachers prepared through these routes.

All of this creates a unique situation for the profession of education. As described in greater detail below, national accreditation for TPPs was originally intended to address deficiencies in state program approval and a perceived low quality of TPPs. It can be argued, however, that accreditation in education has never been fully connected to the levers of professional assessment, entry, and advancement that would make it possible to achieve these goals. Of course, labor market conditions also play a role, and their influence should not be discounted; teacher supply is connected to demand-side issues such as salaries and working conditions. When they arise, shortages of teachers put downward pressure on the standards and qualifications set by TPPs, states, and accreditors. Because teacher shortages have occurred regularly in the United States during the past century (Herbert, 2000; Sedlak & Schlossman, 1986), concurrent with concerns about the low pay and status of teaching, accreditation has always faced significant challenges to its founding goals of setting and maintaining TPP and educator quality.

For all that, accreditation still plays a valuable role as a profession-driven process. Accreditation functions through peer review. It is intended to enable internal program improvement. It also plays a role in defining education as a profession. As Arthur Levine (2006), former president of Teachers College at Columbia University, put it in his well-known (and perhaps infamous) critique of teacher preparation, "no field can be self-regulating until it has established high and explicit standards for itself, created a vehicle for enforcing them, and incorporated the highest quality institutions in its field as members and participants in peer review" (p. 110). Levine was critical of both the reach and rigor of TPP accreditation, yet he also cited the process as the most significant influence—even above TPP faculty and state education agencies—on TPP curricula.

In this paper, we trace the historical development of this professional standards setting and quality assurance process in teacher preparation. We then examine in depth a set of performance- and outcomes-based measures that have emerged in the past few decades as potential ways to assess TPP effectiveness and to support program improvement. Both the historical review of accreditation and more contemporaneous review of performance- and outcomes-based measures inform our recommendations for how TPP accreditors might more effectively incentivize, support, and provide assurance of teacher preparedness and effectiveness. Our goal is to present suggestions for improving accreditation that are well informed by lessons learned but also forward thinking, with an emphasis on advancing TPP improvement and strengthening the teaching profession. Our recommendations are also shaped by the awareness that, even as the past affects the present, context evolves, especially our understanding of the strengths and limitations of available evidence relating to TPP performance.

A BRIEF HISTORY OF ACCREDITATION

The Beginning: Accreditation for Consistency and Quality in Teacher Preparation

National TPP accreditation was born from dissatisfaction with state program approval. Launched in 1954, the National Council for Accreditation of Teacher Education (NCATE) was the original TPP accreditor. Intended to improve upon loose and inconsistent state oversight (Sedlak, 2008), NCATE was created by a collaboration of organizations representing teachers, teacher educators, and state agency officials: the National Education Association (NEA), AACTE, and the National Association of State Directors of Teacher Education and Certification (NASDTEC). These organizations had ambitious goals for NCATE even beyond replacing state program approval. They meant to improve the quality and consistency of TPPs nationwide and do so by enforcing national standards (Feuer et al., 2013; Sedlak, 2008; Tamir & Wilson, 2005).

By the early to mid-1960s, NCATE was accrediting about one-third of the nation's TPPs and was driving change in state program approval systems (Gubser, 1980; Moore et al., 1994; Tamir & Wilson, 2005). Yet, NCATE was already attracting the ire of multiple constituencies, accused of encroaching on faculty professional autonomy (Travelstead, 1963), even as it was critiqued for insufficient rigor. The National Commission on Accrediting (NCA), the earliest forerunner of the Council for Higher Education Accreditation (CHEA) (Collins & O'Brien, 2003), issued a report in 1965 that urged TPPs to forego NCATE accreditation and proposed that a replacement accreditor place greater

emphasis on the academic disciplines, national teacher exams, and supervisors' and employers' ratings of teachers (Accreditation Study Report, 1965; Accrediting commission advises members not to apply to NCATE, 1965). NCA's critique also exemplified the long-standing and ongoing conflict between TPPs and other academic departments at institutions of higher education (IHEs) by taking issue with NCATE requirements placing authority over teacher preparation in schools or colleges of education (Accreditation Study Report, 1965).

This academic infighting was, and is, related to the impulse within higher education to critique both teacher preparation and K-12 education. Such critique was on full display throughout the 1950s and 1960s in works such as *Educational Wastelands: The Retreat from Learning in Our Public Schools* (1953) by Arthur Bestor of the University of Illinois, *The Education of American Teachers* (1963) by James B. Conant of Harvard University, and *Anti-intellectualism in American Life* (1963) by Richard Hofstadter of Columbia University, as well as *Life* magazine's (1958) "Crisis in Education" issue (Wilson, 1958). The issues raised then remain familiar: schools do not focus enough on academics and are deficient in rigor; teachers do not have sufficient subject-matter training or hands-on experience; teacher preparation faculty are too distant from academic departments; and the sheer number of teachers needed precludes the development of a well-qualified workforce (Cochran-Smith & Zeichner, 2009; Herold, 1974). Many of these issues would find echoes in *A Nation at Risk: An Imperative for Educational Reform* by the National Commission on Excellence in Education (1983) and would be repeatedly emphasized in the ensuing focus on accountability in education policy.

The preview of later controversies provided by these early critiques is remarkable. Similarly, accreditation reform efforts in the mid-20th century established a pattern of development that re-emerged in similar forms in the following decades. First, the national accreditor, facing a fragmented field of uncertain quality, takes steps to establish a cross-state system of standards-based quality control. Next, the accreditor comes under fire for engendering stultifying centralization through standardization. Its value for TPPs is called into question by teacher educators. Soon after, or even at the same time, the accreditor is critiqued for insufficient rigor while the evidence base on which accreditation decisions are made also comes under attack. At this point in the pattern, replacement with a different accrediting body is a solution commonly proposed and occasionally implemented.

The 1970s brought both expansion and increasing critiques of accreditation, accompanied by continuing doubts about the efficacy of state program approval systems (Boyd, 1973). The expansion was impressive, if incomplete. By the early 1970s, NCATE was accrediting nearly 40% of TPPs and preparing three-quarters of the nation's teachers (Gay & Daniel, 1972). Still, the critiques leveled against NCATE inspired a series of reforms focusing on improving consistency and rigor. These led to a rise in the rate of initial denials for TPPs seeking accreditation to more than 30% (Gubser, 1980), though many of these programs were granted accreditation after a round of program revisions. Still, none of these changes headed off a comprehensive review of NCATE at the end of the 1970s that was driven in part by a threat from AACTE to create a new accrediting agency more responsive to its concerns (Imig & Imig, 2008; Moore et al., 1994; Wheeler, 1980). That review found that the NCATE process was useful for sorting out the lowest

performing programs and informing TPPs' internal improvements, but still concluded that its effectiveness was hobbled by vague standards that were inconsistently applied.

Indeed, NCATE came in for some harsh treatment. One prominent critic described NCATE's standards as "a mess" (Tom, 1980, p. 113) and questioned its basis in research. Elsewhere, NCATE was panned for high costs, overly long and complex accreditation reports, and innovation-stifling centralization along with too much focus on process and too little on candidate quality (Goodlad, 1990; Tamir & Wilson, 2005; Tom, 1980; Watts, 1986; Wisniewski, 1981). Recommended solutions included improved standards, stronger connections to state education agencies, and new links between accreditation and federal funding (Wheeler, 1980). Floden (1980), citing historical improvements in medical education, argued for increased involvement in accreditation by non-educators and the use of less structured observations and evaluations. Still, others maintained that the NCATE standards, as products of a broad and varied field, would always be complex and unfinished (e.g., Gubser, 1980; Wisniewski, 1981).

A Nation at Risk: The Emerging Failure Narrative and Accountability Response

The pressure brought to bear on NCATE in the late 1970s anticipated the era that followed the release of *A Nation at Risk* in 1983. By the 1980s, states had begun increasing teacher standards and expectations while experimenting with new models of preparation and professional advancement (Mitchell, 1989). The publication of *A Nation at Risk* marked the start of a movement toward accountability in education, one driven by a general loss of public confidence and a "profound disappointment" in the results of educational interventions launched in the 1960s (Fullan, 1994, p. 1), including those directed at recruiting teachers (Mitchell, 1989). The report established a narrative of failure, still a driving force in educational policy, through which teacher preparation was elevated to the front rank of public policy problems (Cochran-Smith, 2021; Mitchell, 1989).

The proposals for improving teacher preparation made at this time set the tone for later reform efforts. Market-based and deregulatory solutions were aligned to broader political movements, while competing proposals from within teacher preparation followed well-established traditional versus progressive lines (Cochran-Smith et al., 2017; Fraser & Lefty, 2018b; Imig & Imig, 2008; Mehta, 2013; Sedlak, 2008). Interest in alternative route preparation programs began expanding, finding early support in the Reagan administration's new emphasis on privatization while following the well-established pattern of prioritizing subject-matter knowledge over preparation from teacher educators (Cooperman & Klagholz, 1985). This era also included the birth of the current systems of licensure assessment through the movement toward "minimum competency" testing for teachers that was kicked off in the mid-1980s, with then-Governor Bill Clinton taking on a key leadership role (Clinton, 1986).

At this same time, reform movements from within education established their own priorities, calling for higher standards for entry and performance connected to new career ladder structures and increased professional autonomy (Darling-Hammond, 1985; Williamson et al., 1984). Such proposals, which included *Tomorrow's Teachers* by the Holmes Group (1986) and *A Nation Prepared* by the Carnegie Task Force on Teaching as a Profession (1986), advocated for professionalization, and these proposals gained

traction. NBPTS was founded in 1987 following the recommendations of the Carnegie Task Force. That same year, the Council of Chief State School Officers (CCSSO) commissioned the Interstate New Teacher Assessment and Support Consortium (INTASC) to create a set of standards for initial teacher licensure aligned to the NBPTS standards. NCATE carried out a major revision to its own standards in 1986-1987—an effort referred to as "the redesign"—with a specific focus on promoting research-based best practices integrated into a professional quality assurance system. The implementation of the new standards brought another spike in accreditation denials (Watts, 1986; Wise & Leibbrand, 1996). Still, by 1986, 80% of new teachers were being prepared by NCATE-accredited institutions, even though NCATE was only accrediting about 45% of TPPs (Fisher, 1986).

Even with this progress, some recommendations for building the teaching profession did not catch on. The creation of NBPTS provided for the recognition of advanced practice, but related proposals such as the implementation of teacher career ladders and differentiated roles for educators gained little traction (Fraser & Lefty, 2018b). Meanwhile, policymakers continued to move a competing agenda, tinkering with teacher preparation to address systemic education problems. By the end of the 1980s, three-quarters of the states had instituted licensure testing, two-thirds had added TPP entrance exams, and almost half had created alternative routes to licensure (Congressional Budget Office, 1993). As these policies were being put into place, teacher educators wrangled among themselves over questions of reform; even the oft-cited Holmes Group represented something of a factional agenda, rooted as it was in the dissatisfaction with NCATE and AACTE coming out of research-oriented universities.

Over time, the promotion of alternative routes—characterized by brief preservice preparation and the completion of TPP coursework, clinical experience, and state requirements during full-time teaching (Stoddert & Floden, 1995)—continued to gain popularity. While such routes can vary significantly, they almost always follow an on-the-job learning model in which candidates are employed as full-time teachers of record. This definition does not include school-based residency programs in which candidates work full time as apprentices in the classrooms of experienced teachers, but it does include emergency or temporary licensure programs. For our purposes, "alternative" describes the routes to teaching involving this sort of on-the-job TPP completion and fulfillment of licensure requirements.

Alternative routes are notable for their embodiment of the contradictory yet still popular tendency to raise standards for teacher preparation and licensure while simultaneously creating pathways into the classroom that bypass them. This has real effects on the teacher workforce. Research has shown that teachers who enter the profession with little or no preservice preparation, such as through alternative routes, show higher rates of turnover and lower classroom effectiveness than comprehensively prepared teachers (Carver-Thomas, 2018; Carver-Thomas & Darling-Hammond, 2017; Podolsky et al., 2016). Furthermore, underprepared teachers who enter teaching through such routes are concentrated in schools serving the highest need students and students of color (Cardichon et al., 2020). Meanwhile, evidence of the importance of high-quality preservice clinical experiences—with features such as guided practice, frequent feedback, experienced mentors, and coordinated coursework—continues to mount (Hollins

& Warner, 2021; Ronfeldt, 2021), highlighting what is missing from many alternative route programs.

This tendency has also undermined policymakers' efforts to increase rigor in teacher preparation while complicating efforts from within education to professionalize teaching. Proposals to raise the bar for TPP accreditation and teacher licensure, also aimed at increasing the effectiveness, status, career options, and professional autonomy of teachers, were not made in a vacuum. All of these proposals were predicated on the assumptions that better pay and working conditions for teachers would enable the implementation of higher standards and that these standards would not be undermined by the creation of alternative routes (Darling-Hammond, 1985; Goodlad, 1990; Williamson et al., 1984). It is not coincidental that alternative routes have continued to proliferate amid teacher shortages driven by low pay and unattractive working conditions.

The 1990s: Reaching for Professionalization, Landing on Accountability

The 1990s saw the continuation of multiple approaches to raising teacher quality during which accreditation increasingly became aligned with the "professionalization agenda" (Zeichner, 2019). The decade brought a fresh batch of media reports calling into question the qualifications and quality of the nation's teachers (Feuer et al., 2013), while state programs continued to be characterized as inadequate (Jones, 1991). A related spirit carried over into federal policy as the 1992 reauthorization of the Higher Education Act (HEA) added new requirements for states and TPPs to assess candidate quality, though HEA also supported the parallel drive for professionalization by providing funding for INTASC and NBPTS (Bales, 2015). It was during the 1990s that NBPTS began certifying teachers, INTASC issued its first set of standards, and both efforts saw broad uptake at the state level (Interstate New Teacher Assessment and Support Consortium, 1992; National Board for Professional Teaching Standards, 1992). In 1990, Arthur Wise was appointed president of NCATE, signaling increasing alignment with the professionalizing agenda given his teacher quality work at the RAND Corporation (Viadero, 1990; Wise & Verstegen, 2000; Zeichner, 2019). The National Commission on Teaching & America's Future (NCTAF), a nonpartisan advocacy group founded in 1994, proposed an agenda for the education profession resting on a "three-legged stool" of accreditation, licensure assessment, and advanced certification in a developmental continuum emphasizing performance-based assessment (Darling-Hammond, 1999; Sykes & Plastrik, 1993; Wise & Verstegen, 2000).

NCTAF's recommendation that all TPPs seek national accreditation in lieu of state program approval brought NCATE further into the fold of reform by professionalization (Moore et al., 1994; Ponessa, 1997). The developmental, standards- and performance-based professional system embodied by the NCTAF agenda gave accreditation an important role: establishing the foundations of the profession, which included "a shared body of knowledge, based on research, and public confidence that professionals are fit to practice" (Wise, 2005, p. 319). This systemic agenda was, in some ways, a refashioning of the original 1950s mission of NCATE to bring TPPs across the country in line with national standards, and NCATE made these connections clear at the time. NCATE issued newly revised standards in 1995, aligning them to the INTASC standards and strengthening the links between accreditation and licensure reform (Darling-

Hammond, 1999; National Council for Accreditation of Teacher Education, 2008; Wise & Leibbrand, 1996).

It was also in this round of revisions that NCATE adopted the same "performance-based standards" approach taken by INTASC (Paliokas et al., 2011, p. 37), which requires TPPs to use performance assessments and to view preparation in terms of competencies and outcomes (Bradley, 1997; Wise & Leibbrand, 1996). The performance-based standards movement also translated into changes in state policy across the country, showing the most success in states such as North Carolina and Connecticut, which combined the more widely adopted standards-based licensure and program approval and performance-based assessment policies with investments to raise teacher salaries (Darling-Hammond, 1997; Wojcikiewicz & Darling-Hammond, 2020). Proficiency-based program approval was not new, having been pioneered in the 1970s alongside competency-based K-12 instruction and—with federal support—teacher preparation (Gay & Daniel, 1972; Roth, 1974), but the systemwide approach of the 1990s built a new professional infrastructure to support standards- and performance-based TPPs.

In that moment, NBPTS, INTASC, and NCATE seemed to be riding on the same professionalization wave together (Imig & Imig, 2008; Wise & Leibbrand, 1996). By 1997, NCATE had established partnerships with 41 states and was effectively handling program approval for 10 of them; 20 states were using NCATE standards for all TPPs, even those not nationally accredited; and 9 states had switched their program approval systems to a competency-based approach (Darling-Hammond, 1999; National Research Council, 2001; Ponessa, 1997). Approximately 40% of all programs were nationally accredited. In 1997, NCATE began another standards review process, maintaining its previous efforts to align with INTASC and NBPTS (National Council for Accreditation of Teacher Education, 2002).

And yet, as in previous decades, dissatisfaction was growing alongside success. NCATE's increasing connection to the professionalization movement and fairly broad support within the teacher education community at this time did not mean the accreditor was spared periodic complaints about high costs, complex processes, and inadequate standards (Gardner et al., 1996; Gideonse, 1993; Sykes & Plastrik, 1993). Some saw the NCATE process not only as too prescriptive but also unsuitable for small liberal arts programs and large, research-intensive universities (Feuer et al., 2013; Goodlad, 1990; Hasbun & Rudolph, 2016; Murray, 2005). Other detractors pointed to centralized, standards-based reform as monopolistic, an unnecessary limit on educator supply, or as oppressive (Hail et al., 2019; Tamir & Wilson, 2005). And so, in 1997, the long-threatened creation of a rival accreditor finally came to pass with the formation of the Teacher Education Accreditation Council (TEAC).

TEAC quickly became the go-to accreditor for the research-intensive universities and liberal arts colleges most dissatisfied with NCATE's approach (Hasbun & Rudolph, 2016). The creation of TEAC also highlighted a continuation of the traditional versus progressive divide in teacher preparation, which emerged in this context as centralized and standardized versus locally defined and inquiry-based processes (Loews & Sawchuk, 2017), or as "investigative" versus "collaborative" approaches (Murray & Wise, 2009, p. 3). The TEAC approach centered on program-level inquiry and required TPPs to provide evidence of their capability to maintain quality (Loews & Sawchuck,

2017; Murray, 2005; Will, 2021; Wise, 2005). In contrast to NCATE, TEAC allowed TPPs, within certain boundaries, to set their own standards (Feuer et al., 2013). This opened up TEAC to accusations of being overly lenient and derailing efforts toward shared professional accountability (Tamir & Wilson, 2005), though TEACs' defenders argued that local knowledge production was no less professional than centralized standardization (e.g., Murray, 2005).

As this accreditation drama played out, the transition from the Reagan-Bush years to the Clinton administration did not interrupt the movement toward market-based accountability in education. Indeed, the Clinton administration played a key role in shifting to policy-driven accountability and away from the professionalization agenda (Imig & Imig, 2008), though it should be noted that the administration did act to preserve federal funding for NBPTS (Moore, 2002). Still, it was the 1998 reauthorization of HEA that established the Title II national reporting system still in place today (Feuer et al., 2013). Along with expanding the federal role in educator preparation accountability, the 1998 reauthorization also funded the expansion of alternative routes to licensure (Imig & Imig, 2008).

Even though the "centralized-localized" (Imig & Imig, 2008, p. 894) professional system that included NCATE, NBPTS, and INTASC seemed to be gaining traction throughout the 1990s, it was soon pushed aside, along with any possibility that NCATE might grow along with it. Even the success of state-level policies in the 1980s and 1990s, in which a combination of higher standards for teachers with greater investments in teacher preparation and the workforce both raised student achievement and narrowed achievement gaps (Darling-Hammond, 2019; Wojcikiewicz & Darling-Hammond, 2020), could not interrupt the failure narrative. And so, just as the major reform push of the 1980s ignored progress in the previous decade, the 1990s were followed by the complete overshadowing of professionalization in education, and in accreditation, by the passage of the federal education policy known as No Child Left Behind (NCLB).

