

# Competing Theories for Improving Teacher Preparation Programs: The Case of North Carolina

ECNU Review of Education  
2019, Vol. 2(2) 117–136  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2096531119853662  
journals.sagepub.com/home/roe



**G. Williamson McDiarmid**

The University of North Carolina at Chapel Hill  
East China Normal University

## Abstract

**Purpose:** Written from the perspective of a school of education dean, the article describes how deans worked with researchers and university system staff to develop multiple measures of teacher preparation program (TPP) quality.

**Design/Approach/Methods:** As a key participant in the development of a multi-methods approach to evaluating (TPP) quality, the author frequently interacted with the university system oversight staff, the researchers tasked with measuring TPP quality, and the school of education deans. The article draws on his experience in collaborating with individuals in these three groups over an 8-year period. The article also reviews the controversial history of hierarchical linear models as tools for evaluating TPPs.

**Findings:** The author argues that over time and as a result of frequent conversations and discussions, the three groups collaborated in creating a data dashboard that provides TPP faculty with evidence that enables them to make data-informed improvements to their programs and satisfies policymakers' interest in informing the public of TPP quality.

**Originality/Value:** Few, if any, articles have been published that describe a collaborative process for creating a data repository that can inform the continuous improvement of TPPs and

---

## Corresponding author:

G. Williamson McDiarmid, 2308 Cherry Laurel Lane, Chapel Hill, NC 27516, USA.  
Email: [bmcd@email.unc.edu](mailto:bmcd@email.unc.edu)



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

address policymakers' concerns about teacher quality. Educators, researchers, and policymakers elsewhere may learn about how to develop a partnership focused on generating and using data in program improvement as well as the value of multiple measures in evaluating TPPs and informing policymakers.

## Keywords

Educational policy, program evaluation, teacher candidate evaluation, teacher education

Date received: 2 February 2018; accepted: 8 May 2019

## Rising tensions

... (T)he great majority of American colleges [of education] are so incompetent and vicious that, in any really civilized country, they would be closed by the police... In the typical American State they are staffed by quacks and hag-ridden by fanatics. Everywhere they tend to become, not centers of enlightenment, but simply reservoirs of idiocy....

(Mencken, 1928, p. 5)

Although current critics of U.S. teacher education cannot rival Mencken for sheer vitriolic bombast, the sentiments that the eloquent curmudgeon expressed live on. Belief that teacher education is, in many instances, a weak intervention has led policymakers and critics to challenge, over the past several decades, the fields' grip on program accountability (Feiman-Nemser, 2001). As a result, tensions between teacher educators and some critics and policymakers have escalated. Stoking the tensions in some states has been the imposition of regulations requiring the use of pupil-learning data to determine program effectiveness (Fleener & Exner, 2011). Educators, for their part, point to problems with causal attribution as well as with the student data and the statistical models, particularly value-added models (Baker et al., 2010). These problems, they believe, undercut the validity of the results and claims about teacher effectiveness (Baker et al., 2010; Darling-Hammond, Amrien-Beardsley, Haertel, & Rothstein, 2011; Green, Baker, & Oluwole, 2012).

The release of the new, more stringent standards of the Council for the Accreditation of Educator Preparation (CAEP) has increased the pressure that many teacher educators are feeling (CAEP, 2018). The standards require programs to produce valid and reliable evidence of their graduates' impact on students' learning as well as other outcome data:

**4.1 REQUIRED COMPONENT** The provider documents, using multiple measures, that program completers contribute to an expected level of student-learning growth. Multiple measures shall include all available growth measures (including value-added measures, student-growth percentiles, and student learning and development objectives) required by the state for its teachers and available to educator

preparation providers, other state-supported P-12 impact measures, and any other measures employed by the provider (<http://caepnet.org/standards/standard-4>).

As has been the case with state and federal regulations aimed at teacher education programs, teacher educators have criticized the new standards (The American Association of Colleges for Teaching Education [AACTE], 2015; Sawchuk, 2015). Many contend that collecting the requisite data requires capacity, resources, and data access they simply do not possess (Sawchuk, 2015).

Concerned with the new regulations and standards, many—perhaps a majority—of teacher educators have been pushing back against the heightened level of accountability. Concerns extend beyond the validity of the data policymakers may use to evaluate the quality of their work. Some view public accountability requirements, at a minimum, as ill-informed affronts to their professional competence. Others suspect more sinister intent, regarding both public accountability and new accrediting standards as existential threats. In the face of these pressures, many have adopted a compliance stance, focusing primarily on meeting accountability and accrediting requirements (Peck & McDonald, 2014).

Amid these concerns and push back, some teacher educators have been looking beyond compliance to focus on program improvement. They see the availability of new data and the urgency that new regulations and standards have generated as opportunities to engage faculty and other program personnel in evidence-informed change. They have accepted that heightened accountability is the new normal (Imig & Imig, 2008). Some of these programs appear to exemplify a form of collaborative sensemaking common to communities of practice (Brown & Duguid, 2000; Wenger, 1998). This approach is intended to capitalize on rather than resist the tensions that may be inherent in an era of heightened accountability.

