

5

ASSESSING STATE AND FEDERAL POLICIES TO EVALUATE THE QUALITY OF TEACHER PREPARATION PROGRAMS

Ken Zeichner¹

States like Louisiana are leading the way in building the longitudinal data systems that enable states to track and compare the impact of new teachers from teacher preparation programs on student achievement over a number of years . . . Louisiana is using that information to identify effective and ineffective programs for the first time—and university-based teacher education programs are using the outcomes data to revamp and strengthen their programs . . . Louisiana is the only state in the nation that tracks the effectiveness of its teacher preparation programs. Every state should be doing the same . . . It's a simple but obvious idea—colleges of education and district officials ought to know which teacher preparation programs are effective and which need fixing. (Duncan, 2009, p. 5)

In this chapter, I examine the warrants for various existing and proposed state and federal policies related to accountability in preservice teacher education programs in the U.S. Given that it is very clear that currently little or no empirical evidence exists that supports the efficacy of particular accountability policies and processes used in state program approval and national accreditation (Wilson, Floden, & Ferrini-Mundy, 2001; Wilson & Youngs, 2005; National Research Council, 2010), I will examine the warrant for specific policies and practices based on a number of other criteria. These include how preservice preparation programs are assessed in other professions, the state of our current methods for assessing teachers' knowledge and teaching skills, the costs and projected benefits associated with particular practices and what are reasonable ways to hold teacher preparation programs accountable for their work. With regard to projected benefits, I will give attention to the likely ability of an accountability practice both to illuminate the quality of teacher education programs and contribute to improving programs.

Although the analysis will discuss a number of different policies and practices,² I will give particular attention to two accountability practices that are under intense discussion in the current policy context: the development of a rigorous teacher performance assessment that would be used for completion of a preservice program and initial teacher licensing and, as noted in the quote above, the evaluation of the quality of a teacher education program based on a value-added analysis of the standardized test scores of elementary and secondary school pupils taught by graduates of specific programs. The latter practice, referred to as the "positive impact mandate" (Hamel & Metz, 2005, p. 158), has received extensive and largely uncritical coverage in the national print and broadcast media (e.g., Abramson, 2010; Glenn, 2010). Both of these practices have been endorsed by the current federal education department (e.g., Duncan, 2010) and it is important that they receive a careful examination while their implementation is still limited. Since the Secretary's talk in October 2009, several other states (e.g., FL, TN, TX, DE) have made moves to implement the positive impact mandate as a form of teacher education program accountability. Federal funding streams such as Race to the Top encouraged more states to join the effort. Given that a new and widely disseminated report was released on accountability in teacher education while this chapter was being written (Crowe, 2010), I will specifically comment on the recommendations made in that report.

Government Policies Related to the Quality of Teacher Preparation Programs

In the last 30 years, both state education departments and the federal government have enacted various policies aimed at assessing the quality of teacher preparation programs that prepare teachers for initial certification. Until the reauthorization of Title II of the Higher Education Act in 1998, and the Elementary and Secondary Education Act in 2001, and despite efforts by the federal government to encourage particular forms of teacher education by providing competitive funds for the use of certain practices (e.g., Clarke, 1969; Earley, 2000a), it was mostly the states and not the federal government that formulated policies and regulations regarding accountability in teacher education (Bales 2006; Imig & Imig, 2008).

Prior to the 1980s, states emphasized an input driven model of program evaluation and approval that judged the degree to which teacher education programs contained the components that were required in a particular state either in terms of opportunities for teacher candidates to study particular topics, the presence of required courses (e.g., a course in teaching reading), or the required number of credit hours devoted to particular topics (e.g., nine credit hours in literacy teaching). These requirements also typically included a minimum number of hours that had to be spent in clinical experiences prior to a full-time student teaching or internship experience and a required minimum number of hours for the full-time teaching experience. For many years, states have licensed individual

teachers based on their completion of a state-approved teacher education program (Cronin, 1983; Darling-Hammond et al., 2005).³

During my first encounters with the state program approval process in Wisconsin in the 1970s, the process, which occurred every five years, consisted of state education department staff and several K-12 educators auditing syllabi of required courses in teacher education programs to see that the required topics were listed, and checking that the required number of minimum credit hours or time periods for different program components like student teaching or academic minors were in existence. During this period, most states left it up to teacher education institutions to make judgments about the quality of candidates' teaching. The effectiveness of candidates in the classroom was usually judged solely by the observation-based assessments made by college and university-based and school-based supervisors and school-based mentors and administrators, a practice which has been shown to be highly unreliable for measuring teacher effectiveness (Porter, Youngs, & Odden, 2001; Chung Wei & Pechione, 2010; Wilson, 2009).

Unlike some other countries where there are national standards related to licensure and program quality (Wang et al., 2003), individual states in the U.S. set their own policies. There is some degree of overlap in state requirements however, as a result of voluntary national accreditation requirements (www.nrcare.org or www.teac.org),⁴ and consortia of states that have agreed on the use of a number of common standards with regard to teaching and teacher education programs (www.nasdtdec.org, www.ccsso.org). Despite these areas of overlap, individual states' ability to set their own policies with regard to accountability for teacher education programs has resulted in accountability and licensing requirements that have been called "haphazard" in the most recent report on teacher quality by the U.S. Secretary of Education (U.S. Department of Education, 2009).

Beginning in the 1970s in several southern states, and then moving to other areas of the country in the 1980s, states began to require a variety of tests to enter and complete teacher preparation programs and generally became more prescriptive about the teacher education curriculum (Cronin, 1983; Prestine, 1989).⁵ These tests include basic skills tests (currently 27 states),⁶ tests of professional knowledge and pedagogy (currently 28 states), and tests of core academic subject-matter content (currently 37 states) and other subject-matter content (currently 32 states) (NASDTEC, 2010). Currently there are about 1,100 different tests used for initial teacher licensure throughout the U.S., with each state choosing its own tests and setting its own passing scores (Crowe, 2010). According to a report by the Education Trust (Brennan, 1999), most state subject-matter licensure tests are viewed as too easy and not relevant to ensuring that teachers have the academic skills that they need to be successful in raising student achievement. They also have very little predictive validity with regard to future success in teaching (Goldhaber, 2010; Wilson & Youngs, 2005). Despite these and other criticisms about the value of the current teacher tests (Berry, 2010), pass rates on teacher licensure tests are used as a component of the accountability system in 32 states.

For example, in New York, 80% of program completers from individual programs must pass the required tests for the programs to avoid sanctions by the state (NASDTEC, 2010).

During the 1970s, states began to introduce performance assessment in teacher education, and competency-based or performance-based teacher education (C/PBTE) was required or there were plans to require it in over 20 states for program approval and the initial licensing of teachers (Gage & Winne, 1975). At one point, all National Teacher Corps projects were required to use performance-based assessment (Houson & Howsam, 1972) and the American Association of Colleges for Teacher Education encouraged teacher education programs to become competency-based. AACTE provided a wealth of resources to help them do so in specific aspects of their programs such as multicultural education (e.g., AACTE, 1974). C/PBTE⁷ was advocated as an alternative to making teacher education programs accountable according to whether they contained all the state required coursework and fieldwork. Uncoupling courses and credits from state licensure requirements with C/PBTE was supposed to enable programs to innovate and to develop different approaches, and was a key factor in the movement toward alternative routes to teaching (Sykes & Dibiener, 2009).

For a variety of reasons, including the cost of implementation and the lack of solid research supporting the connections between teacher competencies and student learning (e.g., Heath & Nielson, 1974), C/PBTE temporarily disappeared from U.S. teacher education with the exception of a few states like Florida and Georgia and programs like Alverno College in Milwaukee (Zeichner, 2005). Around 2000, C/PBTE once again gained momentum in teacher education accountability communities in the U.S. with the adoption of performance-based assessment by NCATE and the implementation by some states of performance standards for initial teacher licensing and program approval (Valli & Rennett-Artev, 2002). State teaching standards in 16 states in this current incarnation of C/PBTE are based in part on the standards developed by the Interstate New Teacher Assessment and Support Consortium (INTASC) that is a part of the Council of Chief State School Officers (www.ccsso.org).

For example, while I was working in 2004 in Wisconsin, state program approval shifted from a system that focused only on program inputs (e.g., Are the required topics and credits in the teacher education curriculum?) to an accountability system that emphasizes performance-based assessment of teacher candidates. Some states like Wisconsin examine the quality of the performance assessment systems in teacher education institutions for program approval, while other states like Washington also want to see evidence in candidate performance assessments that teacher candidates have achieved a certain level of competency on the state teaching standards. Currently, approximately 19 states require a performance assessment of teaching for initial licensure (NASDTEC, 2010).⁸

There have been various responses by teacher educators to the shift toward performance-based assessment as a part of initial teacher licensure and state program

approval. On the one hand, there is a concern that performance assessment negatively affects the ability of teacher educators to engage in the practices that they think are needed to educate beginning teachers well by diverting the attention of teacher educators and the limited resources of their institutions to activities they perceive as not related to their core mission (Berlak, 2010; Kornfeld et al., 2007; Rennett-Arvey, 2008). On the other hand is the argument that teacher performance assessment data (unlike value-added assessment data) potentially provide teacher educators with useful information that they can use in improving their programs (Peck, Gallucci, & Sloan, 2010) and serves as a form of learning for teacher candidates (Darling-Hammond, in press; Diez & Haas, 1997; Chung Wei & Pechione, 2010). Mostly, however, C/PBTE has not been fully implemented in many teacher education institutions despite state requirements, because of the costs and other issues associated with a genuine implementation of the idea (Zeichner, 2005). State departments of education have experienced cuts in their budgets and staff over the years and do not have the capacity in many cases to monitor and enforce a genuine performance-based system (Darling-Hammond, 2005).