The Accountability Era, the Formation of CAEP, and the Push for Federal Regulation

The passage of the No Child Left Behind Act of 2001 solidified the ascendance of test-based accountability as a central feature of education policy. That year, NCATE was accrediting around 40% of the nation's TPPs, preparing 70% of the nation's teachers, and these numbers were on the rise (Vegari & Hess, 2002). The next year, Secretary of Education Rod Paige put out a scathing report that drew on HEA Title II data to claim that "schools of education and formal teacher training programs" were "failing to produce … highly qualified teachers" in accordance with NCLB (U.S. Department of Education, 2002, p. 493). Even though the report emphasized that NCLB would eliminate the hiring of underqualified teachers, the concentration of these teachers in high-poverty schools, and the widespread use of alternative routes to licensure, none of this signaled support for TPPs. Indeed, the report described alternative routes as avoiding "the burdensome requirements of the traditional system" (p. 494), and, in describing teacher preparation, emphasized the importance only of subject-matter coursework.

The Clinton administration also backed this rhetoric with action. A new test-only alternative route to licensure, the American Board for Certification of Teacher Excel-

lence, was launched in 2001 by the also new National Council on Teacher Quality (NCTQ), with $5 million in support from the U.S. Department of Education (ED) (Thomas B. Fordham Foundation, 2002). Then, the requirement for teachers to be "highly qualified," a key NCLB feature, was undercut when the law was implemented. ED allowed alternatively certified teachers working toward full licensure to meet the Highly Qualified Teacher requirements, and this workaround remained in place, in one form or another, until NCLB was supplanted by the next reauthorization of the Elementary and Secondary Education Act (Education Next, 2010; Ziechner, 2013).

Meanwhile, the criticisms of TPPs and their accreditors were piling up from other sources. In 2004, the Superintendent of Boston Public Schools, who later played a role in the founding of the Boston Teacher Residency, gave an address at the AACTE annual meeting entitled "Should Teacher Preparation Take Place at Colleges and Universities?" (Zeichner & Paige, 2007). In 2005, *The New York Times* article titled, "Who Needs Education Schools?" claimed that TPPs faced "pressure to improve from all directions" for their "ideological bias and low admissions standards" (Hartocollis, 2005). Levine's scathing critique of teacher preparation institutions, published in 2006, also took TPP quality assurance systems to task. "Neither the states nor the accreditation process," he claimed, "has been able to assure minimum quality standards in teacher education programs" (Levine, 2006, p. 22). Remarkably, traditional teacher preparation came under attack just as the idea that teachers are the most important in-school factor in student learning was becoming a widely discussed research finding (e.g., Darling-Hammond & Bransford, 2005).

Even though Levine, like the field's reform-minded supporters of previous decades, indicated that higher salaries would be key to higher standards, this received less attention than other aspects of his report. Test-based accountability was "in" as the solution to teacher quality issues. The 2008 HEA reauthorization added more measures of TPP performance, including average licensure test scores (Feuer et al., 2013). Even NCATE took on some of the language and spirit of the times, with the 2008 NCATE standards document claiming that "accountability and improvement in teacher preparation are central to NCATE's mission" (National Council for Accreditation of Teacher Education, 2008, p. 1) and that "education reform must include the reform of teacher preparation" (p. 3).

The transition from the second Bush administration to the Obama administration in 2009 recapitulated the Bush-Clinton transition of the 1990s. While many changes swept through national politics, this did not greatly affect the status of teacher preparation in the eyes of policymakers. Secretary of Education Arne Duncan (2009), speaking at Teachers College at Columbia University, claimed that "many if not most, of the nation's 1,450 schools, colleges, and departments of education are doing a mediocre job of preparing teachers," and indicated the Obama administration's readiness to change that. Its first attempt was through the Race to the Top (RTT) grant program in 2009, which included incentives for states to evaluate educator preparation programs using student achievement data (Feuer et al., 2013). RTT also incentivized alternative preparation, carrying on once more with the contradictory process of simultaneously raising and lowering TPP standards (Weiss, 2013). While only 19 states received RTT funding, 34 modified their policies to meet RTT goals (Crowe, 2011; White House Archives of President Barack Obama, n.d.).

From 2010 onward, the pace of change in educator preparation accreditation accelerated rapidly. NCATE and TEAC were, together, accrediting more than half of TPPs when it was announced in 2010 that these rivals would merge into a single organization, CAEP. The formation of CAEP was meant to bring unity, and thus greater status, to the profession, while also holding TPPs to higher standards, incentivizing their internal improvement efforts, and increasing the focus on outcomes and evidence of teacher candidate and K-12 student learning (Council for the Accreditation of Educator Preparation, 2013; Feuer et al., 2013; Hail et al., 2019). CAEP's new system relied on an intersecting set of assumptions about measuring program quality: that program and completer impact and outcomes were more important than inputs or program processes; high-quality impact data on program completers and their students would be available and applicable; and outcomes-based accountability would spur program improvement (Cochran-Smith et al., 2017; Hasbun & Rudolph, 2016; Imig & Imig, 2008; Will, 2018; Worrell et al., 2014).

CAEP soon attracted attention from supporters and critics. Proponents explicitly linked more rigorous accreditation, improved TPPs, improved K-12 student performance, and reduced achievement gaps (e.g., Worrell et al., 2014). CAEP was seen as an opportunity for TPPs to restore public confidence and even to reverse a developing trend of declining enrollment (DeMonte, 2013; Will, 2018). Levine (2015) portrayed CAEP as an opportunity for teacher preparation to prove itself capable of self-regulation. The merger was met with significant anticipation, and at its founding CAEP had already lined up partnerships with 27 states (Cochran-Smith et al., 2017; Loews & Sawchuk, 2017). Meanwhile, even more change was under way as the 2011 revision of the InTASC standards (Council of Chief State School Officers, 2011) moved the field further along toward a focus on performance-based standards.

As CAEP was ramping up, teacher preparation was increasingly being portrayed as a field in need of reform. A decade after the passing of NCLB, a lack of progress in raising student achievement or closing achievement gaps created a demand for scapegoats. TPPs fit the bill. As a report from the Center for American Progress (CAP) put it, "weak teacher-preparation programs," were "key to the failure of public education to improve instruction for all students" (DeMonte, 2013). As had occurred a decade earlier, educator preparation was once again taken to task in a series of reports. This time the authoring organizations were different, including CAP, CCSSO, the American Federation of Teachers, and the American Psychological Association. NCTQ, with a decade of experience criticizing TPPs and promoting alternatives like American Board for the Certification of Teacher Excellence, partnered with *U.S. News & World Report* to release ratings—viewed with great skepticism at many TPPs—of teacher training institutions.

With this narrative in place, policymakers acted accordingly. Across the country, states raised program approval standards, efforts in which CCSSO played a role (Council of Chief State School Officers, 2011; Worrell et al., 2014). The pressure was also mounting from alternative route programs, which were still held to different standards from traditional, IHE-based TPPs already facing years of declining enrollment (Imig & Imig, 2008). In 2011, the Obama administration made another move targeting educator preparation by setting in motion the process for promulgating new regulations for HEA Title II. These new regulations would, in the manner of RTT, link the evaluation of TPPs to the test scores of students taught by program completers. Even though the

political momentum was on the side of bringing NCLB-style accountability to bear on educator preparation, the move was controversial, and the initial rulemaking process stalled out in 2012.

When the CAEP standards were rolled out in 2013, their focus on outcomes, and particularly on teacher effectiveness data, also drew considerable criticism. Linking accreditation to the test-based accountability of the NCLB era was an unpopular move by itself, and it was even more unpopular in combination with data availability and analytical capacity issues (Croft et al., 2015; Farenga & Ness, 2017; Sawchuk, 2016; Will, 2019). Critics claimed that K-12 student outcome measures were not applicable to TPPs and that putting weight on such measures negatively impacted TPPs preparing educators to work in the highest needs schools. TPPs complained about their difficulties in acquiring outcomes data on their graduates' performance and pointed out that, even when the data were available, the effort required to analyze them would exceed TPPs' available resources. A new 3.0 GPA admissions standard also raised controversy, but this standard was quickly changed due to its anticipated effects on TPP enrollment and the diversity of the teaching force (Evans, 2017; Sawchuk, 2016), as well as its requirement that programs be assessed based on the outcomes achieved by their graduates.

Though the new standards were not universally opposed—the imperative to collect data was, in one example, portrayed as a way of incentivizing TPP–district collaboration (Goodson, 2018)—the declining popularity of NCLB was driving a tendency to cast all outcomes measures in a negative light. Still, the Obama administration pressed ahead, restarting rulemaking for HEA Title II teacher preparation accountability regulations in 2014. The regulations, representing a vast increase in the federal role in educator preparation accountability, were widely unpopular (Cochran-Smith et al., 2017; Flores, 2015; Sawchuk, 2016; VanHouten, 2015), largely because of the link to test-based teacher evaluation systems that the administration was also seeking to incentivize across the states. The regulations were also taken to task for asking states and TPPs to do too much too quickly without regard for data quality and access while allowing alternative route programs to meet a lower expectation (Flores, 2015; Tatto et al., 2016; Weingarten, 2015; Weiss, 2013).

These regulations were, however, supported by the CAEP leadership. This support caused significant blowback from TPP leaders and faculty, many of whom were also up in arms about the new CAEP standards. This blowback was a contributing factor in CAEP's 2015 leadership change, leaving the new accreditor with its second president before it had issued its first accreditation decision (Burns, 2016; Erickson, 2015). The controversies attending the launch of CAEP were important, not only because they were indicative that the views of TPP leaders and faculty were out of line with the prevailing political consensus around TPP accountability, but also because these controversies were generated by the close relationship between the sole accreditor of TPPs and governmental regulatory systems. Accreditors are recognized as empowered by governmental bodies, but accreditation is not the same as regulation. CAEP's method of addressing calls for the reform of teacher preparation was to intertwine the profession-driven quality assurance and improvement process of accreditation with emerging test-based accountability mandates. The results of this response are still working themselves out today.

A Shifting Landscape: CAEP and AAQEP Compete as Teacher Preparation Evolves

More change and instability were in store for accreditation and TPPs. In 2016, the new HEA Title II regulations mandating outcomes-based accountability for TPPs were put into effect and CAEP began accreditation reviews (Kreighbaum, 2016; Sawchuk, 2016). It seemed that a new and comprehensive accountability regime had arrived, but it did not even last 1 year. In March 2017, the HEA Title II regulations were rescinded by the U.S. Senate. Then, in the fall of that year, came the news that a brand new accreditor was being launched. This new accreditor, AAQEP, was created in part as a response to CAEP's controversial promises of outcomes-based rigor, though CAEP also suffered from lingering organizational issues that led to confusing and inconsistent guidance that further fed dissatisfaction at TPPs (Groves, 2019).

AAQEP initially launched with a small staff, the majority of whom had worked for TEAC. Relying on committees of volunteer TPP faculty members, AAQEP built a collaborative, inquiry-focused accreditation process designed to account for the unique contexts of individual TPPs (Association for Advancing Quality in Educator Preparation, 2021a; Cochran-Smith & Reagan, 2021; Will, 2019). This process was based on an entirely new set of standards, independent from, and fewer in number than, the CAEP standards. The AAQEP process was set up to require TPP engagement in articulating the warrant for evidence provided during the accreditation process, another feature that distinguished it from CAEP, though it echoed TEAC's approach. AAQEP emphasized innovation and demonstrated a commitment to collaborative inquiry by placing TPPs seeking accreditation into cohorts that could work together on improvement processes (Association for Advancing Quality in Educator Preparation, 2021b).

Even as a brand new organization, it did not take long for AAQEP to end up on the receiving end of criticism; indeed, the criticism began right away. AAQEP's creation was troubling to some because it seemed to defeat the purpose of the NCATE and TEAC merger. The establishment of a single, rigorous accreditation process based on high standards was, it was claimed, particularly needed as teacher shortages put downward pressure on standards and quality (Will, 2019). The defenders of AAQEP, in turn, painted CAEP's accreditation process as top-down, rigid, and compliance-based while questioning its heavy emphasis on outcomes measures. In this back and forth there were many echoes of past battles, with CAEP standing in for the centralized, standardized, overly complicated NCATE, and AAQEP taking on the role of TEAC as the outsider threatening professional unity, rigor, and status (Ponessa, 1997; Will, 2019).

At the time that this paper goes to publication, the status of educator preparation accreditation, and educator preparation generally, remains unsettled. As of 2021, AAQEP has accredited 44 programs and claimed a total of 160 programs, located in 27 states, as members (Association for Advancing Quality in Educator Preparation, 2021a). CAEP has accredited a total of 423 programs since 2016, including 60 programs in 2021 (Council for the Accreditation of Educator Preparation, 2021). As CAEP's new standards were under review in 2020-2021, a process meant to improve clarity and relevance, AAQEP received recognition by CHEA, opening doors to partnerships with a larger pool of states (Will, 2021). Amid all of this, enrollment in TPPs has been declining for more than a decade while pressure grows from superintendents who need to fill

classrooms and alternative routes to licensure continue to proliferate (Fraser & Lefty, 2018a; Partelow, 2019). Even before the onset of the COVID-19 pandemic, states were lowering licensure requirements to meet these shortages (Will, 2019). Now, as the pandemic may be accelerating teacher turnover, the pressure on teacher preparation and TPP accreditors is mounting.

Recurring Themes in Teacher Preparation Accreditation

The historical development of educator preparation accreditation spans nearly seven decades and encompasses vast changes in education and in the world. This complicates any effort to locate recurring themes in this development, because the danger of oversimplification is real. At the same time, such themes have great potential value in informing current conversations on the future of accreditation, if only to provide some notion of what has already been tried—or not tried. These themes reveal how broader policy trends and imperatives, in education and beyond, have shaped both the perception and regulation of TPPs.

The existence of such linkages is, in fact, a theme itself. Key moments in the development of TPP accreditation were shaped by the status of education according to the public, the media, and policymakers. Teacher preparation has long struggled under the perception of low quality and lack of rigor, going back to the founding of NCATE in the 1950s. The narrative of failure, which took hold in the 1970s and took off in the 1980s, has since spawned decades of reactive policy though it has also spurred improvement efforts from within the profession. The implementation of market- and outcomes-based accountability alongside the adoption of professional standards- and performance-based assessments make it difficult to point to one set of policies or a single trajectory of improvement on which to build. This history should, however, put the field on notice that today's education issues are likely to become tomorrow's teacher preparation issues.

When calls for educational reform do arise, changes to TPP accreditation are one common response. Increasing centralization through accreditation has been a trend since the creation of NCATE in 1954, though in recent years it has included a far greater federal role. Alongside this centralization has been the countervailing pressure to resist standardization, an issue linked in some part to questions of faculty autonomy. Both moves toward centralization and questions about autonomy have certainly played into the creation of rival accreditors. Still, however the field and the public might feel about national-level quality assurance for TPPs in any particular historical period, the general drift has been toward rather than away from centralization.

A similar drive can be seen in the concept of outcomes-based assessment of TPPs. The hints of a movement toward outcomes popped up periodically in the first four decades of NCATE's existence, and then gained a great deal of momentum in the 1990s and 2000s. Even though policymakers and educators had a variety of views of just what "outcomes" meant, there was broad acknowledgment of the general concept. Over time, it has come to be normalized, though not without resistance. Unlike some of the other themes that play a frequent role in conversations about accreditation without being settled one way or the other—such as questions of rigor, specificity, and the knowledge base for teaching—it seems more likely that the long-developing move

toward accreditation based on outcomes is here to stay. However, as we describe in detail below, debates about which outcomes should be used, and how, do not lend themselves to simple answers.

Accreditation does emerge in a consistently positive light in at least one sense, namely, in comparison to state-level program approval. Stinging assessments of state processes can be found across the decades. These critiques often make the point that states face a great deal of pressure to avoid closing underperforming TPPs, or indeed any TPPs, and that state agencies are often under-resourced for this function. It should be considered, however, that seemingly inconsistent state actions regarding TPP and teacher quality are driven by labor market pressure. Demand for teachers seldom goes down. Even when salaries and working conditions do not attract an adequate number of trained educators, classrooms cannot be left empty. Additionally, it is important to note that the persistent negative characterization of state program approval hides real progress and innovation, both past and present. Shifts to competency- and performance-based systems in the 1970s and 1990s laid the groundwork for contemporary state shifts toward performance-based program approval systems. Even now, states are wrestling with how to balance outcomes, quality, and improvement in ways that offer important lessons for accreditors. Although our focus is on accreditation, it is important to consider the historical and contemporary relationship between accreditation and program approval.

THE ROLE OF PERFORMANCE- OR OUTCOMES-BASED DATA IN THE ACCREDITATION PROCESS

The current orientation toward outcomes- and performance-based assessment in teacher preparation has deep roots. Efforts to measure performance and outcomes for teaching candidates and TPPs have been proposed and even implemented at multiple points in the past 60 years. For example, NCA proposed that a new accreditor survey TPP completers and their employers in the 1960s. Competency-based teacher preparation and licensure were implemented from the late 1960s into the 1970s (Nodine, 2016). The Teacher Work Sample, a performance assessment requiring a cycle of instruction and reflection, was adopted by Oregon in the mid-1980s, then picked up in the 1990s by a consortium of TPPs spanning 10 states (Giovannetti, 2012; Schalock & Schalock, 2011). The development of later performance assessments such as edTPA and the Praxis Performance Assessment of Teachers (PPAT), also described below, established them in licensure and program approval systems across the county.

Even with such roots, outcomes-based assessment systems, applied to teachers and TPPs, have spurred significant controversy. Shifts toward standards- and performance-based systems in the 1990s ran in parallel to the nationwide movement toward accountability based on standardized testing, which culminated in the passage of NCLB. RTT and the 2016 HEA Title II regulations brought NCLB-style, outcomes-based measurements to bear on TPPs, an approach also embraced by CAEP. The punitive use of outcomes to compel improvement was at odds with advocacy for professional standards and performance assessments (Darling-Hammond, 2007). Still, these links to NCLB-style accountability have caused outcomes-based quality assurance efforts

in general to be viewed with deep suspicion by TPP faculty (Croft et al., 2015; Sedlak, 2008; Strauss, 2015).

Despite these challenges, outcomes- and performance-based data are likely here to stay. Evidence of completer performance and program outcomes is required by both national TPP accreditors. CAEP's annual reporting includes four impact measures and four outcome measures while AAQEP's annual reporting includes multiple measures addressing candidate and completer performance. Most states assess TPPs using at least some outcomes data, although the type of data used in these processes vary widely not only across states but also by program type, differing for "traditional" (i.e., preservice) TPPs versus alternative (i.e., inservice) options (Fenwick, 2021). In a 2015 Governmental Accountability Office (GAO) survey, 49 states reported reviewing data about traditional TPPs for program approval, while only 43 states reported reviewing data about alternative programs.

Given this widespread commitment to integrate performance- and outcomes-based data into program approval and accreditation, additional attention is needed to carefully consider the capabilities, strengths, and limitations of various methods of collecting, analyzing, and using such data. The specific measures typically used by accreditors and states reflect the history of accreditation and the larger trends in education policy, such as the standardized test–based accountability movement to assess schools and teachers. A host of researchers and organizations have put forward recommendations about how to incorporate outcome measures into the evaluation of TPPs (for more on these recommendations, see the review by Cochran-Smith & Reagan, 2021), and our goal is not to supplant these. Rather, we aim to focus on the applications of data use specific to TPP accreditation and, to a lesser extent, program approval.

As our analysis shows, performance- and outcomes-based data should be neither unquestioningly applied nor summarily rejected for quality assurance and continuous improvement purposes. Notably, certain measures (i.e., knowledge-based teacher licensure exams, student test scores) have been developed for other accountability purposes and then appropriated into evaluations of TPP effectiveness. Other measures (i.e., completer and employer surveys, labor market outcomes) are explicitly collected for the purpose of evaluating TPPs. We discuss the use of both types of measures for TPP accreditation and program approval.