## **Competing theories of action**

Underlying the tensions between policymakers and teacher educators are competing theories of action (Argyris & Schon, 1974). The words and actions of some state and federal policymakers demonstrate a distrust of teacher educators and the enterprise itself. In their view, the field has proved incapable of policing itself. Educators simply lack the will to do what they know to be best practice (Cuban, 2004). As evidence, critics point to the small number of programs that fail to earn accreditation. They contrast this information with persistently mediocre standardized test results and the gap between the performance of poor students of color and white students. With relatively few tools at their disposal to address these issues, policymakers resort to mandates and sanctions (Cuban, 2004).

The market version of this theory of action holds that the publication of performance data for graduates provides potential candidates with consumer information on program quality. In this

version, enrollment in low-performing programs will tank and they will simply go out of business. Enrollment in higher performing programs and private-sector alternatives will expand the invisible hand at work (Chubb & Moe, 1988; Loeb, Valant, & Kasman, 2011).

Some teacher educators and their allies embrace a competing theory of action. According to this theory, the professional and moral values of program personnel manifest in a commitment to improving the preparation of their candidates. Needed to inform improvement efforts are both credible evidence and a program community engaged in collaboratively and continuously analyzing and interpreting evidence. This may include data on candidate and graduate classroom performance, if available. Other types of evidence—candidate performance assessments, longitudinal case studies, graduate retention rates, graduates' program evaluation, and supervisors' evaluations of graduates, to name a few—also inform program personnel about the impact of their work. Through collective sensemaking of this body of evidence, program personnel identify needed changes. After the changes are instituted, the evidence informs program personnel of the effects of the changes. And the cycle repeats (Cochran-Smith & The Boston College Evidence Team, 2009; McDiarmid & Caprino, 2017).

The assumption that appears to underlie the policymakers' theory of action is that threat of exposure or sanctions will motivate teacher educators to improve their programs. In contrast, underlying the teacher educators' theory of action is that program faculty's commitment to their students' success will motivate them to engage in improving their programs (Peck & McDonald, 2013). Skeptics of the teacher educators' theory of action point out that teacher preparation programs (TPPs), historically, have demonstrated a lack of transparency and accountability to the public (Greenberg & Walsh, 2010). For these skeptics, it follows that high-stakes accountability is necessary if teacher preparation is to improve. Or, as 18th-century man of letters Samuel Johnson is reputed to have said, "When a man knows he is to be hanged . . . it concentrates his mind wonderfully."

Not surprisingly, these competing theories of action generate tensions between TPP critics and policymakers, on the one hand, and teacher educators on the other. Exacerbating these tensions has been the increasing use of value-added models to gauge teacher effectiveness and, by extension, the quality of teachers' TPPs. The use of the latter to determine program effectiveness has been the flash point for many of the tensions (Zeichner, 2011).

## Method

To illustrate the tensions that accountability policies have generated and how institutional leaders have attempted to manage these tensions, we examine, from a dean's perspective, a specific case: North Carolina. In collaboration with deans at other state institutions, I was integrally involved in

efforts to negotiate the implementation of accountability measures in ways that addressed policy-makers' concerns and provided TPP faculty with actionable evidence.

North Carolina presents a rich case for understanding the tensions that the new era of accountability is generating for state-supported programs. This is a story of how all the parties involved—the state university system, researchers, and TPP leaders and faculty—have attempted to manage the tensions. It is a story of an emerging, if at times uneasy, middle ground in the clash between competing theories of action.

The author served as dean at the University of North Carolina (UNC) at Chapel Hill for 7 years beginning in January 2009. As dean, I served on the UNC General Administration (UNC GA) Dean's Council, a thrice-a-semester, all-day convening of the deans from all 15 of the campuses in the UNC system that prepare educators. The Assistant President of Academic Affairs for the UNC system organized and chaired these meetings. Meetings typically consisted of informational updates and member dean announcements, but the bulk of the meetings were devoted to reports from the research team. These reports include the latest research results as well as explanations of research design and methods. Deans asked clarifying questions and offered competing interpretations of the results. Over time, they increasingly questioned results and suggested additional research questions.

I came to the Deans Council conversation from having spent the previous 7 years involved with the “Teacher for a New Era” project at the University of Washington. This gave me a perspective on the UNC research grounded in working with faculty and campus leaders, as well as with colleagues at other TNE sites, on both accessing existing data and generating new data to inform program change. Most importantly, the TNE experience had taught me the importance of the social dimension of the work. That is, the imperative to engage program personnel—the people who actually do the work of teacher preparation—in defining the most urgent questions about their program, the data needed to answer the questions, and, ultimately, in making sense of the data. In the words of Dr. John Goodlad, senior advisor on the project, those who grind the wheat should make the bread.