What is a Reasonable Approach to Accountability in U.S. Teacher Education?

Teacher Licensure Exams and System Coherence

Given the lack of empirical evidence related to particular accountability policies and processes in teacher education, one way to begin to formulate a position on an accountability system for teacher education is to look at how the quality of other professional schools are assessed. When one examines how other professions evaluate the readiness of individual candidates to practice and assess the quality of the preservice programs that prepare them, it is clear that there is much more uniformity across the country with regard to how other professionals are licensed. Crowe (2010) and Neville, Sherman, and Cohen (2005) discuss licensing and program approval requirements in a number of professions such as medicine, law, accountancy, nursing, and engineering, and all of these other professional schools have either a national licensing exam or a state exam with a national component before candidates are allowed to practice. Some professional schools also use performance assessments, and the structured observation and evaluation of clinical practice.

Crowe (2010) calls for both a major overhaul of teacher licensing exams and greater uniformity across the nation in teacher standards, policies, and program approval processes. Both of these recommendations are reasonable ones given the practices in other professions. As Berry (2010) points out, however, merely raising the cut scores on current teacher licensing exams, as some have suggested, will not necessarily lead to improvements. For example, research by Goldhaber (2007) showed that raising the cut scores on the North Carolina licensing exam up to the level used in Connecticut would eliminate teachers who have proven that they

can produce higher student achievement on standardized tests. Other analyses have shown the disproportionate failure rates on some exams by minority teacher candidates (Citomer, Laham, & Ziomek, 1999; Villegas & Davis, 2008). Crowe's (2010) recommendation to engage in a major overhaul of teacher licensing exams and to make them more uniform in content and cut scores across the nation seems warranted as a general recommendation.

We have to keep in mind, though, the growing empirical evidence related to the importance of building a more ethnically and racially diverse teaching force in terms of its positive impact on student learning, particularly learning for students of color (Villegas & Davis, 2008). We also have to remember that the purpose of initial licensure tests is to separate those candidates who are minimally competent from those who are not. The National Research Council (NRC) report on teacher testing concluded that "a set of well designed tests cannot measure all of the prerequisites of competent beginning teaching" (Mitchell et al., 2001, p. 165). This group concludes that multiple measures of beginning teacher effectiveness are needed and that decisions about licensing should not be made on licensure tests alone. So, while Crowe's (2010) recommendation that we need to apply higher standards in a new set of teacher licensure tests that are more uniform across the nation makes sense up to a point, there are real dangers in raising the cut scores too high. "Setting substantially higher passing scores on licensure tests is likely to reduce the diversity of the teaching applicant pool" (Mitchell et al., 2001, p. 167) and, as Goldhaber's (2007) research noted above concluded, keep potentially effective teachers out of the classroom.

Assessments of Teacher Effectiveness in the Classroom

Throughout the history of formal American teacher education programs, teacher candidates have had to demonstrate their competence in a classroom as part of program completion (Fraser, 2007). Throughout much of this history, these judgments were made by school-based or college- and university-based supervisors and mentors based on brief classroom observations. The unreliability of these assessment measures of teaching quality has been demonstrated in the literature (e.g., Chung Wei & Pechione, 2010; Porter, Youngs, & Odden, 2001; Wilson, 2009). Crowe's (2010) recommendation that accountability systems in teacher education "should include a measure of teacher effectiveness that reports the extent to which program graduates help their K-12 students to learn" (p. 12) is a reasonable one that can be approached in a number of different ways. One way to obtain assessments of teachers' ability to promote student learning is to strengthen the weak systems of student teacher assessment that exist in many clinical preparation experiences across the nation.⁹

When I began my career as a university teacher educator in the 1970s, efforts were made to infuse some of the more structured classroom observation instruments into student teacher/intern supervision (e.g., Simon & Boyer, 1974) and to

build a body of research and sound practices in supervising clinical experiences in teacher education (e.g., Goldhammer, 1969). The goal in these efforts was to raise the quality of the mentoring and assessment of teacher candidates during their clinical experiences with a focus on students and their learning.

Today there is wide consensus that the quality of supervision and assessment in clinical experiences in preservice teacher education is highly uneven (AACTE, 2010). Currently, there are a number of efforts like the Working with Teachers to Develop Fair and Reliable Measures of Effective Teaching project funded by the Bill and Melinda Gates Foundation (<http://www.gatesfoundation.org/highschools/Documents/mef-framing-paper.pdf>) to develop higher quality classroom observation-based assessments of the quality of teaching. Other notable efforts to make direct assessment of actual teaching in the classroom a central feature of educational accountability include the Classroom Assessment Scoring System or CLASS (Pianta & Hamre, 2009) and an observational framework based on the ETS Praxis III performance assessment (Danielson, 1996). Improving the quality and consistency of supervisor and mentor teacher assessments of teacher candidates is an important part of a strategy to measure the effectiveness of teachers in the classroom before giving them an initial teaching license or allowing them to serve as teachers of record. Using observational frameworks designed for research purposes in classroom-based assessments in preservice clinical experiences will require some adaptations, but we know from adaptations of parts of systematic observation instruments during the era of teacher effectiveness research in the 1970s that this is a doable task (e.g., Acheson & Gall, 1980).

It is not very common for either college- and university-based field supervisors or school-based mentors to be required to receive preparation for their work as supervisors of teacher candidates. In fact, the P-12 teachers who provide the bulk of mentoring and assessment of teacher candidates in most programs rarely receive the compensation and support that are justified by the important role that they play and the time that they spend on this work in many teacher education programs (Zeichner, 2006).¹⁰ Improving the consistency and the quality of field supervision for teacher candidates should be a priority in efforts to raise the quality of how we assess the quality of teacher candidates' teaching.

Another strategy for including a measure of teaching effectiveness that includes the ability to be successful in achieving student learning as a part of initial licensing is to utilize a high quality teacher performance assessment. Berry (2010) and Darling-Hammond (2009, in press) lay out a convincing case for the use of such an assessment based on research evidence from the beginning teacher assessments in Connecticut, and the National Board assessments (also see Darling-Hammond & Chung Wei, 2009). During the last several years, researchers at Stanford have led the development of a rigorous teacher performance assessment (Performance Assessment for California Teachers or PACT) that is used in over 30 California teacher education institutions. Despite some concerns about the assessment and about the lack of funding to support its implementation (Bertak, 2010), this assessment has been shown in some cases to

be able to predict teacher effectiveness according to student learning and to support teacher learning and teacher education program improvement (e.g., Chung Wei & Pechione, 2010; Newton, Walker, & Darling-Hammond, 2010; Pechione & Chung, 2006; Peck, Gallucci, & Sloan, 2010).

AACTE and the CCSSO are currently supporting a project involving 20 states that is developing a nationally available performance assessment based on the PACT that meets high standards of reliability and validity and that can be used in a variety of states for candidates to demonstrate their mastery of state teaching standards.¹¹ This assessment combines embedded signature assessments in individual teacher education programs with a capstone teaching event used across all institutions (Dietz, 2010). The capstone teaching event which is usually done in the final student teaching or internship experience engages candidates in documenting their practice in relation to academic language, planning, teaching, assessing, and reflecting according to a set of guiding questions and structures. The responses of the candidates are then evaluated by trained scorers according to a set of carefully designed and field-tested rubrics. Extensive reliability and validity studies have been, and continue to be, carried out on the PACT (Chung Wei & Pechione, 2010; Pechione & Chung, 2006) and on the new nationally available performance assessment that is based on it.¹²

Recently, I had my first direct experience with the PACT in the elementary and secondary teacher education programs at the University of Washington-Seattle that I currently direct. The implementation of this assessment is understandably a more complicated and expensive enterprise than what currently exists in most programs, with the need for scorer training, building opportunities to learn into the teacher education curriculum, coordinating the assessment with the placement schools and so on (a good assessment requires resources), and the kind of data about our candidates' teaching that emerged from this assessment on the performance of our teacher candidates was invaluable. For example, in our secondary program, we devoted several program meetings to discussions of various forms of the teacher performance assessment data (including artifacts from the assessment) that involved both university- and school-based teacher educators. These discussions led to revisions in the program curriculum for the next cohort of candidates. For example, a number of our secondary teacher candidates scored low on the academic language component of the assessment and some revisions were made in the curriculum and candidate assignments to address these areas of weakness.