Current Data Sources and Their Potential for Use in Accreditation

Although specific data sources vary across states and accreditors (Fenwick, 2021; Government Accountability Office, 2015), there are similarities across TPP assessment contexts, including CAEP and AAQEP requirements (see Table 1). CAEP requires four impact measures that capture completers' performance when they are working as educators (i.e., student achievement scores, observation or student survey ratings, completer satisfaction surveys, employer satisfaction surveys) and four outcome measures that assess candidates' experiences in their TPP or immediately after completion (i.e., licensure test passing rates, graduation rates, employment rates, student loan default rates). CAEP gives discretion to TPPs to determine exactly how to measure each type of impact across the mandated categories. AAQEP, by contrast, provides more flexibility in

TABLE 1 Performance-Based/Outcomes-Based Evidence Requirements for CAEP and AAQEP Accreditation Processes

Panel A: CAEP's Performance-Based/Outcomes-Based Reporting Standards (Council for the Accreditation of Educator Preparation, 2020a

Standard	Suggested Measures
4.1 Impact on P-12 Student Learning and Development	Direct measures of student achievement or growth for students of program completers who are now employed as teachers. These measures can be drawn from state or district data, such as: • Student learning, growth, or value-added measures linked to state teacher evaluations • Student growth measures from state- or district-collected assessments (e.g., Northwest Evaluation Association's [NWEA's] MAP assessments) • Case studies using TPP-created measures
4.2 Indicators of Teaching Effectiveness	Measures capturing the professional knowledge, skills, and dispositions of completers who are now employed as teachers through: • Scores from validated observation instruments • Student perception surveys such as those conducted in the Measures of Effective Teaching study
4.3 Satisfaction with Preparation as Viewed by Employers	Measures capturing employer satisfaction or employment outcomes for completers who are now employed as teachers, such as: • Surveys, interviews, focus groups, or case studies with employers • Employment rates in high needs schools • Retention or promotion rates
4.4 Satisfaction with Preparation as Viewed by Completers	Measures capturing completers' perception of their preparation and its relevance to their responsibilities on the job. CAEP suggests that questions are most relevant when they focus on particular aspects of preparation and relate back to specific benchmarks or norms. Perceptions can be gathered through: • Surveys, interviews, focus groups, or case studies with completers
5.4 Continuous Improvement	Four outcome measures are required as part of CAEP's annual reporting process and, along with the impact measures above, are intended to inform program improvement: 1. Graduation rates (i.e., percent of candidates who completed the program over a specified time period) 2. Ability of completers to meet licensing (certification) and any additional state requirements (i.e., Title II data or other state- or TPP-collected data on certification exam passing rates) 3. Ability of completers to be hired in education positions for which they have been prepared (i.e., employment rates of completers) 4. Student loan default rates and other consumer information (i.e., institutional student loan default rate as reported by the U.S. government and any additional cost information supplied by TPP)

continued

TABLE 1 Continued

Panel B: AAQEP's Performance-Based/Outcomes-Based Reporting Standards (Association for Advancing Quality in Educator Preparation, 2021b)

Standard	Suggested Measures
1. Candidate/Completer Performance	Multiple measures of candidate and completer knowledge and ability, including direct performance measures that address this question: At the end of the program, are completers ready to fill their target professional role effectively? These measures must include data from multiple perspectives, including program faculty, P-12 partners, program completers, and completers' employers. TPPs must include direct measures of candidate performance in a field or clinical setting appropriate to the program. Suggested data sources include: • Grades in content, pedagogical, and professional courses • Licensing or certification examination results • Observations and summary ratings in field placements or internships • Performance assessment results • Survey, interview, or focus group data from completers, cooperating teachers, P-12 employers
2. Completer Professional Competence and Growth	Multiple measures of completers' effectiveness when employed in their professional roles meant to address this question: Were completers prepared to work in diverse contexts, have they done so successfully, and are they growing as professionals? Longitudinal evidence of performance evaluations or student achievement of completers are encouraged, but AAQEP recognizes the challenge of gathering these data and measurement challenges with its use. Suggested data sources include ones similar to those described for Standard 1 but for completers as they begin and continue through their professional career.

its evidence requirements. TPPs must provide multiple measures capturing candidate and completer performance to address AAQEP's first two standards. AAQEP's process requires measures from multiple perspectives (e.g., completers, program faculty, K-12 partners, employers), but the accreditor emphasizes the importance of performance measures that capture candidates' and completers' effectiveness in field placements or the classroom. Unlike CAEP, AAQEP does not outline specific outcome measures but offers a set of suggestions about potential types of measures that could be used to fulfill its requirements.

States' program approval processes also make use of similar data sources. One snapshot of states' reported use of data for TPP assessment, the results of the GAO survey conducted in 2015, is summarized in Table 2. Aligning with federal Title II reporting requirements, pass rates for knowledge-based licensure exams were by far the most commonly employed metric in TPP assessment, with 49 states reporting that they use these metrics to assess at least some traditional programs and 38 states reporting that they use these metrics to assess at least some alternative programs. Just over half of the states also reported using candidate performance assessments, graduation and/or completion rates, or completer surveys as part of their approval process for traditional programs. At least one-quarter of the states reported using satisfaction surveys of school

TABLE 2 Data Used for Teacher Preparation Program Approval by States

	Traditional Programs		Alternative Programs	
Data Type	Used to Evaluate All	Used to Evaluate Some	Used to Evaluate All	Used to Evaluate Some
Licensure assessment pass rates	48	1	35	3
Candidate performance assessments from preservice clinical practice (e.g., edTPA)	29	4	21	4
TPP graduation and/or completion rates	29	3	27	3
Surveys of some or all recent completers' satisfaction with the preparation they received from the TPP	27	6	20	5
Surveys of K-12 schools' satisfaction with the performance of recent TPP completers (completed by principals, district personnel, or others)	23	8	18	5
TPP completer placement rates	19	5	17	4
Teacher evaluation results for recent completers teaching in public schools within the state	16	1	13	3
K-12 student assessment results to measure teacher effectiveness for recent completers teaching in public school within the state	14	1	12	2
Amount or proportion of TPP completers who stay in the teaching field	13	3	12	2

SOURCE: Adapted from Government Accountability Office, 2015.

personnel, graduate and/or completer placement rates, teacher evaluation results, K-12 student assessment results, or retention rates to assess their traditional TPPs.

While many different factors influence how and whether certain measures are useful, we briefly summarize two key considerations: (1) validity and other measurement considerations and (2) equity considerations. For the first, we focus on research examining whether each measure captures how well prepared TPP completers are to begin practicing (i.e., construct validity) and predicts program completers' later performance as teachers (i.e., predictive validity). Numerous reports have summarized the measurement properties of data used to assess TPPs (e.g., see Feuer et al., 2013; National Research Council, 2010; Worrell et al., 2014) but the past decade has seen significant advances in the research on evaluating TPP effectiveness. For the second consideration, we examine how these measures may hinder or support equity-centered approaches to teacher education (Cochran-Smith & Reagan, 2021) and the diversification of the teacher workforce. In the subsections below, our analysis highlights how measures may create unfair comparisons depending on the context and capacity of TPPs.

For each of the performance- and outcomes-based measures we review, we summarize the measurement and equity considerations in Table 3. Important points from Table 3 inform our discussion of the relevance of each measure to accreditation and program approval processes. Our primary focus is understanding how these measures can be used to ensure TPPs are aligning their practices and outcomes to professional

TABLE 3 Measurement and Equity Considerations for Performance- and Outcomes-Based Measures of TPP Effectiveness

Measure and Objectives	Validity and Other Measurement Considerations	Equity Considerations
Knowledge-based licensure exams (e.g., Praxis I and II scores) **Objective as designed:** To screen teaching candidates for a minimum level of competency in certain subject-area or pedagogical knowledge	• **Construct validity:** Testing often occur before candidates begin coursework or clinical experiences so scores may primarily measure candidate selection mechanisms rather than skills and knowledge developed during the TPP (Blue et al., 2002; Comb et al., 2021; Goldhaber & Ronfeldt, 2020; Pool et al., 2004). • **Predictive validity:** Scores have small, positive associations with the achievement results of their future students (Clotfelter et al., 2007, 2010; Cowan et al., 2020; Goldhaber, 2007; Shuls, 2018) or no discernable relationship with achievement results (Boyd et al., 2008; Buddin & Zamarro, 2009; Henry et al., 2013). Relationships vary based on the type of test and the subject area, with licensure exam scores being more predictive of future achievement results in math and science than in English language arts (Clotfelter et al., 2010; Cowan et al., 2020; Goldhaber, 2007). These analyses only include teachers in tested subjects.	• **History of racial bias and differential scores:** There is a long history of bias in teacher exams that disadvantage potential teachers of color, especially Black teachers (Carver-Thomas, 2018). Many studies find differences in scores and pass rates by race/ethnicity (e.g., Goldhaber & Hansen, 2010; Petchauer et al., 2018). Using scores for TPP evaluation could unfairly disadvantage TPPs that enroll more candidates of color. • **Reflection of differing opportunities to learn prior to TPP:** Scores likely pick up differences in opportunities to learn prior to enrollment in TPP (e.g., Nettles et al., 2011) and, in their current form, do not reflect growth and change in knowledge or skills during enrollment in TPP. Current system could discourage TPPs from enrolling students (e.g., first generation students) who had fewer opportunities due to systematic disadvantages. • **Predictive validity varies by race and gender:** One study from North Carolina found that licensure scores are only predictive of later student achievement for female and White teachers but not for male and Black teachers (Goldhaber & Hansen, 2010).

continued

TABLE 3 Continued

Measure and Objectives	Validity and Other Measurement Considerations	Equity Considerations
Performance assessment of teacher candidates (e.g., edTPA) **Objective as designed:** To assess teaching candidates for their performance in instructional skills	• **Construct validity:** edTPA assessment is aligned with InTASC standards (Sato, 2014), and states have aligned their own performance assessments with state teaching standards (California Teaching Performance Assessment, 2021; Massachusetts Department of Elementary and Secondary Education, 2019). One study from North Carolina found that aspects of the TPP experience are associated with differences in performance assessment scores, suggesting that the scores do pick up on differences in TPP experiences (Bastian et al., 2021). • **Predictive validity:** Performance scores can be significant predictors of student achievement or value-added although results vary across studies and subjects (Bastian, 2018; Chen et al., 2021; Darling-Hammond et al., 2013; Goldhaber et al., 2017; Newton, 2010). Studies in Massachusetts and North Carolina found that performance assessments taken during a candidate's TPP predicted graduates' observation scores as teachers (Bastian, 2018; Chen et al., 2021). • **Reliability:** Because performance assessments require individual raters to assess teaching performance based on one or limited observations, there are also important considerations about the reliability of these scores, the need for careful training of raters, and ongoing assessment of interrater reliability (e.g., Gitomer et al., 2021; Whittaker et al., 2018).	• **Concerns over racial equity:** Some evidence of differential scores and passing rates by race/ethnicity of completers (e.g., Goldhaber et al., 2017; Petchauer et al., 2018; Williams et al., 2019), but smaller than those reported in prior analyses of Praxis exams (Carver-Thomas, 2018; Pecheone, 2019). • **Differences in program capacity and context:** Programs have differing capacity to provide support to their candidates for performance assessments (De Voto et al., 2021; Peck et al., 2021) and programs with more limited capacity may struggle to provide adequate support. Similarly, one study in North Carolina found that performance assessment scores are associated with characteristics of student teaching placement schools and mentor teachers (Bastian et al., 2021). It may be valuable to account for differences in capacity or context when using these scores to assess TPP performance.

continued

TABLE 3 Continued

Measure and Objectives	Validity and Other Measurement Considerations	Equity Considerations
Completer and employer surveys **Objective as designed:** To gather information about candidate experience and completer and employer satisfaction about preparation (objectives can vary across surveys)	• **Construct validity:** The validity of the surveys depends on the specific survey measures and the representativeness of the respondents. Some states have designed survey questions to ask completers or employers about preparation in specific aspects of teaching practice captured in state standards (e.g., California, Ohio). There is limited information and research on construct validity. • **Predictive validity:** There is limited evidence on predictive validity. Studies in New York City and North Carolina found that certain survey questions predict later student achievement and/or teacher evaluation ratings (Bastian et al., 2021; Boyd et al., 2009). One study in Chicago found that student teachers' own perceptions of preparedness did not predict later outcomes but that their cooperating teachers' perceptions of their preparedness did predict eventual observational ratings (Ronfeldt et al., 2021). • **Response bias and representativeness:** Survey responses may not accurately reflect the overall perceptions of completers or employers if survey respondents differ considerably from the survey's target population. One study in North Carolina found that the predictive relationships between survey measures and teachers' performance outcomes varied based on the institution's survey response rate (Bastian et al., 2021).	• **Useful tool for assessing equity:** TPPs, accreditors, and state agencies can use surveys as tools for assessing whether completer perceptions or reported experiences differ systematically by demographic groups (e.g., Ronfeldt et al., 2013). Surveys can also include questions about how TPPs address educational equity, diversity, and inclusion in their programs.

continued

TABLE 3 Continued

Measure and Objectives	Validity and Other Measurement Considerations	Equity Considerations
Labor market outcomes (i.e., placement or retention rates) **Objective as designed:** To track what proportion of completers are employed as classroom teachers at a given point in time (can also be used as part of a broader effort to evaluate teacher supply and demand in a given region)	• **Construct validity:** These labor market metrics have high face validity, in that they directly measure one of the main goals of TPPs (i.e., preparing candidates to become classroom teachers). In terms of construct validity, it is difficult to disentangle the potential effect of TPP effectiveness, broader labor market conditions, and eventual teaching context when examining employment and retention rates. Many factors outside of TPPs' control influence teacher hiring and retention. One study in Washington State found meaningful differences in retention rates across TPPs but used 20 years of data and a variety of analytic methods to account for differences in teaching context (Goldhaber & Cowan, 2014).	• **Disadvantaging TPPs whose graduates work in higher needs schools:** Organizational conditions of schools and teaching context consistently predict teacher retention (e.g., see Carver-Thomas & Darling-Hammond, 2017; Nguyen et al., 2020). Poorer working conditions are often associated with schools that serve more students of color or more low-income students, and such conditions are often associated with higher rates of teacher turnover (Carver-Thomas & Darling-Hammond, 2017; Johnson et al., 2012). Thus, assessing TPPs using retention rates of graduates may systematically disadvantage TPPs whose graduates are more likely to work in higher needs schools. • **Useful tool for tracking teacher diversity or equitable distribution of teachers:** These metrics can track the demographics and distribution of teachers. For example, Tennessee's report card for TPPs includes the percentage of completers who identify as teachers of color (Tennessee State Board of Education, 2020). North Carolina's dashboard for TPPs allows users to breakdown placement and retention rates by completer race/ethnicity and gender for individual TPPs. Such metrics should be used with caution given that teachers of color tend to work in higher needs schools and may have higher turnover rates (Carver-Thomas, 2018).

continued

TABLE 3 Continued

Measure and Objectives	Validity and Other Measurement Considerations	Equity Considerations
Classroom performance metrics for graduates of TPPs (e.g., student test scores, teacher observation ratings) **Objective as designed:** To measure and compare teachers' contribution to student achievement or teachers' effectiveness in certain instructional and professional practices as part of the teacher evaluation system	• **Construct validity:** There is ongoing debate about how well classroom performance measures capture teaching performance influenced by TPPs. Student test scores are only available for a subset of grades and subjects, making it difficult to use them to evaluate the broader performance of all TPP graduates. Classroom observations, especially those using established rubrics, arguably have higher construct validity because they can be used for all teachers and are meant to score teachers' effectiveness in particular areas of teaching practice (Grissom & Youngs, 2015; Hill & Grossman, 2013). Many states have aligned their classroom observation rubric with state standards for teaching, although some systems have been critiqued for their lack of suitability in certain subject areas such as special education (Jones et al., 2021; Sledge & Pazey, 2013). • **Error, reliability, and stability:** Teacher value-added measures are often estimated with considerable error and vary across years or across different estimation models (Ballou & Springer, 2015; Corcoran, 2010; Sass et al., 2014). Research examining differences across TPPs in terms of their graduates' effects on student achievement have emphasized methodological challenges with this approach and often concluded that meaningful differences across TPPs are hard to detect (Boyd et al., 2009; Gansle et al., 2012; Koedel et al., 2015; Mihaly et al., 2013). For observations, there have been documented challenges with observer bias and reliability although certain practices and systems can partially ameliorate them (Hill & Grossman, 2012; Ho & Kane, 2013; Kane & Staiger, 2012). One study in Tennessee and another in North Carolina found meaningful differences across TPPs in terms of graduates' observations although each study highlighted important methodological choices such as accounting for differences in teaching contexts (Bastian et al., 2018b; Ronfeldt & Campbell, 2016).	• **Importance of accounting for teaching context:** Because professional working conditions are an important factor explaining teacher effectiveness (Kraft & Papay, 2014), adding contextual characteristics or using models that examine differences within schools is important to try to disentangle the effects of TPPs from the schools in which their completers go on to teach (Bastian et al., 2018b; Boyd et al., 2009; Goldhaber et al., 2013). Systems that do not account for these differences could disadvantage TPPs whose graduates are more likely to teach in high needs or under-resourced school districts. • **Concerns about bias in observational ratings:** Research has found that teachers in classrooms with more students of color, lower-performing students, or lower-income students may be more likely to receive a lower observation rating regardless of their teaching effectiveness (Campbell & Ronfeldt, 2018; Garrett & Steinberg, 2015; Grissom & Bartanen, 2020; Jiang & Sporte, 2016). One study in Tennessee and one study in Chicago found that teachers of color and male teachers also receive lower evaluation ratings, and these differences cannot be explained by other measures of teaching effectiveness (Grissom & Bartanen, 2022; Jiang & Sporte, 2016). Using observation ratings for TPP evaluation could unfairly disadvantage TPPs that enroll more candidates of color or whose graduates are more likely to teach in schools with more students of color or low-income students.

accreditation standards and to support continuous program improvement based on such standards. We discuss how measures may be better suited to certain applications and the trade-offs that exist when they are used to meet multiple goals. While we discuss each measure separately, we recognize that multiple measures should be used in combination for quality assurance and continuous improvement.

Knowledge-Based Licensure Exams

Teacher candidate assessment is a long-standing practice. While there was a National Teacher Exam as early as 1940, knowledge-based testing in its modern form dates from the late 1970s. The widely used Praxis series was developed in the 1990s, and that same decade saw an increasing focus on candidate licensure test scores in HEA Title II reporting. States require teacher candidates to take knowledge exams in various content areas as part of TPP admission and teacher licensure processes, typically setting state-specific cutoff scores for passing. While a few states administer their own exams (e.g., California and New York), most states use teacher exams created by external testing companies. For example, at least 32 states require teacher candidates to take Praxis tests administered by ETS as part of their licensure requirements (Fenwick, 2021).

Current assessments have generally fallen into one of several categories: assessments of general knowledge or basic skills (e.g., California Basic Education Skills Test), assessments of content-specific or subject-matter knowledge (e.g., ETS's Praxis Subject Assessments), and assessments of pedagogical knowledge (e.g., Pearson's National Evaluation Series Assessment of Professional Knowledge tests). As of 2021, 15 states require basic skills tests for admission—down from 25 half a decade ago—while 40 require subject-matter tests for secondary teachers and 25 for elementary teachers (Putman & Walsh, 2021). Many states have different testing requirements—in terms of which tests must be taken or when they must be taken—for candidates who are enrolled in traditional and alternative programs.

The primary objective of these knowledge-based assessments is to screen teaching candidates for minimum levels of competency in certain subjects as determined by state licensure requirements. It is less clear what aggregate data from these assessments (e.g., average scores or passing rate on a specific assessment for teaching candidates) signal about TPP effectiveness. Many teacher candidates take initial basic skills tests (e.g., Praxis I) before they begin their teacher preparation or take subject-matter tests before beginning their post-baccalaureate teacher training. Research on whether these exam scores predict future teaching performance is mixed, as described in Table 3. Also, unlike K-12 student achievement, there are limited options for developing growth measures for teacher candidates' knowledge and skills that could be more directly tied to their experiences in preparation programs. In addition, studies examining Praxis scores have found that candidate experiences and achievement measured prior to program entry are highly predictive of their exam scores (e.g., Blue et al., 2002; Pool et al., 2004). These exams may thus serve as a better measure of TPP selection than TPP effectiveness in preparing candidates (Comb et al., 2021; Goldhaber & Ronfeldt, 2020).