I was fortunate in finding dean colleagues in the UNC system who shared my interest in using valid evidence of program effects to inform change as well as my uncertainty about the initial direction of the Teacher Quality Research Initiative (TQRI). This included a colleague who, as a dean in Louisiana, had experience with value-added data as part of program evaluation. The account of the TQRI work that follows is a product of my experiences and numerous conversations with dean colleagues and other teacher education leaders in the state. As with any account of the past, this is told from my particular perspective. Others involved—deans, faculty, policymakers, researchers, UNC system administrators—no doubt could provide their own account of events shaped, like mine, by their institutional position. I did ask colleagues who were key players to review drafts and have incorporated their suggestions when appropriate.

Before presenting the NC case and to better understand how these tensions have evolved, tracing the development of federal and accountability policies in some states and the concomitant appearance of new tools intended to measure teacher effectiveness is useful.

We also describe in some detail recent efforts from within the field to move beyond the debate about heightened accountability, accept it as the new normal, and focus on collecting and using valid evidence to improve the enterprise.

## Policy and tools developments

The reauthorization of the *Higher Education Act (HEA)* (1998), the *Race to the Top (RTTP)* initiative (2010), and waivers for *No Child Left Behind (NCLB)* required TPPs to collect and make public a range of outcome data appeared amid a debate over the future of university-based teacher preparation (Education Week, 2015; Gatlin, 2008; Imig, 2011; Sawchuk, 2015). Adding fuel to the fire were, as noted above, the new CAEP accreditation standards.

As policymakers increased accountability pressures on TPPs, tools that they believed could be used to determine program effectiveness—notably, value-added and student growth models—became available. For many policymakers, these tools represent economical and putatively objective means to determine program quality. As argued above, the use of statistical models for measuring teacher classroom effects as a tool to evaluate TPP effectiveness underlies an implicit theory of action: If TPPs are required to make public the value-added or growth model data on their graduates' classroom effectiveness, low-performing programs can be identified and closed, leading to higher quality teachers in classrooms who will, in turn, help students perform better on test.

### *The evolution of federal accountability policies*

As Cuban (2004) has observed, U.S. public schools have been accountable since their inception. Similarly, state-supported TPPs have been held accountable by state authorities since the establishment of normal schools in the 19th century (Fraser, 2007). The passage of ESEA in 1965 instituted annual evaluations of schools receiving Title I funds, marking a new level of public accountability (Cuban, 2004).

In the 1980s, public accountability and critical scrutiny of schools and teacher education shifted into a higher gear as states and the federal government flexed their regulatory muscles. Ushering in the modern era of accountability was the publication, in 1983, of *A Nation at Risk*. The authors of this widely disseminated and widely quoted report accused the U.S. of committing an act of unthinking, unilateral educational disarmament (National Commission on Excellence in Education, 1983). The call for standards and accountability appears to mark the moment that, as McCluskey (2011) claims, the self-interest of federal policymakers and educators started to

diverge. Further fueling the accountability movement were data from the Carnegie Foundation for the Advancement of Teaching (CFAT): “[A]n indictment of the schools, CFAT reported that three-quarters of the corporations surveyed had to provide their employees instruction in computation, reading, and math” (Eurich & Wade, 1986, p. 21).

Against this backdrop in the 1980s and 1990s, several states embarked on standards-based reforms, developing their own standards, assessment, and accountability systems (Carnoy & Loeb, 2002). Texas, California, Kentucky, North Carolina, and South Carolina were among the states that led the way. Carnoy and Loeb (2002) identify as the hallmarks of these reforms: aligning standards, curriculum, and assessments and improving the capacity of educators. These elements of state-level standards-based reforms foreshadowed the passage *NCLB* legislation in 2001.

At the federal level, the 1994 reauthorization of ESEA institutionalized standards-based accountability for schools, but TPPs were largely spared until the reauthorization of the *HEA* in 1998. A significant impetus for this legislation was the publication of *What matters most: Teaching for America’s future* by the National Commission on Teaching and America’s Future (Earley, 2001). The addition of Title II to the original 1965 *HEA* legislation imposed new reporting requirements on institutions that educate teachers. By spotlighting research on teachers’ impact on student achievement, *What Matters Most* emboldened critics of university-based teacher preparation. Given research that suggested TPPs are a weak intervention, some critics questioned the need for licensure requirements at all (Walsh, 2006).

The final Title II regulations required states to report annually on (1) program graduates’ scores on the state’s licensure examination; (2) enrollment in preparation programs; (3) length of required school experience; (4) state program approval requirements; and (5) whether the state had labeled any of the TPPs as low performing. The latter required the states to develop indicators for identifying low-performing programs and to make information on TPPs publicly available. Other requirements included (1) details of licensing requirements; (2) alignment between K-12 content standards and licensure requirements; (3) licensure examination results for each program; (4) licensure waivers approved; (5) descriptions of alternate routes and licensure test pass rates for participants; and (6) information on required subject–matter tests (Earley, 2001).