There are other performance-based assessments of teaching besides PACT, such as the ETS Praxis III assessment¹³ and the protocols for evaluating candidate work samples developed originally at Western Oregon University (McConnery, Schalock, & Schalock, 1998).¹⁴ The goal of the CCSSO and AACTE project is to develop a more uniform approach to performance-based assessment in teacher education than the current approach of allowing each state to choose what assessments they will use. Even if it turns out that there is more than one performance assessment

used by states, there should be a requirement that all of the assessments used to assess the quality of teaching effectiveness of program completers meet a set of common standards with regard to their psychometric quality.

The use of portfolios for the assessment of the quality of teacher candidates' teaching is widespread in U.S. teacher education programs (Delandshere & Petrosky, 2010), but most of the portfolios that are used are relatively unstructured compared to the PACT and the Teacher Work Sample Methodology developed at Oregon State and do not have the psychometric quality to be used effectively as a summative assessment tool (e.g., Chung Wei & Pechone, 2010; Wilkerson & Lang, 2003).

Finally, another way to assess the teaching effectiveness of teacher candidates after they complete their preparation programs, and to supposedly judge the quality of these programs, is to use value-added analysis (VAA) to link growth in standardized test scores of pupils to the programs from which teachers graduated and then to rank teacher education programs in each state according to the alleged contribution of their graduates to student learning. Currently, as pointed out earlier, the national media have been obsessed with this strategy (e.g., Abramson, 2010; Glenn, 2010; Honowar, 2007; Kelderman, 2010) and the Secretary of Education, as illustrated in the opening quote of this chapter, travels the country promoting the idea. Louisiana is continually identified as the model for other states to follow in this area (Noell & Burns, 2006) along with Florida that has already begun ranking teacher education programs according to the value-added test scores of pupils taught by graduates from the different programs in the state (Glenn, 2010).

In the last few years, there has been much debate about the wisdom of using VAA to tie growth in students' standardized test scores to specific teachers and teacher education programs. For example, researchers have shown that using value-added student achievement scores to measure teaching effectiveness requires at least three years of data (McCaffrey, Lockwood, Mariano, & Setodji, 2005). Other researchers have questioned the assumptions on which VAA models are based and warn about their careful use (Rothstein, 2010). Some researchers have also raised questions about the tests that are used as measures of student learning (Darling-Hammond & Chung Wei, 2009). Finally, because the results one gets in VAA vary according to the decisions researchers make about how to handle the data, there is wide consensus that VAA should not serve as the sole basis for making decisions about teachers (Braun, 2005).

The National Research Council (2010) report on teacher education in the U.S. examined the relevance of VAA for evaluating teaching and teacher education programs. The report acknowledges some of the concerns that have been raised about this method, including:

That value-added methods do not adequately disentangle the role of individual teachers or their characteristics from other factors that influence student achievement...there are concerns about measures of student outcomes and accurate measurement of teacher education attri-

butes...Another concern is that student achievement tests developed in the context of high stakes accountability goals may provide a distorted understanding of the factors that influence student achievement.

(2010, p. 29)

After acknowledging these and other concerns, the report concludes:

As with any research design, value-added models may provide convincing evidence or limited insights depending on how well the model fits the research question, and how well it is implemented. Value-added models may provide valuable information about effective teacher preparation, but not definitive conclusions and are best considered together with other evidence from a variety of other perspectives.

(2010, p. 29)

Very few of those who have advocated the use of VAA to evaluate the quality of teacher education programs have advocated their use alone as a measure of effectiveness, including Crowe (2010). Levin (1980) advocated the use of a cost-utility analysis for evaluating the wisdom of using particular components in both teacher licensing and teacher education program accountability systems. When one follows Levin's advice, the question arises as to whether it is worth the time and expense to gather value-added data for the purposes of program accountability given the lack of consensus about the wisdom and/or the reasonableness of doing so, and the questionable quality of the information it provides. Couldn't a rigorous and consistent system of teacher education accountability be created that pays attention to teachers' abilities to teach students effectively, utilizing all of the other ways to assess teacher education program quality discussed above and implementing them in a more consistent and rigorous manner than is currently the case?

There are several arguments that should be raised and at least discussed related to the wisdom of VAA as a component of teacher education accountability. None of the popular press articles in the *Chronicle of Higher Education*, or *Education Week* (e.g., Glenn, 2010; Honowar, 2007), the piece on National Public Radio (Abramson, 2010), or articles in local newspapers (Matus, 2009), discusses the concerns that scholars have raised about the methodology in any detail or the fact that scholars disagree about whether and/or how it should be used. They also do not discuss the reasonableness of the approach as a way to evaluate professional schools. In one of the recent statements about the recent Center for American Progress report, it is implicitly asserted that, unless a state is using VAA to evaluate and rank its teacher education programs, it does not "actively hold teacher preparation programs accountable for the effectiveness of the teachers they produce" (Center for American Progress, 2010, p. 1).

The first question that should be asked about VAA is whether there are any other professional schools evaluated in this way on the basis of student/client/patient

outcomes after the candidates have completed their preparation programs. Using VAA as a required component of a teacher education accountability system would be analogous to evaluating medical schools according to how well graduates of particular medical schools were able to help particular patients get well, or how many cases graduates of particular law schools won or lost, or how many clients of accountants from particular business school programs were audited by the IRS, and so on. While Crowe (2010) and other critics of teacher education are eager to draw on the accountability systems for other professional schools to advocate for more uniformity in practice, no one has mentioned the fact that there is not a single profession where preparation programs are held accountable in the accreditation process for student/client/patient outcomes beyond the point of graduation. As has been pointed out in discussions of accountability in other professions (e.g., Neville, Sherman, & Cohen, 2005), uniform licensing exams that sometimes include a performance assessment component are standard practice for assessing candidates' readiness to practice and for assessing the quality of medical preparation. To require that teacher education programs be held to a standard of accountability that no other professional school is held to require is a practice for which a justification has not been provided.

A second question that should be raised about the wisdom of using VAA for teacher education program accountability is the usefulness of the data that it provides about the elements of teacher education programs that are related to positive results. Despite the fact that teacher education program administrators are often quoted in popular press articles promoting the use of VAA results in stimulating program improvement (e.g., Matus, 2009), the fact is that a ranking of institutions using VAA provides very little, if any, information about the particular features of programs that are linked to the outcomes. Although there are a few examples of research projects that use VAA in combination with other methods to illuminate the particular features of teacher education programs that are linked to positive and negative outcomes (Boyd et al., 2008), the kind of analyses that have been produced to date in Louisiana and Florida are not sophisticated enough to produce data that illuminates the particular aspects of preparation programs or teachers' practices that would be useful to program improvement. On the other hand, as has been discussed earlier, there are examples of how specific information about candidates' teaching from a rigorous teacher performance assessment can be used to support program renewal and improvement (e.g., Peck, Gallucci, & Sloan, 2010).

The comparative costs of implementing a VAA-driven accountability system and of developing high quality teacher performance assessments to be used for program completion, together with the value of the data produced for stimulating program renewal and improvement, suggests that strengthening classroom observation-based assessment and developing high quality performance assessments are much more worthy activities to undertake than investing in VAA to assess the teaching effectiveness of teacher education program graduates.

Although Harris and McCaffrey (2010) argue that, given the current system of standardized testing in the U.S., the cost of creating value-added (VA) measures is

quite low, they also acknowledge that the costs associated with calculating the VA measures are only a part of what is needed to adopt a VAA approach to evaluate teaching. In addition to calculating the actual measures, they argue that educators need to be trained in how to use VA measures and to understand their limitations and investments need to be made in overcoming some of the technical problems that have limited the usefulness of VA data to date.

Discussion

In this chapter, I have briefly discussed a number of existing and proposed policies and processes for strengthening the system of teacher education accountability in the U.S. While I have supported certain specific recommendations and general principles advocated in the Center for American Progress's recently released report on teacher education accountability (Crowe, 2010)—such as engaging in a major overhaul of teacher testing, creating greater uniformity throughout the nation in policies and practices, using high quality assessments of candidates' teaching as part of initial licensure and program approval, and holding all teacher education programs to the same accountability standards—I have argued against the ranking of teacher education programs based solely on the VAA of pupil test scores of their graduates as a reasonable, cost-effective and useful way to assess candidates' teaching effectiveness and to evaluate the quality of teacher preparation programs.

No other professional school is held accountable for student/patient/client outcomes after program completion in this way, and the data that are produced by VAA, lacking information about the specifics of teaching and the contexts in which it takes place, do not contribute to the improvement of teaching or teacher education programs. It would be a wiser strategy to invest in improving both classroom observation-based assessments in clinical experiences and to develop a high quality teacher performance assessment to be administered at the completion of a pre-service program. Both of these types of assessments would provide much more specific information about the ability of teacher candidates to effectively produce student learning and, although they are expensive, they will have a much greater impact on improving the quality of teacher education in the U.S. than a VAA approach.

How to Achieve Greater National Uniformity

There are different ways in which we can move toward greater uniformity in initial teacher licensure and program accountability throughout the U.S. Some (e.g., Darling-Hammond & Baratz-Snowden, 2005) have argued for mandatory national accreditation of teacher education programs pointing to mandatory national accreditation of preparation programs in other professions. It is also potentially possible to bring about greater consistency in state requirements and policies by voluntary cooperation among state education departments and professional standards boards. Given the criticisms that have been leveled at the bureaucratic nature

of national accreditation of teacher education in the past (e.g., Johnson et al., 2005)¹⁵ and the fact that there is little existing empirical evidence as to the value of national program accreditation (National Research Council, 2010; Wilson & Youngs, 2005), the voluntary cooperation approach advocated by Crowe (2010) seems like a reasonable approach at least for now.