Equity considerations also arise here. Critics of teacher testing argue that licensure exams serve an unnecessary gatekeeping function that keeps potential teacher candi-

dates, especially potential teachers of color, out of the profession (Bennett et al., 2006). There is a long history of cultural bias in teacher exams that disadvantaged potential teachers of color, especially Black teachers (Carver-Thomas, 2018; Goldhaber & Hansen, 2010; Petchauer et al., 2018). Given this evidence of racial bias, using passage rates on TPP exams to make accreditation decisions could potentially disadvantage TPPs enrolling more candidates of color or candidates with fewer opportunities to learn prior to their TPP experience.

Overall, knowledge-based licensure exams may have limited use for accreditation decisions or program improvement efforts. Given the difficulty of using these exams to measure how candidates develop subject area or pedagogical knowledge as part of their TPP experience, their utility to inform programmatic decision-making or continuous improvements efforts seems limited to considerations about candidate selection. TPPs could use candidate-level scores to support candidates struggling to pass relevant exams, but this could lead to "teaching to the test" and curricular narrowing that may not benefit teacher candidates. If states continue to require prospective teachers to pass these exams, there is some logic to incorporating these metrics into state program approval decisions because a primary goal of teacher preparation is that completers of TPPs should become certified teachers. Still, between concerns about racial/ethnic inequities and validity summarized in Table 3, it is hard to argue that using passing rates on licensure exams offer a fair evaluation of TPPs.

Performance Assessment of Teacher Candidates

For the past three decades, many TPPs and states have developed and adopted performance-based assessments of teacher candidates. These new assessments have been hailed as an important improvement over knowledge-based licensure exams in that they focus on assessing important pedagogical skills of teachers and many were developed by educators based on standards of professional practice (Darling-Hammond et al., 2013). The Performance Assessment for California Teachers (PACT), the predecessor of edTPA, was developed at the Stanford Center for Assessment, Learning, and Equity (SCALE) with engagement from TPP faculty across California. Now administered nationally by Pearson, edTPA evaluates TPP candidates in three areas: planning, instruction, and assessment (Stanford Center for Assessment, Learning, and Equity, 2019). When edTPA was launched nationwide in 2013, it attracted support from key players such as NEA and AACTE (Pinsky & Fry, 2013; Robinson, 2013; Stanford Center for Assessment, Learning, and Equity, 2013, 2014; U.S. Department of Education, 2013). ETS also developed a standards- and accreditation-aligned performance assessment, PPAT, launched in 2015, which includes tasks focused on knowledge of students and the learning environment, assessment, and instructional design, implementation, and analysis (Educational Testing Service, 2017).

Neither national accreditor requires the use of performance assessments in the accreditation process. CAEP lists edTPA and other performance assessments as one type of evidence of content and pedagogical knowledge and indicates that pre- and post-instruction student data captured as part of performance assessments can be used as evidence of candidates' impact on student learning (Council for the Accreditation of

Educator Preparation, 2020a). AAQEP considers performance assessments like edTPA, PPAT, and the California Teaching Performance Assessment (CalTPA) to be direct measures of candidate performance and highlights them as one of the strongest measures available in the field to measure both candidate and program effectiveness (Association for Advancing Quality in Educator Preparation, 2021b). As of 2019, edTPA was in use at more than 900 TPPs across 32 states, 21 of which had policies in place linking edTPA to licensure or TPP approval (Pecheone, 2021). PPAT is currently in use in seven states (Putman & Walsh, 2021).

Both national and state performance assessments have been developed to align with teaching standards (see Table 3). The evaluation of teacher candidates according to professional standards is thus an important benefit of the performance assessment approach, as it can create a common language and set of expectations across the profession to guide teacher preparation (Darling-Hammond, 2010; Peck & McDonald, 2013; Peck et al., 2014). Such alignment could make performance assessment scores a useful tool for TPPs to provide targeted support to candidates (Bastian et al., 2018a) and to support program evaluation and learning (De Voto et al., 2021; Peck et al., 2021), particularly when those scores are broken down by specific teaching domains. However, because of their complexity, external resources and internal capacity building may be necessary for TPPs to engage meaningfully with performance assessments (De Voto et al., 2021; Many & Bhatnagar, 2017; Peck et al., 2021).

Although the research on construct and predictive validity of performance assessments is still emerging (see Table 3), findings suggest that candidate scores may differ based on aspects of the TPP experience itself (e.g., Bastian et al., 2021) and may also serve as a useful signal of candidates' later performance as teachers (e.g., Bastian, 2018; Chen et al., 2021). Passing rates on performance assessments may thus provide one helpful metric of how well TPPs are preparing their teacher candidates. Clinical placements—including the characteristics of placement schools and mentor teachers—may also influence performance assessment scores (Bastian et al., 2021), and states and accreditors should consider whether and how to account for differences in clinical experience placement across TPPs. As with knowledge-based licensure exams, performance assessments have become part of the teacher licensure process in many states, meaning that data collection mechanisms could already be in place and available for further integration into accreditation and program approval processes.

There are still ongoing concerns about reliability and racial equity with performance assessments (see Table 3), and these concerns must be weighed when deciding how these assessment results may be incorporated into an accreditation process. The racial/ethnic differences found in edTPA performance assessments are smaller than reported differences from prior analyses of Praxis exams (Carver-Thomas, 2018; Pecheone, 2019), yet emerging research suggests that the integration of performance assessments into state licensure systems may have decreased the number of candidates, and especially candidates of color, graduating from TPPs (Chung & Zou, 2021). Some states have begun to back off from their commitment to edTPA—including Washington State, Wisconsin, and Georgia—claiming the assessment negatively affects teacher supply and diversity (e.g., Jacobson, 2020; Washington State Professional Educators Standards Board, 2021; Will, 2020). While performance assessments may outperform traditional

knowledge-based exams in terms of providing a useful measure of TPP performance, equity concerns should be monitored as more research emerges about such assessments.

Completer and Employer Surveys

Programs have long used internal surveys to gauge teacher candidates' perspectives on their preparation experience (e.g., course evaluations or exit surveys). Looking to external evaluation, both national accreditors list surveys as one method for collecting outcomes data, while numerous states use statewide surveys to capture the perspectives of TPP completers or their employers (e.g., principals or district leaders) for use in program approval decisions. CAEP specifically lists surveys as one mechanism for gauging completer and employer satisfaction with TPPs, two of the required outcome measures (Council for the Accreditation of Educator Preparation, 2020). AAQEP also mentions surveys as one method to gather multiple perspectives on TPPs (Association for Advancing Quality in Educator Preparation, 2021b). In the 2015 GAO survey (see Table 2), 33 states reported using completer surveys as part of the assessment process for traditional TPPs and 25 states reported using these surveys when assessing alternative TPPs.

For example, Texas surveys all new teachers who completed a TPP within the state to gather completer feedback on the effectiveness of their program and surveys principals about perceptions of the preparation for all first-year teachers in their school (Texas Education Agency, 2022). The completer survey asks new teachers and principals to assess preparation in six areas: (1) planning, (2) instruction, (3) learning environment, (4) professional practices and responsibilities, (5) supporting students with disabilities, and (6) supporting English language learners. They classify teachers as "meeting standard" if their average survey scores are above a 2.0 on a 4-point scale, and then calculate the proportion of respondents per institution who met this standard (Texas Education Agency, 2022). Similarly, Ohio surveys all preservice teacher candidates enrolled in IHEs, completers who are currently working in Ohio schools, and employers as part of their educator preparation metrics report. Ohio's survey includes questions about candidates' level of preparation in domains of teaching aligned to state standards, then reports institutional and state averages for each question (Ohio Department of Higher Education, 2022). As with performance assessments, alignment with professional standards increases the value of surveys for accreditation and program approval.

Unlike the other metrics reviewed in this paper, surveys are a measurement tool rather than a specific measure. Because they can be used to assess many aspects of TPPs, their validity depends on the specific measurement objectives of the survey, the questions themselves, and the representativeness of the survey sample. There is limited external research on the validity of TPP surveys, but multiple studies have found evidence that certain survey questions can be predictive of future teacher outcomes (Bastian et al., 2021; Boyd et al., 2009). From an equity perspective, surveys can be a useful tool for assessing the extent to which TPPs offer an equitable learning experience for all completers. Survey responses can be disaggregated by candidate demographic groups (e.g., see Ronfeldt et al., 2013), while questions can be tailored to assess exposure to culturally responsive teaching practices or training for completers to support English language learners, students of diverse backgrounds, or students with disabilities.

Surveys are sometimes administered directly by TPPs. This creates a significant burden on TPPs that must develop the instruments, determine administration methods, administer surveys to candidates or completers, analyze the data, and compare survey respondents to all completers to assess if the results are representative. While such an approach does allow TPPs to tailor the survey instrument to their specific needs, it can often be challenging for TPPs to craft valid measures and achieve representative samples in their responses. Both national accreditors recognize these challenges. While CAEP requires TPPs to provide documentation that they have collected sufficiently representative responses (Council for the Accreditation of Educator Preparation, 2020, p. 84), AAQEP highlights the benefit of using statewide surveys rather than those fielded by individual programs (Association for Advancing Quality in Educator Preparation, 2021b).

Different methods of disseminating surveys may increase both response rates and representativeness. In California, the Commission on Teacher Credentialing has embedded its survey about TPP experiences into the online application process for teacher credentialing. As a result, its survey response rates are typically above 90%, much higher than surveys conducted in other states. In Tennessee, all teachers are asked to participate in the annual Tennessee Educator Survey (annual response rates vary between 50-60%), which asks early-career teachers to answer a subset of questions about their TPP experience. Notably, both states developed systems to ensure that TPPs receive the survey data for their institutions and also have systems for reporting overall results. Importantly, these approaches require state-level mechanisms for data collection and dissemination.

Surveys offer serious potential to inform program improvement. The specificity and timeliness of survey data create opportunities for TPPs to get almost real-time perspectives from current students, recent completers, or employers. As previously discussed, there are notable challenges when it comes to survey development and response rates when surveys are managed by individual TPPs. Therefore, in terms of their use for accreditation and program approval, surveys represent an opportunity for national accreditors to work with states to develop survey measures aligned with specific professional standards for teachers and teacher preparation. Such a collaboration, which would put states in the lead on implementation, could also marshal resources not only toward developing and validating survey measures but also toward analyzing results, removing these burdens from individual TPPs.

Hiring, Retention, and Other Labor Market Outcomes

When looking at measures collected after candidates complete their TPP, one category of evidence includes labor market outcomes, such as employment rates of graduates into teaching positions or retention rates after a certain number of years. As illustrated in Table 1, CAEP specifically requires that programs submit employment rates of their completers while AAQEP discusses the need for evidence that indicates career progression for completers after they have begun teaching. In the 2015 GAO survey, 24 states reported that they used completer employment rates in their assessment of at least some traditional TPPs and 21 states reported using them for assessments of alternative TPPs. In terms of retention rates, 16 states reported that they used

the "amount or proportion of TPP completers who stay in the teaching field" in their assessment of at least some traditional TPPs and 14 states reported using these data for assessment of alternative TPPs (Government Accountability Office, 2015).

For example, Colorado's Education Preparation Program Dashboard displays the percentage of completers of initial teaching licensure programs who obtain teaching positions in Colorado public schools, if those completers are teaching within their trained field, and retention and attrition rates (i.e., the percent of completers who remain teaching in Colorado public schools or who have left their teaching positions). Similarly, the North Carolina Department of Public Instruction releases the percent of completers from every TPP in the state who are teaching within North Carolina public schools in a given time frame after graduation. While the specific metrics vary across accreditors and states, we discuss these metrics broadly under the umbrella of labor market outcomes.

These metrics—especially employment rates of completers in teaching positions—have high face validity, in that they directly measure one of the main goals of TPPs: preparing candidates to become classroom teachers (see Table 3). However, TPPs do not control broader forces that affect teacher hiring and retention such as teacher salaries, working conditions at placement schools, and the state of the economy. State agencies and accreditors should consider how to account for these broader forces if they integrate these metrics into their systems. Also, unlike licensure exams with predetermined passing scores, it is harder to decide what constitutes acceptable employment and retention rates to inform accreditation or program approval decisions.

Conditions of schools and teaching contexts are consistent predictors of teacher retention (e.g., see Nguyen et al., 2020). Poorer working conditions (e.g., less supportive administration, less time for professional learning, lower salaries) are often associated with schools that serve more students of color or more low-income students, and such conditions are often associated with higher rates of teacher turnover (Carver-Thomas & Darling-Hammond, 2017; Johnson et al., 2012). Thus, assessing TPPs by the percentage of their graduates who are retained in their school or in the profession may systematically disadvantage programs whose graduates are more likely to work in higher needs schools. Such an approach could undermine the goal of having highly qualified and well-prepared teachers in these schools. It could also function to penalize TPPs that are deliberately focused on preparing teachers for these schools, disincentivizing such a focus and further undermining equity goals.

While individual TPPs do sometimes track their completers over time (Wineburg, 2006), creating widespread metrics related to labor market outcomes requires coordination with state educational data systems. Some states already report on these metrics as part of TPP evaluation while others track placement and teacher retention to inform statewide policymaking around teacher supply and demand. Such calculations are only possible if states have existing statewide longitudinal staffing data in which individual teachers can be tied to their preparation program. Most of these calculations include only teachers in a given state's public school system as most states do not systematically collect staffing data for private schools or share these data with other states. This can create challenges for TPPs that primarily prepare candidates for teaching in private or religious schools or for institutions located near state borders (E. Leicht, Tennessee State Board of Education, personal communication, October 6, 2021).

Although collecting placement and retention data can be quite informative for states, it is less clear how TPPs can use these data for program improvement. There are many external factors that influence labor market outcomes, including the preferences of TPP completers, the broader economic conditions, and the current needs and hiring preferences of schools and districts. While it may benefit TPPs to understand those conditions in order to better counsel their candidates, it may be more difficult to use these data to assess the effectiveness of their programs or drive internal improvement.

Classroom Performance Metrics for Graduates of TPPs

One final category of outcome measures includes classroom performance data for completers who are currently employed as teachers, such as student achievement results (often in the form of teacher value-added scores) and classroom observations that are incorporated into states' teacher evaluation systems (Grissom & Youngs, 2015; Steinberg & Donaldson, 2015). Both CAEP and AAQEP require measures capturing the performance of completers once they are employed as teachers (Association for Advancing Quality in Educator Preparation, 2021b; Council for the Accreditation of Educator Preparation, 2020). As illustrated in Table 1, CAEP specifically requires both measures capturing student achievement and teacher performance (e.g., classroom observations or student surveys) while AAQEP suggests using classroom observations. In terms of states' assessment systems, the 2015 GAO report found that 17 states use classroom observation ratings as part of their assessment of traditional TPPs while 15 states use student achievement results. In recent years, some states have moved away from using student achievement metrics, with only 11 states reporting that they used student growth data to assess TPPs as of 2021 compared to 22 states in 2017 (National Council on Teacher Quality, 2021).

There is considerable debate about how to use these metrics, especially value-added scores, to measure *teacher* effectiveness (e.g., see Ballou & Springer, 2015; Cohen & Goldhaber, 2016; Corcoran, 2010; Darling-Hammond, 2015; Haertel, 2013; Hill & Grossman, 2013; Steinberg & Kraft, 2017). For value-added measures, numerous researchers have cautioned that value-added measures are estimated with a large degree of error and often vary substantially across years or across different estimation models (e.g., see Ballou & Springer, 2015; Corcoran, 2010; Goldhaber, 2015; Sass et al., 2014). Researchers comparing TPPs using student achievement or value-added scores have identified similar concerns. Many studies highlight the difficulty in identifying meaningful differences across TPPs (Boyd et al., 2009; Goldhaber et al., 2013; Koedel et al., 2015) and show how specific methodological choices make a considerable difference in how TPP effectiveness is captured and reported (Mihaly et al., 2013). These challenges are compounded by the fact that student growth or value-added measures can typically be calculated only for teachers in public schools teaching certain tested subjects. Thus, many completers will never have achievement or value-added measures as teachers, and it can be particularly difficult to use such measures to assess TPPs with small enrollments. Much of the research comparing across TPPs sets certain inclusion criteria for the minimum number of completers who have performance data, and some states have adopted rolling averages that look across multiple years of completers when calculating performance-based measures.

Classroom observations, especially those using established and well-tested rubrics, arguably have higher construct validity because they are meant to score teachers' effectiveness in particular areas of teaching practice (Grissom & Youngs, 2015; Hill & Grossman, 2013). Although implementation varies across states (Steinberg & Donaldson, 2015), many states have aligned their classroom observation rubric with state standards for teaching. Such alignment potentially strengthens the argument for using classroom observation ratings in the accreditation and program approval process. As with performance assessments, observation rubrics create common expectations and language around teaching that could be used by TPPs to organize instruction and feedback for teaching candidates.

However, there are still important concerns to consider. Universal, cross-subject observation systems have been critiqued for their lack of suitability in certain subject areas (Jones et al., 2021; Sledge & Pazey, 2013). Recent studies have found that teachers of color, male teachers, and teachers working in classrooms serving more low-income students or students of color receive lower observation ratings, and these differences cannot be explained by differences in teacher qualifications or other measures of effectiveness (e.g., Campbell & Ronfeldt, 2018; Garrett & Steinberg, 2015; Grissom & Bartanen, 2020; Jiang & Sporte, 2016). Studies that use classroom observation ratings of TPP completers to measure TPP effectiveness and compare across programs have emphasized the importance of accounting for teaching contexts (e.g., school level, school size, suspension rates, per-pupil expenditures, student demographics) in their analyses (Bastian et al., 2018b; Ronfeldt & Campbell, 2016). With either value-added measures or observation ratings, using measures that do not account for differences in teaching context could potentially disadvantage TPPs whose graduates are more likely to teach in high needs or under-resourced school districts, or who deliberately prepare teachers to work in those schools and districts.

Both national accreditors stress the importance of capturing the performance of TPP completers once they enter the classroom, and value-added measures and observation ratings offer the most direct measures commonly available to do that. Because these measures are built into states' teacher evaluation systems, using the same metrics in TPP approval creates a more coherent accountability framework. As with licensure exams, accreditors or state agencies tasked with program approval could examine how many graduates are considered at or above the professional standard set by a given state's teacher evaluation system. Still, to be able to effectively use these measures to evaluate TPPs, data systems must be able to link student achievement results or teacher observation ratings to individual teachers, link those teachers to their preparation programs, and then be able to appropriately aggregate that data for TPPs. There are many methodological considerations in doing so, with different choices potentially resulting in substantively different ratings for TPPs. Finally, these data must come from graduates after they have been employed in the classroom for at least 1 year, and research suggests that using multiple years of data creates more stable estimates of effectiveness. So, there is a significant time lag in terms of using these data for accreditation. Given these challenges, it may be hard to use teacher performance data for internal program improvement.

RECOMMENDATIONS FOR THE FUTURE OF ACCREDITATION

The historical evolution of TPP accreditation has been characterized by a series of recurring, and conflicting, themes. The drive for national standards in teacher preparation has been consistently matched by doubts about centralization and standardization. Policymakers have repeatedly made simultaneous moves to regulate and deregulate educator preparation, though the former is more often focused on traditional, preservice TPPs and the latter on alternative, inservice TPPs. Accreditors have been called on to increase their rigor and, often contemporaneously, to decrease the cost, complexity, and inflexibility of their institutional reviews. Finally, to an increasing degree in recent decades, efforts to identify and implement measures of TPP quality have been complicated by debates about the validity, usefulness, and unintended consequences of educational assessments at multiple levels.

As the previous section demonstrates in detail, there are few simple answers to questions about measurement of TPP performance for either quality assurance or program improvement. What is clear is that long-standing debates about data use have often failed to capture the nuances of research findings on these issues related to validity, equity, and data accessibility. Even as accreditation has been reviewed, updated, improved, and challenged, broad concerns about "rigor" or "accountability" seem to dominate the discussions of data and evidence, to the exclusion of specific attention to the relationship between the purposes of accreditation and the appropriate use of available data and evidence.