At the beginning of the 21st century, the passage of *NCLB* ratcheted up accountability pressures even further by including sanctions for P-12 schools that failed to make adequate yearly progress. This represented a departure from the standards-based reforms that states such as Kentucky and California instituted. Rather than sanctioning low-performing schools, these states have offered them additional resources and supports (Mintop & Trujillo, 2007; Steffy, 1993). This approach had proved effective, particularly in Kentucky (David, Coe, Kannapel, McDiarmid, & Mazur, 2003).

*NCLB* required states, for the first time, to link pupil achievement and growth data to their teachers and the teachers, in turn, to their in-state preparation programs. *RTTP* further required the

recipient states to publish report cards for TPPs at state-supported institution that included the student achievement results for program graduates. *RTTT* funds included funding to build state data systems that enabled linking student achievement data with individual teachers and teachers with their preparation pathway.

In short, federal and state policies over the past three decades progressively increased accountability pressures on both schools and TPPs. Central to this progression were standardized student assessment results that, initially, state and federal authorities used to evaluate P-12 schools and, more recently in some states, are using to evaluate teachers and their preparation programs. The Trump administration, however, rolled back many Obama-era regulations, including elimination of the existing Title II of the *HEA*. All TPP reporting requirements enumerated above have been eliminated. Replacing teacher preparation in Title II is a program called expanding access to in-demand apprenticeships (Center for American Progress, 2017).

### *State-level accountability and the role of value-added data*

As the accountability spotlight widened beyond schools to include university-based TPPs, technology emerged that many policymakers regarded as the very tool needed to measure the effects of TPPs on their graduates' performance. In the mid-1990s, William Sanders at the University of Tennessee developed a hierarchical linear model designed to isolate a teacher's influence on student achievement from other variables that could affect student performance (Sanders & Horn, 1994). Citing his work and that of other researchers, Sanders claimed that, based on millions of student achievement records, teacher effectiveness—as determined by value-added methods—contributed more to student achievement than any other factor (Sanders & Rivers, 1996).

In the first decade of this century, several states built the data systems needed to support value-added models designed to determine teacher effects on student achievement and link teachers to their TPPs. Among the first and best known are Louisiana and Tennessee.

In Louisiana, the impetus for data-based reform came from the Board of Regents for the University of Louisiana. In 2003, the Regents recruited LSU Professor George Noell to develop a value-added model as part of their larger teacher preparation reform initiative (Board of Regents State of Louisiana, n.d.). In narrating the story of teacher preparation reform in Louisiana from their perspective as teacher preparation leaders, Fleener and Exner describe the development of an assessment and accountability system that incorporated multiple measures of teacher effectiveness (Fleener & Exner, 2011). They describe a data system that links scores for individual P-12 pupils across time and the scores of pupils to individual teachers. Such links are necessary to make claims about program effects. The model enabled researchers to compare newly prepared teachers in specific content areas to experienced teachers to identify programs that were performing well



(Noell & Burns, 2006). They emphasize that the Louisiana value-added model was designed to provide effect scores for TPPs rather than for individual teachers (Fleener & Exner, 2011).

Tennessee also pioneered the use of value-added student data to hold TPPs accountable. The Tennessee Value-Added Assessment System (TVAAS) provides data on pupil growth that is linked to individual teachers and their TPP. These data are one indicator included on the report card that the state, since 2007, produces for each program. Initially, placement and retention rates and Praxis II scores were included with the TVAAS data. The state currently plans to redesign the report card to include surveys of employers and graduates as well as observational data (Tennessee State Board of Education, n.d.).

As reported by the Data Quality Campaign (2014), by 2014, 22 states share information about how teachers perform in the classroom with educator preparation programs, providing data to inform improvements in teacher training (p. 5). This represents a threefold increase between 2011 and 2014. The picture presented by the Dean for Impact organization is somewhat different (Deans for Impact, n.d.). Among the nonrandom sample of 23 member programs scattered across 17 states, only 6 reported having access to student-achievement data connected to their graduates and less than a third could access performance data, such as classroom observations, for their graduates (p. 6). That teacher education leaders at the campus level have difficulty accessing data on their graduates' performance even in states that collect such data should not be surprising. States no doubt vary in the quality and accessibility of such data just as campuses vary in their capacity and resources to use the data. Even on campuses that have access to data on their graduates' classroom performance, experience with and capacity to use these data to improve can be a challenge (McDiarmid & Caprino, 2017; Moss & Piety, 2007; Peck & McDonald, 2014).

As of this writing, it is too early to determine the impact on state policy of the Trump administration's decision to eliminate all reporting requirements for TPPs. Officials in some states may believe this shifts the burden of ensuring the quality of preparation more squarely on their shoulders. This could lead to even greater scrutiny of programs at the state level where, historically, it has been shared in many states with the national accrediting body (Fraser, 2007).

## **Data for what purpose?**

### *Accountability and program improvement tensions in the case of North Carolina*

North Carolina presents a rich case for understanding the tensions that the new era of accountability is generating for state-supported programs. This is a story of how all the parties involved—the state university system, researchers, and TPP leaders and faculty—have attempted to manage the tensions. It is a story of an emerging, if at times uneasy, middle ground in the clash between competing theories of action.