One can question, though, the likelihood of states voluntarily agreeing to adopt the same licensure standards. In the long run, some form of mandatory national program accreditation will probably be the only way to achieve a more uniform national accountability system for teacher education in the U.S. Priority should be given to the National Research Council's (2010) recent recommendation to undertake an independent evaluation of program accreditation in teacher education. This evaluation could lead to a revision of the current system and then to a requirement that all teacher education programs be nationally accredited. If the redesigned system is streamlined and made more manageable and cost effective, is meaningful in the sense of getting at actual program quality,¹⁶ and helps contribute to the improvement of programs, it will likely be positively received by teacher education institutions.

It is very interesting how commentators like Crowe (2010) draw on other professions in a very selective way. Although they draw on accountability in other professions as a reason for creating greater uniformity in teacher education accountability, they fail to point out that in most of the other professions that are used as illustrations the profession itself plays a significant role in setting and enforcing accountability standards. Whatever accountability system is developed for teacher education in the U.S. must include a significant role for the profession in setting and enforcing standards for teachers and teacher education programs along with greater national uniformity.

High Quality Teacher Education Accountability Is Expensive

Another issue that needs to be faced in the creation of higher quality standards for initial teacher licensure and program accountability is that a higher quality system will cost more than what is currently in place. The National Research Council report (2010) on teacher education in the U.S. and the most recent teacher quality report by the U.S. Secretary of Education (U.S. Department of Education, 2009) indicate that somewhere between 70% and 85% of new teachers today have been prepared by a college and university program of some kind. Given the consistent decline over a number of years in state support to the public universities where most college and university educated teachers are prepared (Lyall & Sell, 2006), as well as the continuing cuts in the budgets of the state education agencies that would implement a substantial part of a strengthened accountability system alone or in conjunction with national accreditation bodies (Darling-Hammond et al., 2005), the question of how a higher quality accountability system in U.S. teacher education will be funded is a serious issue that needs to be resolved.

One strategy used in the past to fund components of the accountability system in teacher education has been to shift the costs to prospective teachers. However,

with the widespread use of teacher testing throughout the country at various points in initial licensing and the sharp rise in college and university tuition to offset reductions in state funding, it has become quite expensive for teacher candidates to meet state requirements for initial licensure.¹⁷ The implementation of a high quality teacher performance assessment with good reliability and validity and strengthening classroom observation-based assessment during clinical experiences will also be very expensive. This past year, for example, to administer a version of the PACCT to about 130 elementary and secondary teacher candidates at the University of Washington–Seattle, we spent approximately \$35,500 for the training of scorers, and paying scorers for initial scoring and rescoring (about \$273 per candidate). These costs do not include the salary of a half-time staff person to coordinate the whole process and the substantial staff and faculty time that were spent in designing the infrastructure that was needed to support the assessment and to integrate it into the teacher education curriculum. The state of Washington now has a requirement for an evidence-based performance assessment in all its teacher education programs and discussions are taking place about how the costs of implementing this assessment will be paid. Shifting the costs to teacher candidates who are already paying higher tuition and fees for required basic skills and content exams is problematic given the negative effects this is likely to have on the goal of building a teaching force in the state that is more representative of the population in the state.

When one applies Levin's (1980) cost-benefit analysis approach to the problem of assessing the teaching effectiveness of teacher candidates as a component of teacher education program accountability, the most expensive option in the range of alternatives is the implementation of VAA to rank teacher education programs in each state. The money that would be spent in implementing these analyses in every state, and in training people to use them, could more wisely be spent on supporting directions for reform in teacher education that research shows makes a difference in producing high quality teacher education programs, such as: strengthening the clinical component of preparation and its connection to the rest of the program (e.g., Boyd et al., 2008), supporting the development of higher quality classroom observation and performance assessments, and supporting research and evaluation on teacher education including the needed evaluation of the accountability system in teacher education. The return on investment on these and other expenditures would be greater than that from investing in an expensive VAA accountability system that provides very little data about teaching that can be useful for improving teaching and teacher preparation programs.

Identify and Punish the "Culprits" vs Help Programs Become Better

Sykes and Dibner (2009), in their review of federal policy related to teaching in the U.S. over the last 50 years, make a distinction between sanctions-oriented policies that are designed to identify and punish the culprits (Earley, 2000a) and accountability that is designed to contribute to the improvement of teachers,

schools, and teacher education programs. They argue for policies that provide useful data that contribute to the improvement of teaching and teacher education. There is a certain cynicism among a number of critics of the current teacher education accountability system about the intentions and motives of teacher educators in colleges and universities and there are even accusations in some cases that teacher educators are trying to get away with something dishonestly. A statement by Crowe (2010) is a good example of this cynical attitude. Referring to the 1998 reauthorization of Title II of the Higher Education Act and the required state report cards on candidate pass rates on content exams, Crowe states:

Shortly after the report card statute was established, a significant number of institutions and state agencies joined with the teacher education professional associations—the American Association of Colleges for Teacher Education or AACTE as well as NCATE—to work out a way to beat the reporting system. The trick they devised was requiring teacher candidates to pass all required teacher tests before being allowed to graduate. This allowed programs to report 100 percent pass rates on the teacher tests.

(2010, p. 9)

This statement about “trickery” which criticizes how teacher education institutions responded to requirements reporting on tests that Crowe (2010) has concluded is essentially bankrupt is very interesting.¹⁸ This attitude of “somebody is trying to get away with something” raises a question about what the purpose of the tests are in the first place. Isn’t the goal of requiring candidates to pass licensing exams to ensure that those who receive initial teaching licenses have mastery of basic skills and subject areas in their certification areas at a certain level of competence? Hasn’t this goal been met if teacher preparation institutions do not recommend candidates for initial licensure if they fail to pass the tests?

It seems to me that developing a fair and rigorous system for monitoring the quality of teacher preparation that closes the weakest traditional and alternative programs and that contributes to the improvement of most programs may not be the real goal of some education school critics who advocate for the use of VAA in teacher education. As Diez (2010) argues, too much emphasis on *proving* that teacher education programs work or don’t work can stand in the way of *improving* them. What we should be seeking in an accountability system in teacher education is to get an in-depth and accurate reading of the quality of the teachers that programs are recommending to the state for initial certification and a system that contributes to ongoing improvement of preparation programs.

The idea of publicly ranking teacher education in a state according to teacher candidates’ performance on licensing exams is not a new idea, as can be seen from Table 5.1, which shows the ranking of teacher preparation institutions in Wisconsin in 1863 in the *Wisconsin Journal of Education*. Although it is a reasonable expectation to hold teacher education programs accountable for the performance of their graduates

TABLE 5.1 Summary of Averages

	Whole No. Examined	Int Arithmetic	Written Arithmetic	Algebra	El Sounds	Spelling	Analysis	Grammar	Composition	Reading	Geography	Physical Geography	Physiology	History	Theory and Practice	Pennmanship	Total
Racine High School	6	78.3	87.5	81.7	78.3	86.7	61.7	69.2	60	85	66.7	73.3	75.8	72.5	66.7	75	74.4
Lawrence University	7	75.5	86.4	73.5	68.5	86.4	75.5	77.1	71.4	78.5	56.4	62.9	73.5	65	72.1	77.1	73.3
Allen Grove Academy	13	77.7	76.2	75.5	78.8	80.4	66.5	67.7	52.3	83.1	73.5	68.1	66.2	62.3	62.3	79.6	71.3
Platteville Academy	11	76.8	78.2	64.5	76.3	86.3	60	70.9	62.7	83.2	55.9	63.2	55.4	48.2	65.9	79.5	68.5
Wisconsin Female College	6	72	78.3	67.5	48.3	69.2	73.3	65.8	65	78.3	63.3	67.5	65	63.3	75	62.5	67.6
Fond du Lac High School	4	72.5	72.5	83.7	71.2	81.2	51.2	68.7	63.7	76.2	57.5	50	65	41.2	68.7	80	66.9
Evansville Seminary	20	68.5	74	61	72.7	74	53.2	56.7	57	72.5	61	64.2	62	66.2	70.7	75.5	65.9
Milton Academy	15	76.3	77	58.7	63	76	43.7	59.3	69.3	82.3	52.3	55.3	55	58.3	65	78.7	64.7
Oshkosh High School	3	70	73.3	68.3	70	78.3	46.7	68.3	70	76.7	55	41.6	61.6	50	63.3	71.7	64.3

on licensing exams at the time of program completion, this kind of general ranking of institutions tells us very little about the quality of teaching of the graduates from these institutions. At least it provides some useful information, though, about the relative performance of candidates in the various subject areas covered in the exams that can be used as the basis for examining particular areas of the curriculum.