All of this raises questions about how much progress has been made in accreditation in the 70 years since the establishment of NCATE. Recall that NCATE was expected to align TPPs to national standards for the preparation of teachers and, at the same time, to replace inconsistent state-level program approval systems. In pursuing these goals, it was expected that NCATE would boost teacher quality and improve the field of education generally. This expectation did not account for the lack of any direct connection between TPP accreditation and other potential quality-boosting efforts like higher compensation for teachers, nationwide teaching standards, or national teacher assessments.

Considering the state of the field, it seems hard to argue that the original goals set for TPP accreditation have been achieved. As a voluntary process with a history of inconsistent support from educators and regulators, national accreditation has seldom been sought or achieved by more than 50% of TPPs. National accreditation has not established universally accepted standards for TPPs, nor has it replaced state program approval processes. The recent emergence of a second national accreditor, following a familiar historical pattern, gives the impression that progress in national accreditation has stalled. The stubbornly insoluble disagreements that characterize conversations about data and evidence give the same impression, as does the overly simplistic dichotomy, in these same conversations, between exclusionary reliance on, or complete rejection of, outcomes- and performance-based measures and data.

If, however, the question is whether the original goals of accreditation still stand up today, the answer is not so straightforward. The original problems of vast variation and low quality in TPPs have certainly not been resolved. Indeed, the proliferation of alternative and emergency routes to licensure, and of new forms of TPPs that provide

them, seems to be leading in the opposite direction, with more ways to become a teacher and a growing number of them requiring little or no preservice training. And yet, while such developments do not seem to represent progress, they do suggest that the original goal of establishing national standards applicable to all TPPs, even if it remains elusive, also remains relevant. In other words, the idea behind this goal—establishing a common level of expected quality for TPPs—is still worth pursuing.

In contrast, the second founding goal of replacing state TPP approval systems seems less relevant. For one thing, these state systems show no sign of withering away. Indeed, given the role that state agencies could play in improving accreditation, TPP accreditors would be wise to support their continued existence. The shift to outcomes-focused accreditation and program approval means states occupy a unique place in TPP quality assurance writ large: they have the capacity to gather and share the kinds of data and evidence that TPPs and accreditors require for their continuous improvement efforts. This is particularly true of survey data, employment-related outcomes, and teacher performance measures for TPP graduates. Therefore, it is important to recognize that only states—not accreditors, TPPs, or district partners—can leverage their regulatory authority and their ability to aggregate such data to enable standards- and performance-based accreditation to function.

Our recommendations, then, flow from two conclusions drawn from our analysis. First, there is still a need for accreditation to fulfill a nationwide quality assurance role for teacher preparation. Second, accreditation will not be replacing state program approval systems, but should rather aim to work with them to leverage state capacity and infrastructure for gathering and analyzing data and evidence. To serve both goals, decisions about how to make the best use of data and evidence in quality assurance and continuous improvement on the one hand, and for regulation and accountability on the other, should be made with the important distinctions between these two systems at top of mind. These considerations inform our recommendations below.

Limitations of Accreditation and Program Approval as Levers for Change

Even though TPP accreditation was established with lofty goals, it has never operated in a vacuum. Accreditation is only one piece of our education system, which itself is embedded in broader social and economic contexts. Therefore, it is highly unlikely that changes to TPP accreditation or program approval processes can significantly strengthen the profession by themselves. A broader effort, with nationwide investments in people and systems, would be needed to open up access to teaching for a more diverse pool of candidates while boosting the attractiveness of the profession by raising salaries and improving working conditions. Such investments in the teacher workforce would likely have to be accompanied by a national, cross-state commitment to establishing consistent, effective educator licensure systems. This effort would eliminate the inconsistencies, loopholes, and alternatively regulated pathways to teaching currently allowing professional entry without professional preparation.

Another area for increased investment would be in TPPs themselves, and particularly state-supported IHEs. One aspect of this investment could target TPPs' capacity to use evidence for continuous improvement (Peck et al., 2021). Giving TPP faculty time and support for working together toward improvement goals may also require the

realignment of institutional incentives within TPPs. This would entail not only incentivizing faculty collaboration but also increasing the percentage of full-time faculty, which has been decreasing for decades. It would also involve reversing declines in state funding for IHEs that have been eroding TPP capacity. Strengthening accreditation as a lever for improving teacher preparation may necessitate a both/and strategy of investment in people and systems in higher education, and particularly in TPPs.

The Urgency of Action in Improving Accreditation

A final issue to review is the urgency of the moment. Recent policy developments in educator preparation have not only undermined accreditation but also present challenges to the concept. While it is true that accreditors of TPPs, and TPPs themselves, have come under strong critique many times before, there is some evidence that recent developments represent a break from past challenges. Instead of meeting teacher shortages by creating new routes and pathways to the classroom while preserving a role for TPPs (a strategy commonly used in the past), policymakers are setting up parallel teacher preparation systems disconnected from TPPs and their accreditors.

In one prominent example, the Every Student Succeeds Act (ESSA) of 2015 included a provision authorizing states to spend Title II dollars to establish teacher preparation "academies." These academies do not need to be affiliated with IHEs or seek accreditation; rather, they recommend candidates for licensure based on K-12 student achievement. This creation of funding and authorization channels entirely outside of existing systems has already raised alarms (Zeichner, 2016), and there have been similar developments at the state level. In 2021, the Oregon legislature passed House Bill 2166, allowing, and providing $3.5 million to fund, the establishment of alternative licensure programs that can operate without national accreditation for 4 years after earning state program approval. Mississippi is running a pilot of a performance-based licensure system that links educator licensure to classroom effectiveness ratings, and teachers pursuing this route do not need to complete a TPP (Garcia & Muñiz, 2020; Jackson Public Schools, 2021). North Carolina is considering, but has not yet implemented, a system of performance-based career steps for teachers, starting with apprentice roles that do not require TPP completion (Southern Regional Education Board, 2021).

These developments should raise red flags. If there is a broad movement toward these new teacher preparation systems, the training and licensure of teachers would increasingly not require accredited TPPs nor, by extension, TPP accreditors. This could reduce, or even eliminate, the role that accreditation plays in promoting teaching as a profession. The current moment, in which the long-standing challenges and weaknesses of an industrial era education system are amplified by the stresses of the COVID-19 pandemic, could spur such a movement. Widespread concerns about dropping public school enrollment and teacher shortages could lead to reform efforts targeting TPPs and their accreditors. If the possibility of improving TPP accreditation does not, by itself, provide sufficient impetus for consideration of the recommendations provided here, the state of education today provides additional motivation.

Recommendations to Refocus Accreditation and Data Use on Building the Profession and Improving Teacher Preparation

Over time, accreditation has followed larger trends in K-12 education toward using accountability as the main policy lever to promote improvement. CAEP's approach, with its articulations of greater rigor and its emphasis on outcomes data, especially test score data, was solidly and deliberately aligned with the accountability frameworks such as NCLB and RTT. As described in our historical overview, the pushback to this approach was strong, riling up TPP faculty, costing the president of CAEP his job, and inspiring, in part, the formation of AAQEP. One way to characterize the actions of CAEP, and to contextualize the negative response they produced, is to say that CAEP made an error in attempting to align accreditation entirely with regulation.

Accreditation and regulation are related, and overlap to some extent, but they are not interchangeable. Regulation, at the state or federal level, does fulfill certain quality assurance functions of an accreditor. Still, the accreditor has an additional set of functions, central to its identity: the maintenance and advancement of the profession itself. When accreditors set and maintain standards, they are not simply constructing a floor for TPP performance and practice but also a foundation on which improvements can be built. This is a key reason why accreditation is a peer-review process; peer review is a key feature and indicator of the self-regulating nature of professions.

As described above, we are proceeding from two conclusions; namely, that there is still a need for accreditation to fulfill a nationwide quality assurance role for teacher preparation and that accreditation and program approval must work together, rather than in opposition, to make this happen. To these, we can add a third conclusion: that the concerns of the accreditor are not, and cannot be, entirely aligned with the concerns of the regulator. The former will always be broader, encompassing a wider set of goals and responsibilities. Attempting to align accreditation and regulation, as CAEP has done, overly narrows the scope of accreditation and reduces its potential contributions to the profession.

Drawing these conclusions together, we propose that a deliberate effort to refocus accreditation on its professional roots and functions could serve as a needed course correction. Such an effort would bring much needed quality assurance and improvement functions back to the top of accreditation priorities, provide separate lanes for accreditation and program approval to enable collaboration rather than competition, and give accreditors an opportunity to make new and more extensive contributions to supporting and improving TPPs and the teaching profession. To these ends, we make the following recommendations for TPP accreditors:

Recommendation #1: Re-orient data use toward professional quality assurance and continuous improvement

a. Use measures aligned with professional standards for teaching and teacher preparation
b. Consider data use beyond the evaluation of individual TPPs

Recommendation #2: Partner with states to build system capacity to support accreditation and program approval

a. Inform the expansion of state capacity uniquely suited to meeting accreditation data needs
b. Work with states to align reporting structures to the goals that the data are intended to serve

Recommendation #3: Provide leadership in key efforts to update standards and structures for the teaching profession

a. Advocate for, and participate in, the revision and update of teacher standards
b. Follow the revision of teacher standards with an aligned revision of accreditation standards, and account for updates to the uses of data and evidence
c. Maintain common standards for all routes and pathways to teaching
d. Set in motion the reevaluation of professional certification for teachers to ensure it keeps up with teacher preparation and school staffing structures

Recommendations in Detail

In the following sections, we will describe each of these recommendations mentioned above in detail.

Recommendation #1: Re-orient data use toward professional quality assurance and continuous improvement

Assuming that performance- and outcomes-based measures will continue to be key features of TPP accreditation and program approval, the broader goal of refocusing accreditation on its profession-related functions means accreditors should consider how these measures can support continuous improvement and quality assurance. This goal can drive the processes and structures of accreditation, supplying the "why" that informs the "what" and "how" for uses of data and evidence. The general movement toward outcomes-based accountability has sometimes eschewed the importance of measuring inputs (e.g., the type and content of coursework, length and intensity of preservice clinical experiences), setting these aside in favor of an exclusive focus on outcomes. Approaches such as the impact and outcomes-based framework created by CAEP did serve as a corrective for the overly input-focused nature of previous iterations of TPP accountability and program approval. However, it is difficult from a professional standpoint to conceive of an accreditation process that is entirely agnostic about the subject matter, program features, and preparation practices leading to professional certification.

Professional standards are not simply a matter of outcomes; they are the scope of practice as defined by professionals. Therefore, adopting an exclusively "outcomes-based" approach is not sufficient for professional accreditation, though neither should outcome measures simply be dismissed. Instead, data and evidence on the outcomes side acquire their relevance through connections to professional standards and professionally defined performance. In other words, outcomes-based accreditation is standards-based and performance-based accreditation (see, e.g., Darling-Hammond,

2020). Evaluating programs against standards, and through performance expectations built on those standards, provides TPPs with feedback they can use to move closer to meeting their responsibilities to candidates and to these candidates' future students. It is this kind of feedback that is necessary for continuous improvement. Furthermore, it is this kind of feedback that, along with the channels by which TPPs themselves provide information and ideas, allows for ongoing reflection on standards and expectations necessary for the growth of the profession.

a. Use measures aligned with professional standards for teaching and teacher preparation

Using measures more directly aligned with professional standards for teaching, or measures that could more easily be aligned, offers multiple benefits for strengthening accreditation processes and supporting continuous TPP improvement. Organizing measures around professional standards—both accreditation standards applied to TPPs and teacher standards—holds the promise of focusing the attention of teacher candidates, TPP faculty and program administrators, state agencies, and accreditors on essential and shared expectations for teacher preparation. Emphasizing alignment with TPP and teacher standards could also serve as a corrective for systems that have been overly focused on high-stakes accountability metrics that are tenuously connected to preparation experiences.

Of the measures described above, performance assessments and classroom observations of TPP completers have the most potential. Both are structured around teaching competencies or expectations, which can be aligned with specific professional standards. For individuals, these assessments serve as important markers to identify if those completing TPPs or starting in their classroom career have reached a basic level of competency in various domains of teaching. Performance assessments and observations, aligned to standards, provide a common language and aligned expectations around which TPPs can organize coursework and clinical experiences and assess the effectiveness of those program experiences. These measures can help programs create systems to support individual candidates in their preparation experiences and can also drive programmatic improvement in curriculum and clinical experiences. Still, the goal of quality assurance will demand that TPP accreditors and state program approval systems exercise care in setting expectations regarding these measures, ensuring that these expectations account for differences in teaching contexts and candidate demographics that could affect performance assessment scores or classroom observation ratings.

Similarly, completer or employer surveys can be aligned with professional standards and goals. For example, survey questions can ask about candidates' opportunity to learn in certain important domains of pedagogical or content knowledge and/or ask completers and employers to assess their preparation in specific domains of teaching practice. Indeed, some states already take this approach in their existing surveys (e.g., Ohio and California). Survey questions can also ask about specific programmatic experiences and their alignment with standards, even getting down to the details such as the duration and depth of clinical experiences and the level of support during clinical practice. Accreditors, state agencies, and other professional organizations could invest in developing sets of validated survey questions that would be available for use across

states, opening up the possibility of not just aligning surveys to standards in any given state but also to standards across states, building toward the national quality assurance system envisioned at the founding of TPP accreditation.

b. Consider data use beyond the evaluation of individual TPPs

National accreditors should also consider how to use performance and outcomes data beyond the evaluation of individual TPPs, especially measures that are not as well aligned with professional standards. Some of the problematic features of the data sources described above are mitigated when the data are used in the aggregate or applied to provide information about how TPPs are performing in general rather than individually. If it is difficult to make use of student achievement data to make any meaningful distinctions between the quality of a program, there still might be some value in bringing together as much data as can be gathered to consider how particular types of programs, with different features or pathways to the classroom, can be linked to gains in student outcomes or reductions in opportunity and achievement gaps. A similar approach could be taken when looking at labor market outcomes of teacher candidates. Findings such as these could be fed back to the field for use for improvement of individual TPPs and teacher preparation writ large. Accreditation, in a position of setting and enforcing professional standards, would also be in a natural position to take on the task of assessing the health of the field in this way. Although it would mean a shift in the mission and a building up of capacity for accreditors, along with new relationships with states.

Recommendation # 2: Partner with states to build system capacity to support accreditation and program approval

As discussed in the data sections above, there are not only methodological considerations about how outcomes and performance measures should be employed but also broader questions about the capacity for data use in the field of teacher preparation. Because accreditation happens at the program level, much of the burden for collecting and analyzing data has fallen to TPPs. These responsibilities go beyond simply reporting on candidate outcomes (i.e., calculating the percentage of candidates who pass a certain licensure exam or graduate within a given time period) but also include gathering evidence on candidate performance during their coursework and clinical experiences, locating and making use of employment and retention data, and fielding surveys of candidates, completers, and employers while trying to assess the validity and reliability of these measures and partnering with school districts to gain access to such data (Bastian et al., 2016; Peck et al., 2021; Wineburg, 2006). Both national accreditors require that the evidence provided by TPPs meet certain standards of data quality in terms of validity and reliability, relevance, and representativeness of evidence provided, with some recognition of the challenges embedded within these data requirements (Association for Advancing Quality in Educator Preparation, 2021b; Council for the Accreditation of Educator Preparation, 2020).

Efforts at "outcomes-based" reform have been accused, not without cause, of creating unfunded mandates because of their increasing demands on institutional capacity.

Indeed, attempts to move accreditation toward more outcomes-based and data-driven bases have been resisted, in part, because of the strain such moves place on state agency staff and TPP faculty. Capacity issues also create strong incentives to choose data sources, not because of their usefulness for quality assurance and continuous improvement, but based on convenient access or ease of use, undermining the possibilities for well informed and deliberate use of the data. Capacity issues may also create challenges for TPPs beyond efforts to access data, because the time and effort required to analyze such data, and to employ such analyses to inform program improvement, are also considerable. These challenges grow in inverse proportion to the size of TPPs and the resources available to them. These capacity issues are worsened by increasing reliance in higher education on temporary and adjunct faculty and by state budget cuts to higher education systems, particularly regional universities and colleges that prepare many teachers.

a. Inform the expansion of state capacity uniquely suited to meeting accreditation data needs

While we recognize that TPP-level organizational conditions must be in place to support assessment data use for improvement (see Peck et al., 2021), accomplishing widespread changes in funding and support for state-funded IHEs are not realistic goals for accreditors. Instead, accreditors can seek partnerships with states to build state data system capacity toward collecting and analyzing evidence that benefits accreditors and the broader teaching profession. The measures we recommend as focus areas for accreditors—performance assessments, classroom observations, and surveys—are not only frequently connected to state requirements for teacher licensure and evaluation but are also well suited for state-level data collection and analysis. As mentioned above, states are uniquely positioned to bring resources to bear, at scale, to create and maintain data systems that are useful to TPPs and accreditors. States can also link systems, as done in California, with the incorporation of surveys into the state's credentialing process to boost response rates. This example shows how system design can improve data access without being defined by simple convenience and availability.

Moving these responsibilities to the state level does present some issues. States need resources and internal capacity to capture high-quality data, warehouse the data in accessible data systems, and share the data across the numerous state agencies responsible for assessing TPPs (Fenwick, 2021). Therefore, accreditors should work with states investing in data infrastructure to highlight the in-depth attention needed to the methodological challenges and practical constraints of such systems. Collaboration among accreditors, state agencies, and TPPs would have to go beyond simple data-sharing agreements (although those are necessary) to get to the intersection of state data systems with assessment design. There is a considerable disconnect between academic research, often using complicated statistical models to better isolate differences in outcomes among completers of different TPPs, and the aggregate data reported in most public dashboards created by states to report on TPP outcomes. For example, collaboration is needed to consider how to distinguish between the selection and development of TPP candidates (Goldhaber & Ronfeldt, 2020), how to account for contextual differences across TPPs and the schools in which their graduates or completers go on

to teach, and how to create resources and incentives for such data systems to integrate into TPP practice in meaningful ways (Peck et al., 2021).

b. Work with states to align reporting structures to the goals that the data are intended to serve

Just as measures can be considered for their capacity for alignment to continuous improvement, data use can also meet accountability-, improvement-, and equity-oriented goals in well-informed and balanced ways. Outcome data are often expected to be able to serve multiple purposes with equal facility, but such expectations may exceed what is realistic or possible. Instead, data should be collected and analyzed with particular goals in mind. For example, Tennessee collects and publishes a variety of data for TPP accountability. The state's TPP report cards include both "scored" metrics and "unscored" metrics, and the difference lies in their eventual use: the former are included in the final ratings of TPPs, while the latter are provided for only informational purposes. Along these same lines, the state provides some information for public reporting purposes while also creating more detailed reporting, specific to TPPs, that is provided solely for internal use. California makes a similar distinction, sharing program-level survey results with TPPs while public reporting on the CTC website is at the aggregate level. This matching of goals and uses for data, shaping how and to whom data are reported and how they are applied to accountability and improvement, models a pragmatic approach that goes beyond the simple application of "outcomes" to provide TPP ratings.

Recommendation # 3: Provide leadership in key efforts to update standards and structures for the teaching profession

Bearing in mind the limitations of the influence and leverage of accreditation highlighted above, TPP accreditors have an important role to play in advancing the teaching profession. Indeed, the existence of TPP accreditation is an important indicator that teaching can be considered a profession. That said, since the establishment of NCATE, the actions of TPP accreditors have been largely downstream of trends in educational practice and policy. Because such a stance may not be sustainable in today's unstable and rapidly changing environment, we hold that TPP accreditors would benefit from adopting proactive orientation toward new developments in education, and particularly toward emerging structures and practices in teaching. If changes are coming to schools, and to teaching, accreditors should explore ways to steer these changes and to ensure that they are reflective of professional consensus and engagement. Below we describe several areas where one or more TPP accreditors could take on a leadership role in education and the teaching profession.

a. Advocate for, and participate in, the revision and update of teacher standards

The implementation of national accreditation has been hampered by an ongoing lack of agreement on essential professional requirements for teachers. Recent advances in understanding of the science of learning and development have opened new oppor-

tunities to create educational systems and experiences that give all students access to meaningful, high-quality opportunities to learn and grow (Cantor et al., 2019; Darling-Hammond et al., 2019; National Academies of Sciences, Engineering, and Medicine, 2018; Osher et al., 2020). These ongoing developments in research, applied to practice, are bringing forth proposals to redesign schooling to align to how students learn and develop (Learning Policy Institute & Turnaround for Children, 2021). Taken together, these advances have created an opportunity to revisit the establishment of a common base of understanding around what new teachers should know and be able to do.