*Impetus for UNC TQRI.* The changes in TPPs at UNC system institutions may not have happened or not happened as rapidly were it not for pressure from the UNC GA. The pressure came, initially, from the President of the UNC system president, Erskine Bowles. In 2008, he tasked the Academic Affairs Office in the UNC GA with conducting research on the effectiveness of the programs at the 15 campuses in the system that prepare teachers. He was specifically interested in developing a value-added model for the TPPs in the system.

To conduct the research, UNC GA turned to the newly created Educational Policy Initiative at Carolina (EPIC), housed at the UNC at Chapel Hill. Just as the TQRI got underway, researchers at EPIC also became engaged as evaluators for the newly announced RTTT project in North Carolina. Having already developed a value-added model for the TPPs in the UNC system, EPIC researchers were well positioned to evaluate the impact of RTTT. Given the requirement that performance data on state-supported TPPs be made public, the stakes for the state TPPs became greater than ever. Thus, beginning in 2011, campus and TPP leaders and faculty found themselves under unprecedented scrutiny in a broader political environment that was skeptical of, if not hostile toward, their work.

*Hostile political environment.* As the economy tumbled into a deep recession in 2008–2009 and the state cut funding to higher education, the UNC system leaders, like many others in public higher education, sought ways to reduce costs. Newspapers reported that the UNC Board of Governors was considering closing one or more campuses and duplicative programs. In addition, the political tide that swept conservatives into power across the country in 2010 also washed across NC. Many newly elected legislators were intent on bolstering market-based educational alternatives and cut both higher education and P-12 budgets deeply (NC Policy Watch, 2015). Not surprisingly, deans at several campuses suspected that their school of education and, perhaps, their campus itself were under an existential threat. Despite assurances to the contrary, some understandably believed the goal of the TQRI was to identify and target programs for closure.

The use of EPIC's value-added model to determine the impact that the graduates of each of the 15 UNC institutions had on their students' state test performance reinforced this belief (Henry et al., 2014). UNC GA began, in 2011, to provide provosts and the education and arts and sciences deans at each of the UNC campuses with the aggregate value-added results for their TPP graduates. Several deans reported that the value-added model (VAM) results put them on the defensive. Whether or not their provost understood the statistical model mattered less than their impression of the data: Was their education school or college doing well or not?

*Responses to data varied by institutional type.* The degree of perceived threat and suspicion varied considerably by institutional context. Initially, uncertain of the local reactions to the VAM data, many of the deans worried that campus and system leaders would use the data to evaluate their

performance as well as that of their school. In particular, the deans at several of the minority-serving institutions faced context-specific challenges and scrutiny that were often more threatening than those their colleagues at majority-white institutions were facing.

Across the 15 campuses, institutional contexts and historical missions vary widely. Six of the 15 campuses, including one Historically Black College and University (HBCU), are doctoral-granting institutions with varying levels of research activity. Seven are master's institutions and two are bachelor's institutions. One of the three master's and one of the bachelor's institutions are also HBCUs and one is an institution that historically has served the American Indian communities in the southern part of the state. Perceptions of the VAM data varied by institutional context. In addition, the capacity and resources to interpret and explain the value-added results to campus leaders, faculty, and constituents varied by institution. (These variations also limit the generalizability of any claims based on the data.)

After the initial dissemination of the VAM findings to the campuses, several deans reported that their faculty reacted critically and defensively. Faculty with strong statistical backgrounds criticized what they believed to be technical problems with VAMs that they believed undercut their validity. The concerns they raised echoed those raised by other scholars: nonrandom samples of graduates due both to lack of test scores in non-test grades and subjects as well as graduates who left the state, the lack of baseline data needed for causal attribution, reliance on a single measure of student learning whose validity is itself questionable, stability of teacher effects over time, fixed school effects, the difficulties of understanding and explaining to others the VAM's technical complexities, and so on (Baker et al., 2010; Braun, 2005; Haertel, 2013; Lockwood et al., 2006; Schochet & Chiang, 2010).

To say that, by 2012, teacher educators at the UNC campuses were feeling generally under siege would be an understatement. The literature is replete with examples of P-12 educators' responses to accountability pressures (Finnigan & Gross, 2007; Whitford & Jones, 2000). Research on the responses of teacher educators is scarcer. Given the variety of institutions and contexts, blanket claims about how teacher educators respond seem premature, at best.

From a wide-angle view, research on public sector employees' responses to accountability is, however, relevant. This research suggests, predictably, that the imposition of performance indicators frequently prompts defensiveness as well as unproductive strategies for avoiding negative consequences (Smith, 1995). As Peck and McDonald (2013) found, some teacher educators respond by focusing on compliance—an understandable response given what is at stake.