Table 5.2 shows a recent ranking of education schools in the state of Florida, published in the *St. Petersburg Times* in November 2009. The article begins with a sentence in large type that states that a large local university, the University of Southern Florida, "comes in ninth of the 10 schools when the Florida Comprehensive Assessment Test is used to measure graduates."

This ranking attempts to report on the quality of the ten Florida education schools based on the math and reading scores of students taught by the graduates of different programs. It determined what percentage of graduates from each program had 50% or more of their students make a year's worth of progress. In addition to the fact that this use of value-added analysis did not meet even minimum standards for the use of the method, such as using at least three years of test data, using VAA in combination with other measures of effectiveness, etc. (Berry, 2010; National Research Council, 2010), there is very little useful information provided in this ranking that can be used for program improvement. One could argue that the crude ranking of teacher preparation institutions in 1863 is a more reasonable and useful form of accountability that what was done in Florida in 2009.

The NRC examination warning about the dangers of oversimplification in publically ranking teacher education institutions based on licensure test scores can also be applied to the ranking of institutions by VAA scores alone.

The public reporting and accountability provisions of Title II may encourage erroneous conclusions about the quality of teacher preparation.

TABLE 5.2 Rating Teacher Preparation Programs

<i>University</i>	<i>Percentage of Teachers with 50% or More of Students Making Learning Gains</i>	<i>% "High Performing"*</i>
Florida A&M	80	7
Florida Atlantic	84	19
Florida Gulf Coast	77	14
Florida International	85	23
Florida State	81	20
University of Central Florida	83	20
University of Florida	84	18
University of North Florida	84	11
University of South Florida	76	15
University of West Florida	70	11

Note. * Based on FCAT learning gains that were particularly large.

Although the percentage of graduates who pass initial licensure tests provides an entry point for evaluating an institution's quality, simple comparisons among institutions based on their pass rates are difficult to interpret for many reasons... By themselves, passing scores on licensure tests do not provide adequate information on which to judge the quality of teacher education programs... The federal government should not use passing rates on initial licensure tests as the sole basis for comparing states and teacher education programs or for withholding funds, imposing other sanctions, or rewarding teacher education programs.

(Mitchell et al., 2001, pp. 170–171)

There are Real Problems and How Not to Fix Them

It is clear from both analyses initiated within the teacher education community (e.g., Wilson & Youngs, 2005), from critics of education schools such as Crowe (2010) and from impartial scientific panels convened by the National Research Council (Mitchell et al., 2001; NRC, 2010), that there are real problems with the teacher education accountability system in the U.S. that need to be addressed, including uneven standards for teachers and programs, different accountability rules for different kinds of programs, and the need to include a high quality measure of teaching effectiveness in both the initial licensing process and the assessment of the quality of teacher education programs. No one has argued that the current accountability system for teacher education programs is sufficient and does not need to be improved.

One approach that has become common in recent years is for vocal advocates of deregulation in teacher education and critics of education schools to proclaim themselves as "non-partisan" and issue their own evaluations of teacher education programs and reports on teacher education issues. There is no better example of this than the reports on teacher education programs that have been issued by the National Council on Teacher Quality (NCTQ) in the past few years.

Without any effort to submit either frameworks or "findings" to genuinely impartial peer review, the NCTQ proclaims:

Both program approval standards set by states and accreditation standards set by private organizations provide no indication of the quality of one institution's preparation relative to another... Unfortunately this leaves consumers, aspiring teachers and schools who hire teachers in the dark... As that every child has an effective teacher, NCTQ is stepping into this vacuum to help consumers distinguish between good, bad and mediocre education schools. We do so by setting the bar higher than it has been set traditionally.

(NCTQ, 2010)¹⁹

This allegedly non-partisan body which is not recognized by the federal government or any professional association as an accrediting body has issued its own set of standards for defining a high quality teacher education program²⁰ and has begun to go from state to state in applying its frameworks and issuing reports on the quality of different teacher education programs and on teacher education programs nationally in particular subject areas such as reading and mathematics (Walsh, Glaser, & Wilcox, 2006). The most recent report on teacher education programs in specific states is focused on Texas (NCTQ, 2010). This group, just like other groups, has the right to make its arguments about what makes a good teacher education program. There are two fundamental problems, though, with the current strategy of the NCTQ.

First, although the standards used to evaluate teacher education programs are described as representing consensus thinking from an impartial group, as illustrated in the quote below, the members of the group who developed the standards include some of the most outspoken critics of education schools and advocates of the deregulation of K-12 and teacher education, such as Chester Finn, Michael Podgursky, Frederick Hess, Michael Feinberg, Kate Walsh, and Michelle Rhee (Fordham Foundation, 1999; Hess, 2001; Walsh, 2004). This is hardly a non-partisan group.²¹

The standards were developed over 5 years of study and are the result of contributions made by leading thinkers and practitioners from not just all over the nation but all over the world. To the extent that we can, we look to the practices of higher performing nations; where relevant the practices of other professions, and the best consensus thinking.

(NCTQ, 2010)

Second, for a group that has focused so much on so called "scientific approaches" to teaching reading and mathematics,²² it is ironic that they have not submitted their work to scholarly venues where it can undergo rigorous peer review and critique. Both the recent AERA and National Research Council investigations of research on U.S. teacher education underwent various levels of peer review prior to and after the release of the reports (Cochran-Smith & Zeichner, 2005; National Research Council, 2010). Why is the NCTQ so reluctant to have its judgments of teacher education programs undergo rigorous and impartial peer review in the most highly regarded journals?²³ Their strategy has been to go directly to the media with their so-called scientific reports. The media in turn print accounts about the conclusions of the NCTQ reports implying that they have undergone the usual scientific peer review. Sometimes, as in an article in the *Houston Chronicle* (Mellon, 2010) about the report on Texas, the media quote a few teacher educators who question the NCTQ methodology, but this is "balanced" by quotes from Texas superintendents who endorse the report. The NCTQ website includes a place for superintendents to indicate their support for its work.

This deceptive process of evaluating teacher education programs in an allegedly objective way is driven by a political agenda to deregulate teacher education rather than by any sense of scientific rigor. With regard to both the VAA bandwagon and the uncritical acceptance of the NCTQ reports, the media²⁴ have acted irresponsibly by not publicly discussing existing debates about issues like VAA or indicating whether or not a report has undergone genuine scientific peer review. To be fair, advocates of education schools have sometimes behaved in the same unscientific manner by issuing reports that are not subjected to rigorous peer review, and individuals from both the professionalization and deregulation camps have oversimplified and distorted the positions of their critics to some extent (Wilson & Tannir, 2008), and the media reports have sometimes included brief quotes from educators with different viewpoints.

Rising Above the Bickering

A strong public school system is an essential element of our democratic society, and given what we know about the importance of teachers to the quality of educational outcomes (National Academy of Education, 2009), preparing good teachers for everyone's children who attend our public schools is an extremely important activity that should be above partisan bickering. What our country needs is an accountability system in teacher education that is the result of open and reasoned discussion and debate of different positions on goals and the means to achieve them, and genuine peer review of research findings and policy recommendations. The cost-benefit framework suggested by Levin (1980), which calls for careful consideration of the social benefits and costs associated with particular elements of a teacher education accountability system, would be a useful way to structure this analysis.

The uncritical acceptance by the media of the pronouncements of *any* group on teacher education accountability interferes with the important goal of strengthening our teacher education accountability system. It is urgent for the U.S. Department of Education to commission the impartial evaluation of teacher education accountability called for in the National Research Council (2010) assessment of teacher education in the U.S. and for the Department to insist that the commonly acceptable standards for research including peer review be followed in the allocation of funds to support particular practices and policies in teacher education.

It has become very clear that public policies do not follow in any field in a linear way from research findings and that research can never dictate the specifics of particular policies in education or any other field (e.g., Krison, Harvey, & McCormick, 1998; Stevens, 2007). Advocating for more reasoned and careful examination of the findings of particular inquiries related to accountability in teacher education and the methodologies that were used to produce them does not suggest that these inquiries will be able to translate directly into specific policies. There are legitimate differences that need to be negotiated in views about fundamental aspects of teaching and teacher education related to the purposes of

public education, the role of teachers, how student learning can be measured, and so on that will never be able to be resolved through research alone, even if it is of high quality (Cochran-Smith & Zeichner, 2005). This open debate and discussion about the goals and processes of public education and teacher education are fundamental benefits of living in a democratic society, and we should insist that our government agencies and the media support rather than short-circuit this process.

Former Illinois state superintendent of education Joseph Cronin, in his analysis of the history of state regulation in teacher education in the U.S., warns policymakers about the dangers of supposedly simple solutions²⁵ to problems of teacher education accountability and of avoiding the kind of reasoned discussion and debate that is needed:

Most of all, legislators and study commission members should remember that any change not only may fail to solve the specific problem but may in fact create new problems not anticipated at present . . . Remember the immortal words of H.L. Mencken, those of you who would reform teacher education: For every complicated problem there is a simple solution and it is usually wrong.

(1983, p. 190)

Redesigning the accountability system for teacher education in the U.S. is a complex matter that requires all of us to rise above our own self-interest and to learn to work in more productive ways with those who hold positions different from our own.