Such a base of understanding would take the form of teacher standards that represent a professional consensus, per Imig and Imig (2008), about the requirements for professional certification that signal readiness to practice. These teacher standards would, most likely, take the form of a revision of the widely adopted InTASC standards, which have not been updated in a decade and are due for a refresh. We also recommend that this revision focus specifically on standards for new teachers and on the bar for readiness for professional practice. Such a focus would enable future close alignment with TPP accreditation standards and expectations, because they would define the practice for which candidates are being prepared. Additionally, professional standards need not define every aspect of professional practice, but they should be able to capture the shared values of the profession around essential skills and knowledge for teaching (Wineburg, 2006), a goal furthered by a focus on initial certification.

It could be argued that the creation of new teaching standards and their subsequent adoption in some form by states across the nation lie far outside the purview of TPP accreditors. As a counterargument, we assert that pursuing the goal of an accreditation system that is truly national in scope requires a firm foundation from which to act, and this foundation is formed, in part, by broadly accepted, high-quality standards that represent the professional consensus of the field and have gained the trust of policymakers and the public. The lack of national standards lies at the heart of the issues that weaken accreditation and the teaching profession, including the low percentage of programs accredited, the moves by policymakers to regulate (or deregulate) TPPs without input from the profession, and the uncertain value of accreditation for TPPs, candidates, and the public.

Accreditors cannot create such standards by themselves, but they should be tireless advocates for a broad movement to do so. Recall that the establishment of NCATE in 1954 brought together AACTE, NEA, and NASDTEC; in other words, organizations representing teacher educators, teachers, and state agency staff responsible for teacher licensure. The 2011 InTASC revisions were guided by the same triad of groups—teachers, teacher educators, and staffers from state agencies—that made up the committee responsible for seeing the revisions through (Paliokas et al., 2011). However, the range of organizations involved in the effort in other ways, including nominating the educators and state agency staff for committees, providing feedback as the work progressed, and disseminating the new standards when they were rolled out, was much broader (Council of Chief State School Officers, 2013), as can be seen in this list:

- American Association of Colleges for Teacher Education (AACTE)
- American Association of School Administrators (AASA)

- American Federation of Teachers (AFT)
- Association of Teacher Educators (ATE)
- Council for Exceptional Children (CEC)
- Learning Forward
- National Association for Gifted Children (NAGC)
- National Association of Elementary School Principals (NAESP)
- National Association of Secondary School Principals (NASSP)
- National Association of State Boards of Education (NASBE)
- National Association of State Directors of Special Education (NASDSE)
- National Association of State Directors of Teacher Education and Certification (NASDTEC)
- National Board for Professional Teaching Standards (NBPTS)
- National Commission on Teaching & America's Future (NCTAF)
- National Council for Accreditation of Teacher Education (NCATE)
- National Education Association (NEA)
- National School Boards Association (NSBA)
- National Teacher of the Year Program
- Teach for America (TFA)
- Teacher Education Accreditation Council (TEAC)

While we do not view this list as exhaustive or exclusive, we provide it here as an example, because a similar effort now would require the same engagement level as the 2011 effort.

b. Follow the revision of teacher standards with an aligned revision of accreditation standards, and account for updates to the uses of data and evidence

The alignment of TPP accreditation standards with teaching standards, first established nearly 30 years ago, makes the revision of accreditation standards the next logical step after the revision of teaching standards. This means that the precise shape of new accreditation standards cannot be known until a revision of teaching standards is completed. Even so, we can say upfront that the alignment of outcomes- and performance-based measures to quality assurance and continuous improvement should inform accreditation standards revisions. The measures that we recommend be the focus of new accreditation standards are those emphasized above: performance assessments, classroom observations, and surveys. Because these measures are also tied to our recommendations on building capacity for shared data use, we can also say that it will be important to connect standards revisions to data system improvement efforts.

As with changes to teaching standards, we acknowledge that no major changes to the workings of accreditation are possible without buy-in from across the country in the form of broad state participation and from across the field. It would probably require that one or more of the organizations that have led or been instrumental in past efforts, such as CCSSO, NASDTEC, NBPTS, NEA, and AFT, would have to step up to provide leadership in these efforts, along with the accreditors themselves, CAEP and AAQEP. The long list of organizations involved in teacher standards reform, provided in the section above, also applies here as an example of the scope and scale of participation needed. The exact shape of such a collaborative endeavor will likely only become clear

through implementation. What is certain, however, is that it will demand time, money, effort, and conviction, all on a national scale, and over months if not years.

c. Maintain common standards for all routes and pathways to teaching

Establishing and maintaining the boundaries of the teaching profession requires more than promulgating a set of standards and expectations for TPPs; it also means ensuring that all TPPs are held to the same accreditation standards, and by extension, that all teachers are held to the same standards for entry into the profession. In the abstract, the idea that a profession would require some standardization of preparation and initial practice seems fairly straightforward, but this has seldom been the case for education. As explored in our historical review, common expectations across traditional and alternative TPPs have been hard to establish, frequently attacked, and sometimes undermined by policymakers themselves. Such differential treatment is a feature of state program approval systems, and the 2015 GAO survey results summarized in Table 2 suggest that states maintain different standards even when it comes to the data and evidence used to evaluate traditional and alternative TPPs.

Accreditors, in meeting their responsibilities to define and maintain the teaching profession, should use the same basis of professional standards for assessing both traditional and alternative TPPs. However, different program designs may need to employ data and evidence in ways that honor the context of each type of program. Such differences may influence the timing of data collection (e.g., when do you conduct completer or employer surveys asking about satisfaction with preparation when some candidates finish training before they begin teaching and others participate in concurrent training). In addition, the inferences made from similar measurement tools may be different depending on the context and nature of the TPP. While emergent research on performance assessments in preservice training indicates that the characteristics of a candidate's student teaching placement school may influence their edTPA scores (Bastian et al., 2021), it is reasonable to assume that teaching conditions for those in internship training programs may have an even stronger influence on performance assessments. This is not to say that student teachers and interns should not be assessed based on the same set of professional standards, but that accreditors may need to carefully consider how to modify the accreditation processes to take into account these contextual differences.

By refocusing accreditation's data use away from regulation and toward program improvement, conversations about evidence should focus on how each type of performance- or outcomes-based measure can support the goals and purposes of each TPP. Exceptions to this guideline arise only when TPP goals focus on circumventing standards and expectations for entry into the profession. There are TPPs focused on recruiting nontraditional students, serving high needs schools, and preparing educators for higher turnover fields like special education. Any of these types of programs might be seen to fall short of their peers according to some outcomes-based measures. If this were to occur, it would be up to the accreditor to determine that the TPP was, by virtue of its intended purpose, directly confronting a systemic problem in education.

Accreditors can acknowledge and account for this in how they assess these programs and support their continuous improvement through the conscientious use of evidence.

d. Set in motion the reevaluation of professional certification for teachers to ensure it keeps up with teacher preparation and school staffing structures

Related to the new routes and pathways to the classroom discussed above, there has also been recent policy action on changing K-12 school staffing structures. Residencies, Grow-Your-Own teacher programs, and a variety of alternative programs have for years been expanding available roles for novice educators. Instead of moving from "teacher candidate" to "teacher," they may serve as residents, apprentices, interns, or fellows (e.g., Darling-Hammond et al., 2019; Guha et al., 2016; Hirsch et al., 2001) and also as paraprofessionals or substitutes, with the latter roles providing funding for the former (Yun & DeMoss, 2020). States have already created credentials for teachers in training to accommodate alternative programs, and some are considering further revisions to licensure systems to recognize the levels of preparation and practice that are stepping stones to professional licensure. These include teaching apprentice and resident licenses for candidates with a variety of qualifications and responsibilities (e.g., Rosilez, 2021; Southern Regional Education Board, 2021). Such routes also open up possibilities for the recognition, in licensing and professional advancement, of mentoring and leadership roles for experienced educators.

More recently, differentiated school staffing models, in which teams of educators work together in a variety of roles that distribute the responsibilities of teachers, have been proposed or implemented around the country (Basile, 2022; Chiefs for Change & Johns Hopkins University Institute for Education Policy, 2020; Council of Chief State School Officers, 2020; Hassel & Hassel, 2020; Learning Policy Institute, 2021; Partnership for Children and Youth, 2021; Weber, 2020). While some of these models predate the COVID-19 pandemic, they have been held up as particularly applicable in responding to COVID-19-related instructional challenges and educator shortages. They are also touted as allowing for clinical teacher preparation through the participation of teacher residents in teams. While the makeup of these instructional teams is dependent on the particular model adopted, the integration of fully licensed teachers or lead teachers with a mix of specialists, residents, tutors, assistant teachers, and expanded and remote learning staff across these models represents a significant departure from the one teacher, one classroom model that has long dominated schooling in the United States.

The creation of new roles for educators, and new licensure and preparation systems aligned to these roles, also creates an opportunity for TPP accreditors. As we described in the introduction to this paper, there is more overlap between the functions of certification and licensure in education compared to other professions, and fewer opportunities for specialty certifications. Now, it seems that specialties are being created, and licensure systems are in some cases being revised accordingly, but there is no related effort to tie these specialties to professional certification aligned to professional standards. Working, as they do, at the confluence of preparation and practice, and strongly connected to teacher standards and TPP standards, accreditors are well positioned to advocate for and lead such an effort.

Bringing together a coalition like those that produced the NBPTS and InTASC stan-

dards, a TPP accreditor could convene a task force with the goal of defining the skills and knowledge needed for this emerging constellation of educator specializations. Next steps could include the creation of new professional standards and professional certifications linked to emerging educators' roles. Still, these steps would constitute a large-scale and long-term project that could unfold over time. In the short term, it would be enough for accreditors, by exercising leadership, to put educators in a proactive position related to new staffing structures, with enhancement of the profession as the goal.

Finally, the potential expansion of certification as a tool for defining differentiated knowledge, skills, and roles in education presents an intriguing possibility: what if some of the struggles to define common standards for educator preparation and initial practice have been rooted in a lack of such differentiation? In other words, is it possible that the complexity and scope of the work of teaching, as defined in the one teacher, one classroom model, is a contributing factor in the lack of consensus about what new teachers should know and be able to do? Answering such questions in the affirmative opens up new possibilities for accreditation. If new specialties were created within teaching, and if some of the knowledge and skills currently encompassed by teaching standards were distributed among these specialties, this could simplify, at least somewhat, the task of defining broadly accepted standards for preparation and initial practice. The creation of new specializations and certifications could, therefore, not only build out a new professional infrastructure for teaching but also make possible a new consensus about the what, how, and why of teaching.

CONCLUSION

Historically, efforts to establish, expand, and update national TPP accreditation have been met with mixed and uneven success. The same can be said about the efforts to reach settled agreements about the use of performance- and outcomes-based measures, data, and evidence in assessing TPPs. Having described these developments in detail here, we can make no claim to have picked out the one set of recommendations that creates the perfect formula for reform. However, we have built these recommendations on a vital goal that emerged from our analysis: the necessity of realigning accreditation to its professional roots and purposes. Whether these recommendations present the most direct or effective path to this goal, our historical review of TPP accreditation and close examination of performance- and outcomes-based assessment of TPPs have brought us to the conclusion that the goal itself is sound.

Our recommendations are intended to improve the reach and relevance of TPP accreditation and enhance the quality assurance and continuous improvement efforts in teacher preparation. Of these recommendations, some can be carried out by an accreditor, or perhaps two accreditors, while others will require collaboration with states and other stakeholders across the educational landscape. We did not choose these recommendations for their feasibility or the capacity that accreditors might have to carry them out, though we did take these considerations into account. We did choose them out of a conviction that TPP accreditors, and the teaching profession, must rise to current challenges. As this report goes to publication, our education system is under great strain. And yet, as with any time of crisis, this moment presents not only challenges but also opportunities. If major change is coming to education and to TPP accreditation in

these uncertain times, it is our hope that the recommendations made here can inform that change and make a positive contribution to it.

REFERENCES

Accreditation Study Report. (1965). *The American Mathematical Monthly*, *72*(7), 775-777.

Accrediting commission advises members not to apply to NCATE. (1965). *The Phi Delta Kappan*, *46*(10), 482.

Alder Graduate School of Education. (2021). *Alder Graduate School of Education*. https://aldergse.edu/about/accreditation.

American Bar Association. (n.d.). *Frequently asked questions*. https://www.americanbar.org/groups/legal_education/resources/frequently_asked_questions.

American Board of Medical Specialties. (2021). *What ABMS board certification?* https://www.abms.org/board-certification.

Association for Advancing Quality in Educator Preparation. (2021a). *About us*. https://aaqep.org/about-us.

Association for Advancing Quality in Educator Preparation. (2021b). *Guide to AAQEP accreditation*. https://aaqep.org/files/2021%20Guide%20to%20AAQEP%20Accreditation.pdf.

Bales, B. L. (2015). Restructuring teacher education in the United States: Finding the tipping point. *Athens Journal of Education*, *2*(4), 297-312.

Ballou, D., & Springer, M. G. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher*, *44*(2), 77-86.

Basile, C. G. (2022). Arizona State University: A learning enterprise supporting P-12 education in the COVID-19 pandemic. In F. M. Reimers & F. J. Marmolejo (Eds.), *University and school collaborations during a pandemic sustaining educational opportunity and reinventing education*. Springer.

Bastian, K. C. (2018). *edTPA in North Carolina: Early evidence on candidate performance and predictive validity*. Education Policy Initiative at Carolina. https://epic.unc.edu/wp-content/uploads/sites/1268/2021/08/edTPA-ResearchBrief-2018-r3_Final.pdf.

Bastian, K. C., Fortner, C. K., Chapman, A., Fleener, M. J., McIntyre, E., & Patriarca, L. (2016). Data sharing to drive the improvement of teacher preparation programs. *Teachers College Record*, *118*(12), 1-29.

Bastian, K. C., Lys, D., & Pan, Y. (2018a). 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., Patterson, K. M., & Pan, Y. (2018b). Evaluating teacher preparation programs with teacher evaluation ratings: Implications for program accountability and improvement. *Journal of Teacher Education*, *69*(5), 429-447.

Bastian, K. C., Sun, M., & Lynn, H. (2021). What do surveys of program completers tell us about teacher preparation quality? *Journal of Teacher Education*, *72*(1), 11-26.

Bennett, C. I., McWhorter, L. M., & Kuykendall, J. A. (2006). Will I ever teach? Latino and African American students' perspectives on Praxis I. *American Educational Research Journal*, *43*(3), 531-575.

Bestor, A. (1953). *Educational wastelands: The retreat from learning in our public schools*. University of Illinois Press.

Blue, T. W., O'Grady, R. J., Toro, J. A., & Newell, E. A. (2002). *How do we find the best teachers? A study of the relationships among SAT, GPA, Praxis series test scores, and teaching ratings*. Paper presented at the annual conference of the Association of Teacher Educators. https://eric.ed.gov/?id=ED467764.

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

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). *The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools*. National Bureau of Economic Research.

Boyd, R. G. (1973). The development of accreditation and its influence upon curriculum development in higher education. *Journal of Thought*, *8*(3), 188-196.

Bradley, A. (1997, October 29). Accreditors shift toward performance: NCATE to stress candidate evaluation. *EducationWeek*, *17*(9), 1-2.

Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, *66*(2), 103-115.

Burns, K. (2016, November 14). *Accreditation not the destination for all schools*. *Politifact.* Poynter Institute. https://www.politifact.com/truth-o-meter/promises/obameter/promise/252/require-all-schools-of-education-to-be-accredited.

California Teaching Performance Assessment. (2021). *Performance assessment overview: Overview of instructional cycles and rubrics version 04.1*. California Commission on Teacher Credentialing. http://www.ctcexams.nesinc.com/Content/Docs/CalTPA_Assessment_Overview.pdf.

Campbell, S. L., & Ronfeldt, M. (2018). Observational evaluation of teachers: Measuring more than we bargained for? *American Educational Research Journal*, *55*(6), 1233-1267.

Cantor, P., Osher, D., Berg, J., Steyer, L., & Rose, T. (2019). Malleability, plasticity, and individuality: How children learn and develop in context. *Applied Developmental Science*, *23*(4), 307-337.

Cardichon, J., Darling-Hammond, L., Yang, M., Scott, C., Shields, P. M., & Burns, D. (2020). *Inequitable opportunity to learn: Student access to certified and experienced teachers*. Learning Policy Institute. https://learningpolicyinstitute.org/product/crdc-teacher-access-report.

Carnegie Task Force on Teaching as a Profession. (1986). *A nation prepared: Teachers for the 21st century* (p. 60). Carnegie Forum on Education and the Economy.

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

Carver-Thomas, D., & Darling-Hammond, L. (2017). *Teacher turnover: Why it matters and what we can do about it*. Learning Policy Institute.

Chen, B., Cowan, J. Goldhaber, D., & Theobald, R. (2021). *From the clinical experience to the classroom: Assessing the predictive validity of the Massachusetts Candidate Assessment of Performance*. National Center for Analysis of Longitudinal Data in Educational Research. https://caldercenter.org/publications/clinical-experience-classroom-assessing-predictive-validity-massachusetts-candidate.

Chiefs for Change & Johns Hopkins University Institute for Education Policy. (2020, May). *The return: How should education leaders prepare for reentry and beyond?* https://chiefsforchange.org/wp-content/uploads/2020/06/CFC-TheReturn_5-13-20.pdf.

Chung, B. W., & Zou, J. (2021). *Teacher licensing, teacher supply, and student achievement: Nationwide implementation of edTPA*. EdWorkingPaper: 21-440. Annenberg Institute at Brown University. https://doi.org/10.26300/ppz4-gv19.

Clinton, W. (1986). Why we test teachers in Arkansas. *Teacher Education Quarterly*, *13*(3), 28-30.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673-682.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources*, *45*(3), 655-681.

Cochran-Smith, M. (2021). Relocating teacher preparation to new graduate schools of education. *The New Educator*, *17*, 1-20.

Cochran-Smith, M., Baker, M., Burton, S., Carney, M. C., Chang, W-C., Fernández, M. B., & Stern, R. (2017). Teacher quality and teacher education policy: The U.S. case and its implications. In M. Akiba & G. LeTendre (Eds.), *The Routledge international handbook of teacher quality and policy* (pp. 445-462). Routledge.

Cochran-Smith, M., & Reagan, E. M. (2021). *"Best practices" for evaluating teacher preparation programs*. National Academy of Education Committee on Evaluating and Improving Teacher Preparation Programs. https://naeducation.org/evaluating-and-improving-teacher-preparation-programs-commissioned-paper-series.

Cochran-Smith, M., & Zeichner, K. M. (Eds.). (2009). *Studying teacher education: The report of the AERA panel on research and teacher education*. Routledge.

Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher*, *45*(6), 378-387.

Collins, J. W., & O'Brien, N. P. (Eds.). (2003). *The Greenwood dictionary of education*. Greenwood Publishing Group.

Comb, M., Cowan, J., Goldhaber, D., Jin, Z., & Theobald, R. (2021). *State ratings of educator preparation programs: Connecting program review to teacher effectiveness*. CALDER Working Paper No. 249-0321. https://caldercenter.org/publications/state-ratings-educator-preparation-programs-connecting-program-review-teacher.