*Evolving relationship among system leaders, researchers, and campuses.* Despite these rocky beginnings, the relationship between the UNC deans and faculties of education, on the one hand, and the EPIC researchers and UNC GA leaders, on the other, evolved in ways that offered the promise of

progress in improving programs at many of the UNC institutions. Different institutions are, predictably, at different points in this evolution. Overall, all the actors are attempting to balance the multiple purposes of public performance indicators such as those in North Carolina: public accountability, consumer information, and program improvement (Feuer, Floden, Chudowsky, & Ahn, 2013). Superficially, these purposes may appear compatible. They exist, however, in political, historical, and cultural contexts that, as we have seen in North Carolina, bring them into conflict.

Aware of the broader political context, the EPIC team and UNC GA leaders took steps to alleviate some of the accountability pressure that the deans were feeling. Rather than publicly disseminate campus-by-campus VAM results, EPIC researchers categorized the data by entry portals. These portals included UNC-prepared teachers, out-of-state prepared teachers, in-state private prepared, alternate entry, and so on. This, fortuitously, proved valuable in the political sphere: The researchers found that students of teachers prepared at UNC institutions, in aggregate, significantly outperformed the students of teachers prepared outside of NC as well as lateral entry teachers—or, in some areas, no worse (Bastian, Patterson, & Pan, 2015; Henry et al., 2010).

At the same time, provosts and deans at the UNC institutions received the data for their graduates. EPIC researchers and UNC GA leaders visited each of the campuses to help campus leaders and deans understand the data and discuss how to use the data to improve their TPPs. Subsequently, some deans began to use the VAM data, in conjunction with other evidence, to assess programmatic strengths and weaknesses (Bastian, Fortner, et al., 2015).

In addition, most of the UNC Deans Council thrice-a-semester meetings were devoted to presenting and discussing the VAM data. Initially, many of the deans' questions focused on understanding the model itself and the meaning of the results. This was an instantiation of UNC GA's commitment to transparency and collaboration. Over time, the deans began to ask more critical questions, offer their own interpretations of the data, voice their concerns about the research methodology, and request additional analyses.

For example, the deans hypothesized that the types of schools—beyond the school-level variables included in the VAM—in which their graduates typically taught could help explain the differential VAM results across programs. The resulting analysis revealed that graduates from the selective research campuses were more likely to be hired by schools in which pupils were already scoring above the average of schools in the state than were graduates of less selective institutions (Bastian & Henry, 2015). Several of these less selective institutions were minority-serving institutions. Graduates of these institutions were typically hired by schools whose test scores were below the average for schools in the state. These are schools that, historically, have counted on these institutions to prepare teachers for them (Dilworth, 2012).

These data on the pupil test scores for the schools that hired graduates of each of the 15 campuses in the UNC system provided context for better understanding the value-added results

for the various institutions. Raudenbush, one of the statisticians responsible for the architecture of VAMs (Raudenbush & Bryk, 1986), has pointed out that fixed school effects can significantly impact value-added result for teachers:

A growing body of evidence suggests that schools can vary substantially in their effectiveness, potentially inflating the value-added scores of teachers assigned to effective schools. Schools also vary in contextual conditions such as parental expectations, neighborhood safety, and peer influences that may directly support learning or that may contribute to school and teacher effectiveness. Moreover, schools vary substantially in the backgrounds of the students they serve, and conventional statistical methods tend to break down when we compare teachers serving very different subsets of students. (Raudenbush, 2013)

This confirmed what many of the deans believed, both intuitively and empirically, to be the case. The analysis equipped them with credible evidence that school context impacts pupil achievement in ways unaccounted for by value-added models alone. This helped them explain the results to their campus and community leaders as well as policymakers and other stakeholders.

Gradually, through collaborative work such as this, a greater level of trust developed between the researchers and a number of the deans. Helping to nurture that trust has been the co-constructing, over the course of several years, of a common research agenda and sharing data with the campuses for the primary purpose of program improvement rather than accountability (see Bastian, Fortner, et al., 2015).

*Use of multiple measures.* During this time, at the urging of the deans and with the support of UNC GA, the TQRI expanded beyond the value-added data to include other data sources. These include a common graduate survey and performance assessments, specifically the edTPA. Interaction between the deans and the researchers led to a pilot study correlating VAM data with edTPA scores to test the accuracy of local scoring of portfolios and a study of the noncognitive skills of candidates at one of the campuses (Bastian, Henry, Pan, & Lys, 2015).

The adoption of the edTPA was a gradual, bottom-up process. Concerned about the power of the VAM data to inform and bring about change in their programs, several of the deans, on their own initiative, piloted the use of the edTPA in their programs. They also provided support to colleagues on other campuses who wished to implement the edTPA. Motivating the adoption of the edTPA was a concern among some of the deans that the grain size of the value-added data was too large to inform program changes. As one of them noted, the value-added data provides a 30,000-foot view of a program. Needed, they believed, were data that more closely reflected the work of program personnel and were sufficiently fine-grained to inform program change. This reflects the experience of leaders at other institutions that created a portfolio of instruments to provide various types of qualitative and quantitative measures to both identify program weaknesses and inform changes (McDiarmid & Caprino, 2017).