Notes

- 1 I would like to thank the following people, in addition to the editors, for their helpful comments about earlier drafts: Michael Apple, Linda Darling-Hammond, Mary Diez, Penny Engel, Kerry Kretchmar, Katie Payne, Cap Peck, Sharon Robinson, Cathy Taylor, Sheila Valencia, and Pat Wasley.
- 2 These include program approval, testing of basic skills, content and professional knowledge, and various ways to assess the quality of teachers' teaching.
- 3 As Conant (1963) and Cronin (1983) have pointed out, the teaching profession through the National Commission on Teacher Education and Professional Standards (TEPS), which was an arm of the National Education Association, and the professional standards boards in some states has exerted varying degrees of influence on both teacher certification and teacher education program accountability. They have also pointed out that states have differed in the degree to which they have given higher education a role in determining and monitoring these processes.
- 4 Forty-three states have adopted or integrated criteria for assessing the quality of teacher preparation programs from voluntary national accreditation agencies (U.S. Department of Education, 2009).
- 5 There have always been some tests involved in getting an initial teaching license (e.g., Elsbree, 1939), but many of them were locally administered by school districts or county education officers.

- 6 These figures with regard to the numbers of states using particular kinds of certification tests come from the most recent report on teacher quality from the U.S. Secretary of Education (U.S. Department of Education, 2009).
- 7 The term "C/BTE" is being used in a general way here as it was back in the 1970s to describe a general approach to teacher education that focused on teacher candidates demonstrating mastery of a set of outcomes. In practice, programs ranged from those that focused on discrete bits of isolated aspects of teaching, while others focused on fewer and more integrated aspects of teaching based on a clear conceptual framework (Lisbon & Zeichner, 1991).
- 8 The data on the NASDTEC website on August 1 are from 2004.
- 9 See Wasley and McDiarmid (2004) for a discussion of a number of different ways to connect teacher education, teaching, and student learning.
- 10 The highly publicized Conant Report (1963) emphasized the improvement of the clinical component of teacher education as the most important thing that could be done to raise the quality of teacher education in the U.S. and singled out the lack of preparation and support for cooperating teachers as one of the weakest aspects of the system. For example, "cooperating teachers should have time freed to aid the student teachers; they should also have increased compensation in recognition of their added responsibility and talent" (p. 62).
- 11 <http://aacte.org/index.php?/Programs/Teacher-Performance-Assessment-Consortium-TPAC/teacher-performance-assessment-consortium.html>
- 12 Detailed information about the PACT can be found at <http://www.pactpa.org/main/hub.php?pageName=Home>, and information about the CCSSO and AACTE project to develop a nationally available performance assessment based on PACT can be found at <http://aacte.org/index.php?/Programs/Teacher-Performance-Assessment-Consortium-TPAC/teacher-performance-assessment-consortium.html>
- 13 See <http://www.ets.org/praxis/institutions/praxisiii/>
- 14 See <http://www.wou.edu/education/worksample/twsm/>
- 15 There have been recent efforts to streamline and focus national accreditation more on outcomes (www.ncaate.org, www.iteac.org).
- 16 NCATE is already involved in streamlining its system and strengthening its connection to P-12 student learning and its ability to support the continuous improvement of programs (Cibulka, 2009).
- 17 The current fees for the most widely used tests are: Praxis I (\$130), Praxis II (\$65–115), and a \$50 registration fee. See <http://www.ets.org/praxis/about/fees>
- 18 One could argue that requiring candidates to pass a content test prior to their final student teaching or internship experience (which is a common practice) is a more ethical stance to take, given the effects that lack of minimal content knowledge could have on pupils, and the investment of time and money that candidates need to make to do a full-time clinical experience. I am grateful to Mary Diez for pointing this out to me.
- 19 Retrieved from www.nctq.org/p/response/evaluation_faqs.jsp on August 7, 2010.
- 20 The NCTQ standards focus on admissions and exit requirements from programs and the preparation of teachers to teach reading and mathematics. They do not address the quality of performance of teacher candidates in classrooms (www.nctq.org).
- 21 On its website (nctq.org), its mission is stated as "to provide an alternative voice to existing organizations and build a case for a comprehensive reform agenda that would challenge the current structure and regulation of the profession." The problem is not that NCTQ has involved mostly supporters of a deregulation agenda and critics of education schools to advise them on the development of their standards. I am also not suggesting that some amount of deregulation in teacher education is a bad thing. The problem is that the NCTQ publicly describes its position as non-partisan, purely in service of consumers, and that it is not transparent about the political agenda that propels its work.

- 22 Assertions that are in conflict with the recent analyses of the National Research Council (2010).
- 23 See <http://aacte.org/index.php?/Traditional-Media/Resources/aacte-members-respond-to-nctq-research-efforts.html> for a series of letters from state professional teacher education associations and education school deans detailing some of the methodological and ethical concerns that exist about the NCTQ evaluations.
- 24 This includes some of the major national publications in education such as the *Chronicle of Higher Education* and *Education Week*.
- 25 Teacher warranties where teacher education programs guaranteed the quality of their graduates, and promised to remediate any deficiencies at no cost to school districts, is one of the allegedly simple solutions to teacher education program accountability that did not amount to much (Earley, 2000b).

References

- Abramson, L. (2010). Study tries to track LA teachers. January 4 edition of *All things considered*. Retrieved January 10, 2010 from www.npr.org
- Acheson, K. & Gall, M. (1980). *Techniques in the clinical supervision of teachers*. New York: Longman.
- American Association of Colleges for Teacher Education (1974). *Achieving the potential of performance-based teacher education*. Washington, DC: Author.
- American Association of Colleges for Teacher Education (2010, March). *The clinical preparation of teachers: A policy brief*. Washington, DC: Author.
- Bales, B. (2006). Teacher education policies in the United States: The accountability shift since 1980. *Teaching and Teacher Education* 22, 395–407.
- Berkat, A. (1010). Coming soon to your favorite credential program: National exit exams. *Rethinking schools* 25(1). Retrieved August 30, 2010 from www.rethinking-schools.org
- Berry, B. (2010, May). *Strengthening state teacher licensure standards to achieve teaching effectiveness*. Washington, DC: Partnership for Teacher Quality. Retrieved June 24, 2010 from www.aacte.org
- Boyd, D. et al. (2008). Surveying the landscape of teacher education in New York City: Constrained variation and the challenge of innovation. *Educational Evaluation and Policy Analysis* 30(4), 319–343.
- Braun, H. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service.
- Brennan, J. (1999). *Study finds teacher licensure tests to be mostly high school level: If this is all we expect teachers to know why send them to college?* Washington, DC: The Education Trust.
- Center for American Progress. (2010). Better teachers, better students: Event proposes ways to measure the effectiveness of teacher training programs. Retrieved August 3, 2010 from www.americanprogress.org
- Chung Wei, R. & Pechione, R. (2010). Assessment for learning in preservice teacher education. In M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality* (pp. 69–132). San Francisco: Jossey-Bass.
- Cibulka, J. (2009). The redesign of accreditation to inform simultaneous transformation of educator preparation and P-12 schools. *Quality Teaching* 18(2). Retrieved August 1, 2010 from www.wvate.org
- Clarke, S. (1969). The story of elementary teacher education models. *Journal of Teacher Education* 20(3), 283–293.
- Cochran-Smith, M. & Zeichner, K. (2005) (Eds.). *Studying teacher education*. New York: Routledge.
- Conant, J. (1963). *The education of American teachers*. New York: McGraw Hill.
- Cronin, J.M. (1983). State regulation of teacher preparation. In L. Shulman & G. Sykes (Eds.), *Handbook of teaching and policy* (pp. 171–191). New York: Longman.
- Crowe, E. (2010, July). *Measuring what matters: A stronger accountability model for teacher education*. Washington, DC: Center for American Progress.
- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association of Supervision and Curriculum Development.
- Darling-Hammond, L. (2009). Recognizing and enhancing teaching effectiveness. *International Journal of Educational and Psychological Assessment* 3, 1–24.
- Darling-Hammond, L. (in press). *Evaluating teacher effectiveness: How teacher performance assessments can measure and improve teaching*. Washington, DC: Center for American Progress.
- Darling-Hammond, L. & Baratz-Snowden, J. (Eds.). (2005) *A good teacher in every classroom*. San Francisco: Jossey-Bass.
- Darling-Hammond, L. & Chung Wei, R. (2009). Teacher preparation and teacher learning: The changing policy landscape. In G. Sykes, B. Schneider, & D. Plank (Eds.), *Handbook of education policy research* (pp. 613–636). New York: Routledge.
- Darling-Hammond, L., Pacheco, A., Mitchell, N., LePage, P., Hammermness, K. & Youngs, P. (2005). Implementing curriculum renewal in teacher education: Managing organizational and policy change. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world* (pp. 442–479). San Francisco: Jossey-Bass.
- Delandshere, G. & Petrosky, A. (2010). The use of portfolios in preservice teacher education. In M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality* (pp. 9–42). San Francisco: Jossey-Bass.
- Diez, M. (2010). It is complicated: Unpacking the flow of teacher education's impact on student learning. *Journal of Teacher Education* 61(5), 441–450.
- Dier, M. & Haas, J. (1997). No more piecemeal reform: Using performance assessment to rethink teacher education. *Action in Teacher Education* 19(2), 17–26.
- Duncan, A. (2009, October). *Teacher preparation: Reforming the uncertain profession*. Address given by Secretary of Education Arne Duncan at Teachers College, Columbia University.
- Earley, P. (2000a) Finding the culprit: Federal policy and teacher education. *Educational Policy* 14(1), 25–39.
- Earley, P. (2000b, February). *Guaranteeing the quality of future educators: Report on a survey of teacher warranty programs*. Washington, DC: American Association of Colleges for Teacher Education.
- Elsbire, W. (1939). *The American teacher*. New York: American Book Co.
- Forham Foundation (1999). *Better teachers and how to get more of them*. Dayton, OH: Author.
- Fraser, J. (2007). *Preparing America's teachers: A history*. New York: Teachers College Press.
- Gage, N. & Winne, P. (1975). Performance-based teacher education. In K. Ryan (Ed.), *Teacher education* (pp. 146–172). Chicago: University of Chicago Press.
- Gutormer, D., Laham, A.S. & Ziomek, R. (1999). *The academic quality of prospective teachers: The impact of admissions and licensure testing*. Princeton, NJ: Educational Testing Service.
- Glenn, D. (2010). Education schools are scrutinized for graduates' success as teachers. *Chronicle of Higher Education*. Retrieved on August 5, 2010 from <http://chronicle.com>
- Goldhaber, D. (2007). Everyone's doing it, but does it tell us about teacher effectiveness? *Journal of Human Resources* 52(4), 765–794.
- Goldhaber, D. (2010). Licensure tests: Their use and value for increasing teacher quality. In M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality* (pp. 133–162). San Francisco: Jossey-Bass.