Conant, J. B. 1963. *The education of American teachers*. McGraw-Hill.
Congressional Budget Office. (1993, May). *The federal role in improving elementary and secondary education*. https://www.cbo.gov/sites/default/files/103rd-congress-1993-1994/reports/199305federalroleinimproving.pdf.
Cooperman, S., & Klagholz, L. (1985). New Jersey's alternate route to certification. *Phi Delta Kappan*, *66*(10), 691-695.
Corcoran, S. P. (2010). *Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice. Education Policy for Action Series*. Annenberg Institute for School Reform at Brown University. https://eric.ed.gov/?id=ED522163.
Council for Higher Education Accreditation. (2006). *FACT SHEET #5—Accrediting organizations in the United States: How do they operate to assure quality?* https://www.chea.org/sites/default/files/other-content/Fact-Sheet-US-Accrediting-Organizations-How-They-Assure-Quality.pdf.
Council for the Accreditation of Educator Preparation. (2013). *CAEP accreditation standards*. caep-2013-accreditation-standards.pdf.
Council for the Accreditation of Educator Preparation. (2020). *CAEP consolidated handbook*. http://caepnet.org/~/media/Files/caep/accreditation-resources/caep-handbook-final.pdf?la=en.
Council for the Accreditation of Educator Preparation. (2021). *60 schools from 26 states recognized for national excellence in educator prep*. http://caepnet.org/about/news-room/60-schools-from-26-states-recognized.
Council of Chief State School Officers. (2011). *Interstate Teacher Assessment and Support Consortium (InTASC) model core teaching standards: A resource for state dialogue*. https://ccsso.org/sites/default/files/2017-11/InTASC_Model_Core_Teaching_Standards_2011.pdf.
Council of Chief State School Officers. (2013). *InTASC model core teaching standards and learning progressions for teachers 1.0*. https://ccsso.org/sites/default/files/2017-12/2013_INTASC_Learning_Progressions_for_Teachers.pdf.
Council of Chief State School Officers. (2017). *Transforming educator preparation: Lessons learned from leading states*. https://ccsso.org/sites/default/files/2017-11/CCSSO%20Educator%20Preparation%20Playbook.pdf.
Council of Chief State School Officers. (2020). *Restart & recovery: Considerations for teaching and learning: System conditions*. https://learning.ccsso.org/considerations-for-teaching-and-learning-system-conditions?related_post_from=842.
Cowan, J., Goldhaber, D., Jin, Z., & Theobold, R. (2020). *Teacher licensure tests: Barrier or predictive tool?* Working Paper No. 245-1020. https://files.eric.ed.gov/fulltext/ED609773.pdf.
Croft, S. J., Roberts, M. A., & Stenhouse, V. L. (2015). The perfect storm of education reform: High-stakes testing and teacher evaluation. *Social Justice*, *42*(139), 70-92.
Crowe, E. (2011). *Race to the top and teacher preparation: Analyzing state policies for ensuring real accountability and fostering program innovation*. Center for American Progress. https://files.eric.ed.gov/fulltext/ED518517.pdf.
Darling-Hammond, L. (1985). *Beyond the commission reports: The coming crisis in teaching*. RAND Corporation.
Darling-Hammond, L. (1997). *Doing what matters most: Investing in quality teaching*. National Commission on Teaching & American's Future. https://files.eric.ed.gov/fulltext/ED415183.pdf.
Darling-Hammond, L. (1999). *Reshaping teaching policy, preparation, and practice: Influences of the National Board for Professional Teaching Standards*. National Partnership for Excellence and Accountability in Teaching, American Association of Colleges for Teacher Education.
Darling-Hammond, L. (2007). Evaluating No Child Left Behind. *The Nation*, *284*(20), 11-21.
Darling-Hammond, L. (2010). *Evaluating teacher effectiveness: How teacher performance assessments can measure and improve teaching*. Center for American Progress.
Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher*, *44*(2), 132-137.
Darling-Hammond, L. (2019). *Investing for student success: Lessons from state school finance reforms*. Learning Policy Institute. https://learningpolicyinstitute.org/product/investing-student-success-school-finance-reforms.
Darling-Hammond, L. (2020). Accountability in teacher education. *Action in Teacher Education*, *42*(1), 60-71.
Darling-Hammond, L., & Bransford, J. (Eds.). (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. John Wiley & Sons.

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2019). Implications for educational practice of the science of learning and development. *Applied Developmental Science*, 24(2), 97-140.

Darling-Hammond, L., Newton, S. P., & Wei, R. C. (2013). Developing and assessing beginning teacher effectiveness: The potential of performance assessments. *Educational Assessment, Evaluation and Accountability*, *25*(3), 179-204.

Darling-Hammond, L., Oakes, J., Wojcikiewicz, S., Hyler, M. E., Guha, R., Podolsky, A., Kini, T., Cook-Harvey, C., Mercer, C., & Harrell A. (2019). *Preparing teachers for deeper learning*. Harvard Education Press.

De Voto, C., Olson, J. D., & Gottlieb, J. J. (2021). Examining diverse perspectives of edTPA policy implementation across states: The good, the bad, and the ugly. *Journal of Teacher Education*, *72*(1), 42-55.

DeMonte, J. (2013). *Who is in charge of teacher preparation?* Center for American Progress. https://www.americanprogress.org/issues/education-k-12/reports/2013/06/17/66753/who-is-in-charge-of-teacher-preparation.

Duncan, A. (2009). *Policy address on teacher preparation*. Teachers College, Columbia University. https://www.tc.columbia.edu/articles/2009/october/arne-duncan-full-transcript.

Education Next. (2010). *Kill HQT ASAP*. https://www.educationnext.org/kill-hqt-asap.

Educational Testing Service. (2017). *Fact sheet on the PPAT assessment and edTPA*. https://www.ets.org/s/ppa/pdf/PPAT-Facts.pdf.

Erickson, T. (2015). *CAEP board names Chris Koch interim president*. Council for the Accreditation of Educator Preparation. http://caepnet.org/about/news-room/news-archive/caep-board-names-interim-president.

Evans, C. M. (2017). Predictive validity and impact of CAEP standard 3.2: Results from one master's-level teacher preparation program. *Journal of Teacher Education*, *68*(4), 363-376. https://www.nciea.org/sites/default/files/publications/Evans%202017.pdf.

Farenga, S. J., & Ness, D. (2017). SCALE Down, SCALE Back!: Academic freedom under siege through standards proliferation by para-educational enterprises. In *Alternatives to Privatizing Public Education and Curriculum* (pp. 78-98). Routledge.

Fenwick, L. T. (2021). *A tale of two cities: State evaluation systems of teacher preparation programs*. American Association of Colleges for Teacher Education. https://naeducation.org/landscape-analysis-of-state-evaluation-systems-of-teacher-preparation-programs.

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.

Fisher, R. (1986). Update on NCATE. *The Science Teacher*, *53*(2), 35-37.

Floden, R. E. (1980). Flexner, accreditation, and evaluation. *Educational Evaluation and Policy Analysis*, *2*(2), 35-46.

Flores, A. R. (2015). Docket ID ED-2014-OPE-0057. Public comment letter. Hispanic Association of Colleges and Universities. https://www.hacu.net/images/hacu/govrel/HACU%20Ltr%20on%20TP%20NPRM.013015.pdf.

Fraser, J. W. & Lefty, L. (2018a). *Three turbulent decades in the preparation of American teachers: Two historians examine reforms in education schools and the emergence of alternative routes to teaching*. Institute for Education Policy, Johns Hopkins School of Education. fraserleftyfinal.pdf.

Fraser, J. W. & Lefty, L. (2018b). *Teaching teachers: Changing paths and enduring debates*. Johns Hopkins University Press.

Fullan, M. (1994). *Change forces: Probing the depths of educational reform*. School Development and the Management of Change Series. Falmer Press. https://files.eric.ed.gov/fulltext/ED373391.pdf.

Gansle, K. A., Noell, G. H., & Burns, J. M. (2012). Do student achievement outcomes differ across teacher preparation programs? An analysis of teacher education in Louisiana. *Journal of Teacher Education*, *63*(5), 304-317.

Garcia, A., & Muñiz, J. (2020). Piloting performance-based licensure to ease testing disparities. In *Mississippi's Multifaceted Approach to Tackling Teacher Shortages* (pp. 15-18). New America. http://www.jstor.org/stable/resrep34415.6.

Gardner, W. E., Scannell, D., & Wisniewski, R. (1996). The curious case of NCATE redesign. *The Phi Delta Kappan*, *77*(9), 622-629.

Garrett, R., & Steinberg, M. P. (2015). Examining teacher effectiveness using classroom observation scores evidence from the randomization of teachers to students. *Educational Evaluation and Policy Analysis*, *37*(2).

Gay, L. R., & Daniel, K. F. (1972). Accreditation and performance-based teacher education. *Educational Technology*, *12*(11), 45-50.
Gideonse, H. D. (1993). Appointments with ourselves: A faculty argument for NCATE. *The Phi Delta Kappan*, *75*(2), 174-180.
Giovannetti, J. (2012, August). The chronicles of the Renaissance Group. *The Chronicles of the Renaissance Group*, *1*(1), 11-25. https://educationalrenaissance.org/index.php/edren/article/download/34/pdf.
Gitomer, D. H., Martínez, J. F., Battey, D., & Hyland, N. E. (2021). Assessing the assessment: Evidence of reliability and validity in the edTPA. *American Educational Research Journal*, *58*(1), 3-31.
Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources*, *42*(4), 765-794.
Goldhaber, D. (2015). Exploring the potential of value-added performance measures to affect the quality of the teacher workforce. *Educational Researcher*, *44*(2), 87-95.
Goldhaber, D., & Cowan, J. (2014). Excavating the teacher pipeline: Teacher preparation programs and teacher attrition. *Journal of Teacher Education*, *65*(5), 449-462.
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., & Hansen, M. (2010). Race, gender, and teacher testing: How informative a tool is teacher licensure testing? *American Educational Research Journal*, *47*(1), 218-251.
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.
Goldhaber, D., & Ronfeldt, M. (2020). Toward causal evidence on effective teacher preparation. *Linking Teacher Preparation Program Design and Implementation to Outcomes for Teachers and Students*, 211-236.
Goodlad, J. I. (1990). *Teachers for our nation's schools*. Jossey-Bass Publishers.
Goodson, L. (2018). Karen Symms Gallagher: CAEP chair shares insight into teacher preparation, accreditation. *Educational Considerations*, *44*(1).
Government Accountability Office. (2015). *Teacher preparation programs: Education should ensure states identify low-performing programs and improve information sharing*. GAO-15-598. Government Accountability Office. https://www.gao.gov/products/gao-15-598.
Grissom, J. A., & Bartanen, B. (2022). Potential race and gender biases in high-stakes teacher observations. *Journal of Policy Analysis and Management*, *41*(1), 131-161.
Grissom, J. A., & Youngs, P. (2015). *Improving teacher evaluation systems: Making the most of multiple measures*. Teachers College Press.
Groves, S. R. (2019). *Ethical considerations in the Council for the Accreditation of Educator Preparation*. Doctoral dissertation. University of Oklahoma Graduate College. https://shareok.org/bitstream/handle/11244/320326/2019_Groves_Shelley_Dissertation.pdf?sequence=1.
Gubser, L. (1980). NCATE's director comments on the Tom critique. *The Phi Delta Kappan*, *62*(2), 117-119.
Guha, R., Hyler, M. E., & Darling-Hammond, L. (2016). *The teacher residency: An innovative model for preparing teachers*. Learning Policy Institute.
Haertel, E. H. (2013). Reliability and validity of inferences about teachers based on student scores. William H. Angoff Memorial Lecture Series. In *Educational Testing Service*. https://eric.ed.gov/?id=ED560957.
Hail, C., Hurst, B., Chang, C., & Cooper, W. (2019). Accreditation in education: One institution's examination of faculty perceptions. *Critical Questions in Education*, *10*(1), 17-28.
Hartocollis, A. (2005, July 31). *Teaching the teachers: Who needs education schools? The New York Times*. https://www.nytimes.com/2005/07/31/us/education/teaching-the-teachers-who-needs-education-schools.html.
Hasbun, T. C., & Rudolph, A. (2016). Navigating the waters of accreditation: Best practices, challenges, and lessons learned from one institution. *SAGE Open*, *6*(2).
Hassel, B., & Hassel, E. A. (2020). Rethinking school staffing. In F. M. Hess & B. L. Wright (Eds.), *Getting the most bang for the education buck*. Teachers College Press.
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.
Herbert, B. (2000, January 10). *In America: The teacher crisis. The New York Times*. https://www.nytimes.com/2000/01/10/opinion/in-america-the-teacher-crisis.html.

Herold, J. (1974). Sputnik in American education: A history and reappraisal. *McGill Journal of Education/ Revue des sciences de l'éducation de McGill*, *9*(2).

Hess, F. (2006). *Schools of reeducation? The Washington Post*. https://www.washingtonpost.com/archive/opinions/2006/02/05/schools-of-reeducation/aac40641-0d0a-4d81-ad4d-d4c99af1aaca.

Hill, H., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review*, *83*(2), 371-384.

Hirsch, E., Koppich, J., & Knapp, M. (2001). Revisiting what states are doing to improve the quality of teaching. *An update on patterns and trends*. Center for the Study of Teaching and Policy, University of Washington.

Ho, A. D., & Kane, T. J. (2013). *The reliability of classroom observations by school personnel*. Research Paper. MET Project. Bill & Melinda Gates Foundation.

Hofstadter, R. (1963). *Anti-intellectualism in American Life*. Alfred A. Knopf.

Hollins, E. R., & Warner, C. K. (2021). *Evaluating the clinical component of teacher preparation programs*. National Academy of Education Committee on Evaluating and Improving Teacher Preparation Programs. National Academy of Education.

Holmes Group. (1986). *Tomorrow's teachers: A report of the Holmes Group*. Holmes Group.

Imig, D. G., & Imig, S. R. (2008). From traditional certification to competitive certification: A twenty-five year retrospective. In M. Cochran-Smith, S. Feiman-Nemser, D. J. McIntyre, & K. E. Demers (Eds.), *Handbook of research on teacher education: Enduring questions in changing contexts* (3rd ed., pp. 886-907). Routledge, Taylor & Francis Group, and Association of Teacher Educators.

Interstate New Teacher Assessment and Support Consortium. (1992). *Model standards for beginning teacher licensing, assessment and development: A resource for state dialogue*. Council of Chief State School Officers.

Jackson Public Schools. (2021). *Pathways to teaching: The education programs and partnerships*. https://www.jackson.k12.ms.us/Page/13416.

Jacobson, L. (2020). *Is edTPA standing in the way of getting more teachers into classrooms? K-12 dive*. https://www.k12dive.com/news/is-edtpa-standing-in-the-way-of-getting-more-teachers-into-classrooms/572969.

Jiang, J. Y., & Sporte, S. E. (2016). *Teacher evaluation in Chicago: Differences in observation and value-added scores by teacher, student, and school characteristics*. Research Report. University of Chicago Consortium on School Research. https://consortium.uchicago.edu/publications/teacher-evaluation-chicago-differences-observation-and-value-added-scores-teacher.

Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record*, *114*(10), 1-39.

Jones, D. W. (1991). NCATE has made the grade: It is a valuable contributor to professional quality control. *Teacher Education Quarterly*, *18*(4), 47-56.

Jones, N. D., Bell, C. A., Brownell, M., Qi, Y., Peyton, D., Pau, D., Fowler, M., & Holtzman, S. (2021). *Using classroom observations in the evaluation of special education teachers*. Boston University Wheelock Educational Policy Center. https://wheelockpolicycenter.org/effective-teachers/sped-teacher-eval.

Kaiser Permanent Nursing. (2018). *List of approved specialty certifications*. https://kpnursing.org/_HI/professionaldevelopment/docs/List%20of%20Approved%20Specialty%20Certifications_HNA_5.2018.pdf.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Research Paper. MET Project. Bill & Melinda Gates Foundation.

Koedel, C., Parsons, E., Podgursky, M., & Ehlert, M. (2015). Teacher preparation programs and teacher quality: Are there real differences across programs? *Education Finance and Policy*, *10*(4), 508-534.

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

Kreighbaum, A. (2016, October 13). *New accountability for teacher prep*. *Inside Higher Education*. https://www.insidehighered.com/news/2016/10/13/obama-administration-releases-final-rules-teacher-preparation-programs.

Learning Policy Institute. (2021). *How can states and districts use federal recovery funds strategically? Accelerating learning through expanded learning time*. https://learningpolicyinstitute.org/sites/default/files/product-files/Federal_Funds_ELT_FACTSHEET.pdf.

Learning Policy Institute & Turnaround for Children. (2021). *Design principles for schools: Putting the science of learning and development into action*. https://k12.designprinciples.org.
Levine, A. (2006). *Educating school teachers*. The Education Schools Project.
Levine, A. (2015, April). *Accountability: A gift to teacher education*. *The Huffington Post*. https://www.huffpost.com/entry/accountability-a-gift-to-_b_7011302.
Lilly, M. S. (1992). Research on teacher licensure and state approval of teacher education programs. *Teacher Education and Special Education*, *15*(2), 148-160.
Loews, L. & Sawchuk, S. (2017, October 12). *Yet another group sets out to accredit teacher-prep programs*. *EdWeek*. https://www.edweek.org/teaching-learning/yet-another-group-sets-out-to-accredit-teacher-prep-programs/2017/10.
Many, J. E., & Bhatnagar, R. (2017). *Implementing and analyzing performance assessments in teacher education*. Information Age Publishing.
Massachusetts Department of Elementary and Secondary Education. (2019). *Guidelines for candidate assessment of performance: Assessment of teacher candidates*. https://www.doe.mass.edu/edprep/cap/guidelines.html.
Mehta, J. (2013). From bureaucracy to profession: Remaking the educational sector for the twenty-first century. *Harvard Educational Review 83*(3), 463–488.
Middle States Commission on Higher Education. (2021). *About us*. https://www.msche.org/about-us.
Mihaly, K., McCaffrey, D., Sass, T. R., & Lockwood, J. R. (2013). Where you come from or where you go? Distinguishing between school quality and the effectiveness of teacher preparation program graduates. *Education Finance and Policy*, *8*(4), 459-493.
Mitchell, E. (1989). Teacher education in the Reagan revolution. *Teacher Education Quarterly*, *16*(2), 49-62.
Moore, J. W. (2002). *Perceived barriers to the national board for professional teaching standards certification*. East Tennessee State University. https://dc.etsu.edu/cgi/viewcontent.cgi?article=1867&context=etd.
Moore, K. D., Hopkins, S., & Tullis, R. (1994). NCATE standards: Restructuring teacher education. *Teacher Education Quarterly*, *21*(2), 27-36.
Murray, F. B. (2005). On building a unified system of accreditation in teacher education. *Journal of Teacher Education, 56*(4), 307-317.
Murray, F. B. & Wise, A. E. (2009). *Towards a unified accreditation system for educator preparation*. Teacher Education Accreditation Council. http://www.teac.org/wp-content/uploads/2009/03/unified.pdf.
National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. The National Academies Press. https://doi.org/10.17226/24783.
National Board for Professional Teaching Standards. (1992). *What teachers should know and be able to do*. National Board for Professional Teaching Standards.
National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. National Commission on Excellence in Education.
National Council for Accreditation of Teacher Education. (2002). *Professional standards: Accreditation of schools, colleges, and departments of education*. http://www.hunter.cuny.edu/~ct/ScarletN/Ncate Resources/unit_stnds_2002.pdf.
National Council for Accreditation of Teacher Education. (2008). *Professional standards for the accreditation of teacher preparation institutions*. https://files.eric.ed.gov/fulltext/ED502043.pdf.
National Council on Teacher Quality. (2021). *Program performance measures national results*. State Teacher Policy Database. Data set. https://www.nctq.org/yearbook/national/Program-Performance-Measures-89.
National Education Association. (2021). *Alternative routes to certification state-by-state analysis*. https://www.nea.org/sites/default/files/2021-04/Alternative%20Paths%20Report-Final%20Apr2020.pdf.
National Research Council. (2001). *Testing teacher candidates: The role of licensure tests in improving teacher quality*. National Academy Press. https://doi.org/10.17226/10090.
National Research Council. (2010). *Preparing teachers: Building evidence for sound policy*. The National Academies Press. https://doi.org/10.17226/12882.
Nettles, M. T., Scatton, L. H., Steinberg, J. H., & Tyler, L. L. (2011). Performance and passing rate differences of African American and White prospective teachers on Praxis Examinations: A joint project of the National Education Association (NEA) and Educational Testing Service (ETS). *ETS Research Report Series*, (1), i-82.