Consistent with their theory of action, the deans committed to program improvement recognized that perhaps the most powerful resource for change is their faculty's and staff's deep moral commitments to their students (Cochran-Smith & The Boston College Evidence Team, 2009; Fullan, Cuttress, & Kilcher, 2009; McDiarmid & Caprino, 2017). Seeing in candidates' portfolios detailed evidence of whether their candidates are taking up and using the skills and knowledge taught in the program has proved a powerful spur to rethinking program content, pedagogy, and design (Cochran-Smith & The Boston College Evidence Team, 2009; Peck, Galucci, & Sloan, 2010).

UNC GA and the state, for their part, complied with the *RTTP* requirement to publish a report card for the state-supported preparation programs. At the same time, UNC GA was committed to providing programs with valid data. In collaboration with the SAS Institute, GA rolled out a publicly accessible Educator Quality Dashboard (<http://eqdashboard.northcarolina.edu/>) in (2015). Available on the dashboard are a range of data on UNC system TPPs: selection criteria for candidates; retention of graduates in NC classrooms; descriptions of the TPPs; university-school partnerships; supervisor evaluations; and performance data including program-by-program data from the EPIC-developed VAM. Plans include publishing additional data such as edTPA results. Some campuses have begun using the dashboard data to examine their programs and make evidence-based changes (Bastian, Fortner, et al., 2015). Involving additional campuses in similar work has been hampered by frequent leadership turnover at both the school and campus levels.

As others have observed, gaining access to valid data is just the beginning of the process of bringing about evidence-based change (Cochran-Smith & The Boston College Evidence Team, 2009; Peck et al., 2010). Engaging faculty and staff in using the data to inquire into their questions about the impact of their programs requires leadership and commitment (McDiarmid & Caprino, 2017; Peck & McDonald, 2013). Some deans may still be wary about perceptions of the published data on their campuses, in their communities, and around the state and nation. Some may still be waiting for the other shoe to drop. This reflects the tension at the core of public performance indicators such as value-added results.

## Conclusions

The story of the evolution of TPP evaluation and accountability in NC may provide some useful lessons going forward. Perhaps, most critical has been the development, over time and not without some bumps along the way, of a productive relationship among UNC GA, EPIC researchers, and the deans of education and their faculties. At the same time, pressure to focus on compliance and, secondarily, on program improvement is strong. Where leaders and their faculty have steered a productive course between improvement and compliance offers both proof of concept and hope that such a course is possible (McDiarmid & Caprino, 2017).

Viewing the unfolding story in NC through the lenses of the competing theories of action described at the outset of this article is also instructive. The theories conflict on the assumption about motivations for human behavior—a classic conflict between extrinsic and intrinsic motivation (Ryan & Deci, 2000). The NC story suggests that both may be critical to motivating leaders and program faculty to change. Deciding on exactly what changes will lead to improved outcomes depends on both the quality of the evidence and the process by which the evidence is interpreted. Engaging program faculty in collectively making sense of the evidence appears critical to making the changes in the program believed necessary to improve outcomes (McDiarmid & Caprino, 2017).

The NC story also points out the limitations of value-added data alone to inform program improvement. Leaving aside the considerable technical problems and problems of attribution, two major issues arose as the value-added data on graduates became available to TPPs. Faculty require data at the program component or experience level to know what needs to be changed. In the absence of such finer-grained data, faculty are far less likely to take individual responsibility for the results. Seeing their fingerprints on the evidence engages their moral commitment to their students and to the mission of preparing teachers ready for the challenges of the classroom.

## Epilogue

The following article by Song and Xu (2019) in this volume presents a valuable history of the evolution of governmental policies intended to raise the quality of teacher preparation in China. Viewing the story they tell through the theory-of-action lens described above, Chinese Ministry of Education policies have evolved from the imposition of standards to the use of a rigorous examination as tools intended to improve preparation quality. That is, the focus of policies has changed from a focus on inputs (i.e., standards and standardized curriculum) to a focus on outputs (i.e., written examination performance). These policy approaches share a similar underlying theory of change: To improve preparation programs, externally mandated standards and determination of adequate professional knowledge are required. Chinese policymakers appear to share with their U.S. counterparts' faith that market forces will drive out programs whose candidates perform poorly. In China, the teacher examination is the tool used to measure teacher candidate quality; in the U.S., in several states, the tool is the performance of preparation program graduates' pupils on state tests. In both cases, the results are made publicly available to inform the application decisions of potential students and their families.

In the Chinese case, this theory of change rests on the assumption that performance on the teacher examination validly reflects the skills and knowledge teachers need to succeed. That is, the examination results validly predict teacher candidates' classroom readiness. I do not know if the predictive validity of the teacher examination has been established but the mere fact that this is

a paper-and-pencil (or computer-based) test raises questions. Specifically, is it possible to measure with such an assessment the myriad pedagogical skills—such as responding to pupil questions, managing pupil behavior, adapting subject matter representations to diverse learners on-the-fly, and so on—that teachers require (Ball & McDiarmid, 1990; McDiarmid & Ball, 1988)?