- Goldhammer, R. (1969). *Clinical supervision: Special methods for the supervision of teachers*. New York: Holt, Rinehart & Winston.
- Hamel, F. & Metz, C. (2005). Reforming accountability: A preservice program wrestles with mandated reform. *Journal of Teacher Education* 56(2), 157-167.
- Harris, D. & McCaffrey, D. (2010). Value added: Assessing teachers' contributions to student achievement. In M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality* (pp. 251-283). San Francisco: Jossey-Bass.
- Heath, R. W. & Nielson, M. (1974). The research base for performance-based teacher education. *Review of Educational Research* 44(4), 463-484.
- Hess, F. (2001). *Tear down the wall: The case for a radical overhaul of teacher certification*. Washington, DC: Progressive Policy Institute.
- Honowar, V. (2007). Gains seen in retooled teacher education. *Education Week*. Retrieved October 27, 2007 from www.edweek.org
- Houston, W.R. & Hovsarn, R. (1972). *Competency-based teacher education*. Chicago: Science Research Associates.
- Imig, D. & Imig, S. (2008). From traditional certification to competitive certification. In M. Cochran-Smith, S. Feiman-Nemser, & D.J. McIntyre (Eds.), *The handbook of research on teacher education*, 3rd edition (pp. 886-907). New York: Routledge.
- Johnson, D., Johnson, B., Farenga, S., & Ness, D. (2005). *Thrivifying teacher education: The accreditation squeeze*. Lanham, MD: Roman Littlefield.
- Kelderman, E. (2010). Teacher-education programs are unaccountable and undemanding report says/ *Chronicle of Higher Education*, July 29. Retrieved July 30, 2010 from <http://chronicle.com>
- Kisron, A., Harvey, G., & McCormick, B. (1998). Enabling the implementation of evidence-based practice: A conceptual framework. *Quality in health care*, 7, 149-158.
- Kornfeld, J., Grady, K., Marker, P., & Ruddell, M. (2007). Caught in the current: A self study of state-mandated compliance in a teacher education program. *Teachers College Record* 109(8), 1902-1930.
- Levin, H.M. (1980). Teacher certification and the economics of information. *Education Evaluation & Policy Analysis* 2(4), 5-18.
- Liston, D. & Zeichner, K. (1991). *Teacher education and the social conditions of schooling*. New York: Routledge.
- Lyall, K. & Sell, K. (2006). *The true genius of America at risk: Are we losing our public universities to de facto privatization?* Westport, CT: Praeger.
- Matus, R. (2009). State rates teacher prep programs. *St. Petersburg Times*, November 19. Retrieved January 4, 2010 from www.tampabay.com
- McCaffrey, D.M., Lockwood, J.R., Mariano, L., & Setodji, C. (2005). Challenges for Value-added assessment of teacher effects. In R. Lissitz (Ed.), *Value-added models in education: Theory and applications* (pp. 111-144). Maple Grove, MN: JAM Press.
- McConney, A.A., Schalock, M.D., & Schalock, H.D. (1998). Focusing improvement and quality assurance: Work samples as authentic performance measures of prospective teachers' effectiveness. *Journal of Personnel Evaluation in Education* 11, 343-363.
- Melton, E. (2010). Colleges slammed over teacher preparation. *Houston Chronicle*, April 28. Retrieved August 6, 2010 from texas.houstonchronicle.com, article apr282010.pdf
- Mitchell, K., Robinson, D., Plake, B., & Knowles, K. (2001). *Testing teacher candidates: The role of licensure tests in improving teacher quality*. Washington, DC: National Academy Press.
- National Academy of Education (2009). *Teacher quality*. Retrieved January 2, 2010 from www.naseducation.org
- NASDTEC (2010). Retrieved August 22, 2010 from www.nasdtc.org
- National Council on Teacher Quality (2010, April). Ed school essentials: Evaluating the fundamentals of teacher training programs in Texas. Retrieved on August 6, 2010 from <http://www.nctq.org/edschoolreports/texas>
- National Research Council (2010, April). *Preparing teachers: Building evidence for sound policy*. Washington, DC: National Academies Press.
- Neville, K.S., Sherman, R.H., & Cohen, C.E. (2005). *Preparing and training professionals: Comparing education to six other fields*. New York City: The Finance Project. Retrieved January 6, 2007 from www.financeproject.org
- Newton, S., Walker, L., & Darling-Hammond, L. (2010). *Predictive validity of the Performance Assessment for California Teachers*. Stanford University: Stanford Center for Opportunity Policy in Education.
- Noell, G.H. & Burns, J.L. (2006). Value-added assessment of teacher preparation: An illustration of emerging technology. *Journal of Teacher Education* 57(1), 37-50.
- Pechone, R. & Chung, R. (2006). Evidence in teacher education: The performance Assessment for California Teachers. *Journal of Teacher Education* 57(1), 22-36.
- Peck, C., Gallucci, C., & Sloan, T. (2010). Negotiating implementation of high-stakes performance assessment policies in teacher education: From compliance to inquiry. *Journal of Teacher Education* 61(5), 451-460.
- Planta, R. & Hamre, B. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher* 38(2), 109-119.
- Porter, A., Youngs, P., & Odden, A. (2001). Advances in teacher assessments and their uses. In V. Richardson (Ed.), *Handbook of Research on Teaching*, 4th edition (pp. 259-297). Washington, DC: American Educational Research Association.
- Prestine, N. (1989). The struggle for the control of teacher education: A case study. *Educational Evaluation & Policy Analysis* 11(3), 285-300.
- Remmert-Arrey, P. (2008). The hidden curriculum of performance-based teacher education. *Teachers College Record*, 110(1), 105-138.
- Robstner, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics* 123(1), 175-214.
- Simon, A. & Boyer, G. (1974). *Mirrors for behavior: An anthology of classroom observation instruments*, vol. 3. Philadelphia: Research for Better Schools.
- Stevens, A. (2007). Survival of the ideas that fit: An evolutionary analogy for the use of evidence in policy. *Social Policy and Society* 6(1), 25-35.
- Sykes, G. (2009, March). *Fifty years of federal teacher policy: An appraisal*. Washington, DC: Center on Education Policy.
- Sykes, G. & Dibner, K. (March, 2009). *Fifty years of federal teacher policy: An appraisal*. Washington, DC: Center on Education Policy.
- U.S. Department of Education (2009, November). *The secretary's sixth annual report on teacher quality*. Washington, DC: Office of Postsecondary Education.
- Valli, L. & Remmert-Arrey, P. (2002). New standards and assessments? Curriculum transformation in teacher education. *Journal of Curriculum Studies* 34(2), 201-206.
- Villegas, A.M. & Davis, D.E. (2008). Preparing teachers of color to confront racial/ethnic disparities in educational outcomes. In M. Cochran-Smith, S. Feiman-Nemser, & D.J. McIntyre (Eds.), *Handbook of research on teacher education*, 3rd edition (pp. 583-605). New York: Routledge.
- Walsh, K. (2004). A candidate-centered model for teacher preparation and licensure. In F. Hess, A. Rothenhan, & K. Walsh (Eds.), *A qualified teacher in every classroom?* (pp. 223-254). Cambridge, MA: Harvard Education Press.