Newton, S. (2010). *Preservice performance assessment and teacher early career effectiveness: Preliminary findings on the performance assessment for California teachers*. Stanford University.
Nguyen, T. D., Pham, L. D., Crouch, M., & Springer, M. G. (2020). The correlates of teacher turnover: An updated and expanded meta-analysis of the literature. *Educational Research Review*, *31*, 100355.
Nodine, T. R. (2016). How did we get here? A brief history of competency-based higher education in the United States. *The Journal of Competency-Based Education*, *1*(1), 5-11.
Ohio Department of Higher Education. (2022). *Measuring the performance of educator preparation programs*. https://www.ohiohighered.org/educator-preparation/transparency-accountability.
Osher, D., Cantor, P., Berg, J., Steyer, L., & Rose, T. (2020). Drivers of human development: How relationships and context shape learning and development. *Applied Developmental Science*, *24*(1), 6-36.
Paliokas, K., McWalters, P., & Diez, M. (2011). *The Interstate Teacher Assessment and Support Consortium (InTASC): An update on the model core teaching standards*. http://images.pearsonassessments.com/images/nes_publications/2011_03paliokasmcwaltersdiez.pdf.
Partelow, L. (2019). *What to make of declining enrollment in teacher preparation programs?* Center for American Progress. https://www.americanprogress.org/issues/education-k-12/reports/2019/12/03/477311/make-declining-enrollment-teacher-preparation-programs.
Partnership for Children and Youth. (2021). *The expanded learning workforce: Essential, frontline workers serving kids & families*. https://static1.squarespace.com/static/59f76b95268b96985343bb62/t/603ff28232f79c463ded419e/1614803593798/Expanded+Learning+Workforce+Fact+Sheet+-+02-24-21.pdf.
Pecheone, R. L. (2019). *Educative assessment and meaningful support: 2018 edTPA administrative report*. Pearson Education. https://secure.aacte.org/apps/rl/res_get.php?fid=4769&ref=rl.
Pecheone, R. L. (2021). *Educative assessment and meaningful support: 2019 edTPA administrative report*. Stanford Center for Assessment, Learning, and Equity. https://edtpa.org/resource_item/2019AR.
Peck, C. A., Cuthrell, K., Mace, D. H. P., Sloan, T., & Lys, D. B. (2021). *Using data to improve teacher education: Moving evidence into action*. Teachers College Press.
Peck, C. A., & McDonald, M. (2013). Creating "cultures of evidence" in teacher education: Context, policy, and practice in three high-data-use programs. *The New Educator*, *9*(1), 12-28.
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.
Peck, C. A., Youngs, M. G., & Zhang, W. (2021). *Using teaching performance assessments for program evaluation and improvement in teacher education.* National Academy of Education Committee on Evaluating and Improving Teacher Preparation Programs. https://naeducation.org/wp-content/uploads/2021/11/Revised-Final-pp-for-NAEd-EITPP-Paper-6-Peck_Young_Zhang.pdf.
Petchauer, E., Bowe, A. G., & Wilson, J. (2018). Winter is coming: Forecasting the impact of edTPA on Black teachers and teachers of color. *The Urban Review*, *50*(2), 323-343.
Pinsky, S., & Fry, V. (2013). *National briefing heralds launch of edTPA, releases field test data*. EdPrepMatters. https://edprepmatters.net/2013/11/national-briefing-heralds-launch-of-edtpa-releases-field-test-data.
Podolsky, A., Kini, T., Bishop, J., & Darling-Hammond, L. (2016). *Solving the teacher shortage: How to attract and retain excellent educators*. Learning Policy Institute. https://learningpolicyinstitute.org/product/solving-teacher-shortage.
Ponessa, J. (1997). *Despite rocky road, ed. school accreditation effort on a roll*. *EducationWeek.* https://www.edweek.org/teaching-learning/despite-rocky-road-ed-school-accreditation-effort-on-a-roll/1997/06.
Pool, J., Dittrich, C., Dittrich, C., Pool, K., & Hausfather, S. (2004). An analysis of SAT and Praxis I: Performance of teacher education candidates at three different types of institutions. *Action in Teacher Education*, *26*(2), 60-68.
Putman, H., & Walsh, K. (2021). *State of the states 2021: Teacher preparation policy*. National Council on Teacher Quality. https://www.nctq.org/publications/State-of-the-States-2021:-Teacher-Preparation-Policy#performance.
Reed, M. (2020, September 16). *No more regional accreditors: A small change that portends much larger changes. Inside Higher Ed*. https://www.insidehighered.com/blogs/confessions-community-college-dean/no-more-regional-accreditors.
Relay Graduate School of Education. (2021). *Institution*. https://www.relay.edu/institution.
Robinson, S. (2013, November 8). *AACTE celebrates official launch of edTPA. EdPrepMatters*. https://edprepmatters.net/2013/11/announcing-edtpa-field-test-results-recommended-passing-score.

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

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

Ronfeldt, M., Matsko, K. K., Greene Nolan, H., & Reininger, M. (2021). Three different measures of graduates' instructional readiness and the features of preservice preparation that predict them. *Journal of Teacher Education*, *72*(1), 56-71.

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

Rosilez, A. (2021, November 4). *Nontraditional licensure pathways & the year ahead*. PowerPoint presentation. Oregon Teacher Standards and Practices Commission. https://www.oregon.gov/tspc/Commission/Meetings/2021NOV/Item_1.5a&2.5a.pdf.

Roth, R. A. (1974). *Competency-based teacher education and certification definitions: Syntheses and schema*. Michigan Department of Education. https://files.eric.ed.gov/fulltext/ED095172.pdf.

Sass, T. R., Semykina, A., & Harris, D. N. (2014). Value-added models and the measurement of teacher productivity. *Economics of Education Review*, *38*, 9-23.

Sato, M. (2014). What is the underlying conception of teaching of the edTPA? *Journal of Teacher Education*, *65*(5), 421-434.

Sawchuk, S. (2016). *Teacher prep accreditation group seeks to regain traction*. *EducationWeek*. https://www.edweek.org/teaching-learning/teacher-prep-accreditation-group-seeks-to-regain-traction/2016/08.

Schalock, H. D., & Schalock, M. D. (2011). Teacher work sample methodology at Western Oregon University. In H. Rosselli, M. Brodsky, & M. Girod (Eds.), *Connecting teaching and learning: History, evolution and case studies of teacher work sample methodology* (pp. 1-24). Rowman and Littlefield.

Sedlak, M. (2008). Competing visions of purpose, practice, and policy: The history of teacher certification in the United States. In M. Cochran-Smith, S. Feiman-Nemser, D. J. McIntyre, & K. E. Demers (Eds.), *Handbook of research on teacher education: Enduring questions in changing contexts* (3rd edition). Routledge, Taylor & Francis Group, and Association of Teacher Educators.

Sedlak, M., & Schlossman, S. (1986). *Who will teach? Historical perspectives on the changing appeal of teaching as a profession*. Center for the Study of the Teaching Profession. RAND Corporation. https://www.rand.org/content/dam/rand/pubs/reports/2007/R3472.pdf.

Shuls, J. V. (2018). Raising the bar on teacher quality: Assessing the impact of increasing licensure exam cut-scores. *Educational Policy*, *32*(7), 969-992.

Sledge, A., & Pazey, B. L. (2013). Measuring teacher effectiveness through meaningful evaluation: Can reform models apply to general education and special education teachers? *Teacher Education and Special Education*, *36*(3), 231-246.

Southern Regional Education Board. (2021). *North Carolina pathways to excellence for teaching professionals*. https://www.sreb.org/sites/main/files/file-attachments/ncgraphic_dec8.pdf?1617045981.

Stanford Center for Assessment, Learning, and Equity. (2013, November). 2013 *edTPA field test summary report*. https://secure.aacte.org/apps/rl/res_get.php?fid=827.

Stanford Center for Assessment, Learning, and Equity. (2014). *edTPA breakfast meeting at AACTE's 66th annual meeting questions and answers*. https://secure.aacte.org/apps/rl/res_get.php?fid=1119&ref=edtpa/edtpa.aacte.org/faq.

Stanford Center for Assessment, Learning, and Equity. (2019). *Making good choices: Candidate support resource edTPA*. http://www.edtpa.com/PageView.aspx?f=GEN_GettingStarted.html.

Steinberg, M. P., & Donaldson, M. L. (2015). The new educational accountability: Understanding the landscape of teacher evaluation in the post-NCLB era. *Education Finance and Policy*, *11*(3), 340-359.

Steinberg, M. P., & Kraft, M. A. (2017). The sensitivity of teacher performance ratings to the design of teacher evaluation systems. *Educational Researcher*, *46*(7), 378-396.

Stoddert, T., & Floden, R. E. (1995). *Traditional and alternate routes to teacher certification: Issues, assumptions, and misconceptions*. https://edwp.educ.msu.edu/research/wp-content/uploads/sites/10/2020/11/IP952.pdf.

Strauss, V. (2015, February 13). *No Child Left Behind's test-based policies failed. Will Congress keep them anyway? The Washington Post*. https://www.washingtonpost.com/news/answer-sheet/wp/2015/02/13/no-child-left-behinds-test-based-policies-failed-will-congress-keep-them-anyway.

Sykes, G., & Plastrik, P. (1993). *Standard setting as educational reform: Trends and issues paper No. 8*. ERIC Clearinghouse on Teacher Education.

Tamir, E., & Wilson, S. M. (2005). Who should guard the gates? Evidentiary and professional warrants for claiming jurisdiction. *Journal of Teacher Education*, *56*(4), 332-342.

Tatto, M. T., Savage, C., Liao, W., Marshall, S. L., Goldblatt, P., & Contreras, L. M. (2016). The emergence of high-stakes accountability policies in teacher preparation: An examination of the U.S. Department of Education's proposed regulations. *EducationPolicy Analysis Archives*, *24*(21).

Tennessee State Board of Education. (2020). *2020 Tennessee educator preparation report card*. https://www.tn.gov/content/dam/tn/stateboardofeducation/documents/eprc/2020/2020_Technical_Report.pdf.

Texas Education Agency. (2022). *Principal and teacher surveys to evaluate Texas educator preparation programs*. https://tea.texas.gov/about-tea/news-and-multimedia/correspondence/taa-letters/principal-and-teacher-surveys-to-evaluate-texas-educator-preparation-programs-2.

Thomas B. Fordham Foundation. (2002). *Five-year report, 1997-2001*. http://www.edexcellencemedia.net/publications/2002/200205_tbfffiveyear/report.pdf.

Tom, A. R. (1980). NCATE standards and program quality: You can't get there from here. *Phi Delta Kappan*, *62*(2), 113-117.

Travelstead, C. C. (1963). Second of two opposed positions: NCATE yesterday, today, and tomorrow. *The Phi Delta Kappan*, *45*(1), 38-42.

University of Nebraska Medical Center. (2021). *LCME accreditation*. University of Nebraska Medical Center.

U.S. Department of Education. (2002). *Meeting the highly qualified teachers challenge: The Secretary's annual report on teacher quality*. https://files.eric.ed.gov/fulltext/ED513876.pdf.

U.S. Department of Education. (2013, November 14). *AACTE launches performance assessment for preservice teachers*. *The Teachers Edition*. https://content.govdelivery.com/accounts/USED/bulletins/9047c0.

U.S. Department of Education. (2020, February 26). *Re: final accreditation and state authorization regulations*. Official letter from Diane Auer Jones, principal deputy undersecretary delegated the duties of the undersecretary. https://sacscoc.org/app/uploads/2020/03/State-Authorization-Letter-w-Diane-Signature-2.26.19.pdf.

U.S. Department of Education. (2021). *History and context of accreditation in the United States*. https://www2.ed.gov/admins/finaid/accred/accreditation_pg2.html.

VanHouten, Z. (2015, 29 January). *Panel: Proposed federal regulations could have negative impact on teacher diversity*. https://edprepmatters.net/2015/01/panel-proposed-federal-regulations-could-have-negative-impact-on-teacher-diversity.

Vegari, S., & Hess, F. M. (2002). *The accreditation game*. http://www.nysed.gov/common/nysed/files/principal-project-file-57-the-accreditation-game-can-accreditation-ensure-quality-teacher-training.pdf.

Viadero, D. (1990). NCATE breaks with tradition in selecting RAND's Wise to become president. *EducationWeek*. https://www.edweek.org/education/ncate-breaks-with-tradition-in-selecting-rands-wise-to-become/1990/06.

Washington State Professional Educators Standards Board. (2021, May 5). *edTPA eliminated as a state requirement for teacher certification*. https://content.govdelivery.com/accounts/WAOSPI/bulletins/2d7d4ac.

Watts, D. (1986). Four educators comment on the redesign of NCATE. *The Phi Delta Kappan*, *68*(4), 315-318.

Weber, G. (2020, 24 September). *Addressing teacher shortages with differentiated staffing models*. American Institutes for Research. https://www.air.org/resource/addressing-teacher-shortages-differentiated-staffing-models.

Weingarten, R. (2015). *Re: Teacher Preparation Program NRPM*. Public comment letter. American Federation of Teachers. https://www.aft.org/sites/default/files/ltr_edteacherprepregs.pdf.

Weiss, E. (2013). *Mismatches in Race to the Top limit educational improvement*. Economic Policy Institute. https://www.epi.org/publication/race-to-the-top-goals.

Wheeler, C.W. (1980). *NCATE: Does it matter?* Institute for Research on Teaching, Michigan State University. https://edwp.educ.msu.edu/research/wp-content/uploads/sites/10/2020/11/rs092.pdf.

White House Archives of President Barack Obama. (n.d.). *Race to the Top*. https://obamawhitehouse.archives.gov/issues/education/k-12/race-to-the-top.

Whittaker, A., Pecheone, R., & Stansbury, K. (2018). Fulfilling our educative mission: A response to edTPA critique. *Education Policy Analysis Archives*, *26*(30). http://dx.doi.org/10.14507/epaa.26.3720.

Will, M. (2018). After five years a bold set of teacher prep standards still faces challenges. *EducationWeek, 38*(2). https://www.edweek.org/teaching-learning/after-five-years-a-bold-set-of-teacher-prep-standards-still-faces-challenges/2018/08.

Will, M. (2019, July 24). *Teacher-preparation programs again have a choice of accreditors: But should they*. *EducationWeek*. https://www.edweek.org/leadership/teacher-preparation-programs-again-have-a-choice-of-accreditors-but-should-they/2019/07.

Will, M. (2020, June 12). *Georgia eliminates the edTPA requirement for teacher candidates. EducationWeek*. https://www.edweek.org/teaching-learning/georgia-eliminates-the-edtpa-requirement-for-teacher-candidates/2020/06.

Will, M. (2021, June 22). *The complicated, divisive work of grading teacher-preparation programs: CAEP and AAQEP offer differing routes for proving effectiveness. EducationWeek.* https://www.edweek.org/teaching-learning/the-complicated-divisive-work-of-grading-teacher-preparation-programs/2021/06.

Williams, J. A., Hart, L. C., & Algozzine, B. (2019). Perception vs. reality: edTPA perceptions and performance for teacher candidates of color and White candidates. *Teaching and Teacher Education*, *83*, 120-133.

Williamson, J. L., Backam, C., Guy, M., Kay, P., & Turley, J. (1984). *Emergency teacher certification*. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education. https://files.eric.ed.gov/fulltext/ED241485.pdf.

Wilson, S. (1958, March). *Crisis in education*. *Life Magazine*.

Wineburg, M. S. (2006). Evidence in teacher preparation: Establishing a framework for accountability. *Journal of Teacher Education*, *57*(1), 51-64.

Wise, A. (2005). Establishing teaching as a profession: The essential role of professional accreditation. *Journal of Teacher Education*, *56*(4), 318-331. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1021.9225&rep=rep1&type=pdf.

Wise, A. E., & Leinbbrand, J. (1996). Profession-based accreditation: A foundation for high-quality teaching. *Phi Delta Kappan*, *78*(3), 202.

Wise, A. E., & Verstegen, D. A. (2000). Arthur E. Wise and the promise of equal educational opportunity. *Journal of Education Finance*, *25*(4), 583-596.

Wisniewski, R. (1981). Quality in teacher education: A reply to Alan Tom. *Journal of Teacher Education*, *32*(6), 53-55.

Wojcikiewicz, S. K. & Darling-Hammond, L. (2020). *Restarting and reinventing school: Learning in the time of COVID and beyond*. Priority 9: Prepare Educators for Reinventing Schools. https://restart-reinvent.learningpolicyinstitute.org/sites/default/files/product-files/Restart_Reinvent_Schools_COVID_Priority9_Educator_Preparation.pdf.

Worrell, F., Brabeck, M., Dwyer, C., Geisinger, K., Marx, R., Noell, G., and Pianta, R. (2014). *Assessing and evaluating teacher preparation programs. American Psychological Association*. https://www.apa.org/ed/schools/teaching-learning/teacher-preparation-programs.pdf.

Yun, C., & DeMoss, K. (2020). *Sustainable strategies for funding teacher residencies: Lessons from California*. Learning Policy Institute. https://learningpolicyinstitute.org/product/strategies-funding-teacher-residencies-california-report.

Zeichner, K. (2013). Two visions of teaching and teacher education for the twenty-first century. In X. Zhu & K. Zeichner (Eds.), *Preparing teachers for the 21st century* (pp. 3-19). Springer.

Zeichner, K. (2016). *Independent teacher education programs: Apocryphal claims, illusory evidence*. National Education Policy Center. http://nepc.colorado.edu/publication/teacher-education.

Zeichner, K. (2019). Preparing teachers as democratic professionals. *Action in Teacher Education*, *42*, 38-48.

Zeichner, L., & Paige, L. (2007). *The current status and possible future for "traditional" college and university-based teacher education programs in the U.S. submission to the International Alliance of Leading Education Institutes*. University of Wisconsin–Madison. http://www.intalalliance.org.

AUTHOR BIOGRAPHIES

Steven K. Wojcikiewicz is senior researcher and a policy advisor at the Learning Policy Institute (LPI). He is a member of LPI's Educator Quality team. He is a co-author of the book *Preparing Teachers for Deeper Learning* and of several case studies of educator preparation programs that are part of that project. His focus is on initiatives related to educator preparation research, practice, and policy, including the Educator Preparation Laboratory (EdPrepLab), an initiative of LPI and the Bank Street Graduate School of Education focused on teacher and leader preparation for deeper learning and equity. Wojcikiewicz has two decades of experience in varied roles across the field of education. He started his career teaching high school social studies and band in Tulsa, Oklahoma, and has also worked as an experiential educator aboard sailing vessels around the country. After earning his Ph.D., he served as an assistant and an associate professor of teacher education at Western Oregon University, followed by positions as the assistant director of educational issues at the American Federation of Teachers in Washington, DC, and as the vice president of policy for Deans for Impact in Austin, Texas. Prior to joining LPI, he was the director of the Pacific Alliance for Catholic Education and a faculty member in the School of Education at the University of Portland in Oregon. Wojcikiewicz has a Ph.D. in educational psychology and educational technology from Michigan State University, an M.A. in teaching from the University of Portland, and a B.A. in history and economics from the University of Notre Dame.

Susan Kemper Patrick is a senior researcher at the Learning Policy Institute (LPI). She is a member of LPI's Educator Quality team. After working in schools as a middle school teacher and a volunteer coordinator, her research now examines inequities in the learning opportunities available to both students and their teachers. At LPI, Patrick leads a collaborative research project with the California Commission on Teacher Credentialing. This project focuses on understanding the experiences of California teachers and administrators during their preparation and induction experiences. Prior to joining LPI, she served as a postdoctoral researcher for the Tennessee Education Research Alliance (TERA), a research–practice partnership between the Tennessee Department of Education and Vanderbilt University. At TERA, Patrick led research projects exploring the landscape of professional learning opportunities for educators across the state and examining the effects of COVID-19 on Tennessee students and teachers. Patrick received a Ph.D. in educational leadership and policy from Vanderbilt University, an M.Ed. in learning, diversity, and urban studies from Vanderbilt University, and a B.A. in African and African-American studies from Duke University.

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