- Walsh, K., Glaser, D., & Dunne Wilcox, D. (2006, May). *What education schools aren't teaching about reading and what elementary teachers aren't learning*. Washington, DC: National Council on Teacher Quality.
- Wang, A.H., Coleman, A.B., Coley, R.J., & Phelps, R.P. (2003, May). *Preparing teachers around the world*. Princeton, NJ: Educational Testing Service.
- Wasley, P. & McDiarmid, G.W. (2004, June). *Connecting the assessment of new teachers to student learning and to teacher preparation*. Prepared for the National Commission on Teacher and America's Future Summit on High Quality Teacher Preparation, June 28-30, Austin, TX. Retrieved August 4, 2009 from http://www.nctaf.org/resources/events/2004_summit-1/documents/Wasley-McDiarmid_Final_-_NCTAF.pdf
- Wilkinson, J.R. & Lang, W.S. (2003). Portfolios: The pried pipe of teacher certification assessments: Legal and psychometric issues. *Education Policy Analysis Archives* 11(45). Retrieved February 12, 2004 from <http://epaa.asu.edu/epaa>
- Wilson, S. (2009). Measuring teacher quality for professional entry. In D.H. Gitomer (Ed.), *Measurement issues and assessment for teaching quality*. (pp. 8-29). Thousand Oaks, CA: Sage.
- Wilson, S., Floden, R., & Ferrini-Mundi, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations*. Seattle: Center for the Study of Teaching and Policy, University of Washington, College of Education.
- Wilson, S. & Tanni, E. (2008). The evolving field of teacher education. In M. Cochran-Smith, S. Feiman-Nemser, & D.J. McIntyre (Eds.), *Handbook of research on teacher education*, 2nd edition (pp. 908-935). New York: Routledge.
- Wilson, S. & Youngs, P. (2005). Research on accountability processes in teacher education. In M. Cochran-Smith & K. Zeichner (Eds.), *Studying teacher education* (pp. 645-736). New York: Routledge.
- Wisconsin Journal of Education (1863). Examination of teachers. *Wisconsin Journal of Education* 7, 19.
- Zeichner, K. (2005). Learning from experience with performance-based teacher education. In E. Peterman (Ed.), *Designing performance assessment systems for urban teacher education* (pp. 3-19). New York: Routledge.
- Zeichner, K. (2006). Reflections of a university-based teacher educator on the future of college and university-based teacher education. *Journal of Teacher Education* 57(3), 326-340.

EDITORS' COMMENTARY

This chapter illustrates the complex interaction of institutional, state, and federal teacher education policies. Although there is general support for some form of professional accreditation among teacher educators, opinions in the policy community are mixed. Some state policymakers think accreditation by NCATE or TEAC is important enough to be mandatory, whereas others dismiss it as not relevant at all. The latter perspective is found in Chapter 4, the policy case study of Florida. As Zeichner notes, the link between most accountability measures is tenuous, but that has not stopped federal decision makers from imposing certain of these measures on the teacher education system, specifically through requirements to receive federal funds. One might wonder, then, why federal policy-makers have not required professional accreditation, such as NCATE or TEAC,

for all teacher preparation programs—those based in institutions of higher education as well as alternative route programs found elsewhere.

The answer is twofold. Many federal decision makers, both in Congress and in the U.S. Department of Education, are skeptical of the ability of those within teacher education to objectively judge their peers. Even if there was a direct correlation between professional accreditation and teacher quality, it is unlikely decision makers' skepticism would be diminished. In addition, it is important to note that there are many education lobbying groups offering perspectives on teacher education and other issues. Among these groups are national associations that represent college and university presidents. They include the American Council on Education, the National Association of State Colleges and Universities, and the National Association of Independent Colleges and Universities. Teacher education programs housed in colleges and universities are represented by one or more of these three organizations and they share a common tenet that federal intervention in higher education issues should be minimal. Moreover, their members—the college and university presidents—are wary of needing to respond to the demands of multiple professional accreditation bodies due to the cost of compliance and the demands placed on institutions. The case study of Florida is illustrative of this. Although college presidents were part of a blue ribbon commission on teacher education, they apparently did not defend a requirement that their education units retain NCATE accreditation. The organizations that represent college presidents have opposed federal mandates for teacher preparation accreditation by NCATE or TEAC in the past and it is likely this position will not change. The complex web of connections between federal expectations, interest group positions, revolving state requirements, and what is important to the teacher education community illustrates the sense-making and co-construction perspectives put forward by Darnow and Park (and presented in Chapter 1). As policies are put into place, there are nuanced interactions between policy actors and their decisions, and these interactions are constantly being influenced by the context in which they are implemented.

An important theme in Zeichner's work is that, time after time, policymakers jump to a decision whether there is empirical evidence to support that decision or not. He cites the use of value-added assessments as a contemporary example. Sykes documents many other instances when federal policies not only have no supporting evidence but were enacted in the face of contrary research (see Chapter 1). Unfortunately, if state and federal decision makers do not trust educators and teacher educators in particular, it is not surprising that they would dismiss scholarly findings conducted by them. Another aspect of the evidence and policy disconnect is the use of language. Virtually every major policy initiative since the late 1990s has referenced teacher quality: high quality is good, and poor quality is not. But drawing meaning from these words is problematic. In the Higher Education Act, a qualified teacher is one who passes the state licensing exam, but passing an exam alone does not guarantee successful teaching in all situations. The

No Child Left Behind Act defines characteristics of a highly qualified teacher which includes an academic major or minor in the teaching field. But, as Zeichner observes, how can a college major or minor guarantee an individual will perform well in the classroom? Deborah Stone (see Chapter 1) discusses the problem of multiple understandings of terms as contributors to the paradoxical nature of creating and implementing policies, and in the realm of teacher education policy this clearly is the case.

In his essay in Chapter 2, Hess joins Zeichner in questioning the utility of the current generation of value-added assessments as a means of deciding who is a good teacher and who is not. Zeichner argues that the most promising mechanism for deciding which teachers are effective is to observe them over time as they engage in their work. He acknowledges this is a time-consuming and expensive form of evaluation. This is true. What also is true is that it seems unlikely that a society unwilling to pay teachers a salary commensurate with the challenges of their work will be willing to support investments in costly evaluation mechanisms.

Zeichner offers the recommendation that comprehensive evaluation systems, both of teacher education programs and teachers themselves, be built through partnerships and structures such as professional standards boards. Chapters 6 and 7 present longitudinal policy case studies of partnerships, but the outcomes were not the same. In New Jersey the establishment and implementation of a K-16 partnership between an institution and school district is described from the points of view of a dean, superintendent, and teacher. All attest to the success of the venture. Weisenbach's description of the rise and demise of an independent professional standards and licensure board in Indiana suggests a less optimistic picture. The Indiana standards board, which included representatives from the K-12 and higher education community, seemed to enjoy initial success until key actors and the policy context changed. In New Jersey, there was continuity of leadership in the university during the decade in which the partnership was developed and implemented. The lesson here may be that K-16 partnerships, whether they are a quasi-governmental unit like a practices board or an arrangement between a school district and a university, need to be robust enough to withstand shifts in context and changes in leadership.

Discussion Questions

1. If one of the policy dilemmas is agreeing on common understandings and definitions what terms could be used to describe a teacher candidate who is ready to enter the classroom but whose experience in multiple situations has not been observed? Are there terms to describe teachers observed as effective in multiple settings? How would educators and policymakers reach agreement on these terms and how to quantify them?
2. Given the reality that leaders change jobs and every election cycle brings the opportunity for a new set of policymakers to replace existing ones, how can

3. K-16 educators protect successful partnerships from disruption by contextual changes?
4. If traditional teacher education programs are constrained by state policy and may lack support from college or university presidents, is there a future for collegiate-based teacher preparation?
5. Zeichner provides some critiques of policy and the analysis of policy, suggesting that some—Crawe, for example—selectively and inappropriately use other professions as examples in their criticism of teacher education. Do these arguments seem likely to influence policymakers and others?

TEACHER EDUCATION POLICY IN THE UNITED STATES

Issues and Tensions in an
Era of Evolving Expectations

*Edited by Penelope M. Earley,
David G. Imig, and Nicholas M. Michelli*

 **Routledge**
Taylor & Francis Group
NEW YORK AND LONDON

First published 2011

by Routledge

711 Third Avenue, New York, NY 10017

Simultaneously published in the UK

by Routledge

2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2011 Taylor & Francis

The right of the editors to be identified as the authors of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging in Publication Data

Teacher education policy in the United States : issues and tensions in an era of evolving expectations / edited by Penelope M. Earley, David G. Imig, and Nicholas M. Mitchell.

p. cm.

1. Teachers—Training of—United States. 2. Teachers—Training of—Government—policy—United States. 3. Educational change—United States. I. Earley, Penelope M. (Penelope McGaw), 1945– II. Imig, David, 1939– III. Mitchell, Nicholas M., 1942–
LB1715.T4145 2011
370.71'173—dc22

2010042964

ISBN: 978-0-415-88360-3 (hbk)

ISBN: 978-0-415-88361-0 (pbk)

ISBN: 978-0-203-84359-8 (ebk)

Typeset in Benbo

by RefineCatch Limited, Bungay, Suffolk, UK

Printed and bound in the United States of America on acid-free paper.

On behalf of all who contributed to this volume we dedicate it to our children, grandchildren, nieces, and nephews, who are our future.