The other issue that arises is that of how to improve programs, not merely weed out the ineffective from the effective. The assumption in China and some U.S. states is that the data from teacher examinations—in the Chinese case—or from value-added models—in the U.S. case—will provide teacher educators with the evidence needed to improve their programs. Song and Xu note that Chinese teacher educators do receive item-level data on their candidates' examination performance. For this evidence to motivate and guide teacher educators' program improvement efforts, however, the evidence must be seen to accurately reflect the teachers' required skills and knowledge.

As seems clear from the U.S. example, an approach to determining TPP quality that relies on evidence of questionable validity undervalues and disrespects the faculty knowledge and moral commitment. Song and Xu quote a teacher education faculty member as saying about the teacher examination:

Our institution requires us to carefully study the examinations and teach to them, and even requires students to memorize the standard answers. I really can't accept this. Because there simply is no agreement about different theories in the field of education and there is no standard correct answer to many problems. In the 'no rules method' [for example], the teaching process itself requires the teacher to rationally choose the method for instruction that best fits the needs of the learners; how could a paper examination test that? But now we have no choice. It's like university education is now 'teaching to the test,' too. (Song & Xu, 2019, p. 154.)

This quotation reflects the frustration teacher educators often feel when their professional knowledge is ignored in favor of what they view as an oversimplified and inaccurate measure of teaching competence. As I argued above, most teacher educators are deeply committed to their students' success. This is the motivation for them to improve their programs. To do so requires access to evidence that, they feel, validly reflects the effects of their courses and programs. The Chinese teacher educator quoted above has fundamental concerns about the validity of a paper-and-pencil evaluation of an activity as socially and intellectually complex as teaching. Consequently, this educator—and no doubt, many others—is unlikely to regard the results from the examination as accurately reflective of how well their students have been prepared to teach. They are also unlikely to view the results as a valid and compelling basis for making improvements to their preparation programs.

Using the examination results as a de facto mechanism for driving out of the market programs whose candidates perform poorly seems unlikely to produce a more effective system for preparing well-qualified teachers. It seems more likely to result in a system effective in producing teachers



who are adept examination takers. Arguably, the key to better prepared teachers are better TPPs grounded in evidence of candidates' and graduates' classroom performance. This is the theory that underlies the argument for using performance assessments to determine the extent to which program participants are taking up and using the skills and knowledge they are taught in their program (Peck & McDonald, 2013). This is the evidence teacher educators need.

### **Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### **References**

- Argyris, C., & Schon, D. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco, CA: Jossey Bass.
- Baker, E., Barton, P., Darling-Hammond, L., Haertel, E., Ladd, H., Linn, R., . . . Shepard, L. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: The Economic Policy Institute.
- Ball, D., & McDiarmid, G. (1990). The subject matter preparation of teachers. In R. Houston (Ed.), *Handbook of research on teacher education* (pp. 437–449). New York, NY: Macmillan.
- Bastian, K., Fortner, K., Chapman, A., Fleener, J., McIntyre, E., & Partiaracha, L. (2015). *Data sharing to drive the improvement of teacher preparation programs*. Chapel Hill, NC: Education Policy Initiative at Carolina.
- Bastian, K., & Henry, G. (2015). Teachers without borders: Consequences of teacher labor force mobility. *Educational Evaluation and Policy Analysis, 37*, 163–183.
- Bastian, K., Henry, G., Pan, Y., & Lys, D. E. (2015). *Evaluating a pilot of the teacher performance assessment: The construct validity, reliability, and predictive validity of local scores*. Chapel Hill, NC: Education Policy Initiative at Carolina.
- Bastian, K., Patterson, T., & Pan, Y. (2015). *Teacher preparation program effectiveness report*. Chapel Hill, NC: Education Policy Initiative at Carolina.
- Board of Regents State of Louisiana. (n.d.). *Value-added teacher preparation program assessment model*. Retrieved August 12, 2016, from <http://www.regents.la.gov/page/value-added-teacher-preparation-program-assessment-model>
- Braun, H. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational testing Service.
- Brown, J., & Duguid, P. (2000). *The social life of information*. Cambridge, MA: Harvard Business School Press.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? *A cross-state analysis*. *Educational Evaluation and Policy Analysis, 24*, 305–331.
- Center for American Progress. (2017, December 7). *What you need to know about the house higher education act bill*. Retrieved February 14, 2018, from <https://www.americanprogress.org/issues/education-postsecondary/news/2017/12/07/443915/need-know-house-higher-education-act-bill/>

- Chubb, J., & Moe, T. (1988). Politics, markets, and the organization of schools. *American Political Science Review*, 82, 1065–1087.
- Cochran-Smith, M., & The Boston College Evidence Team. (2009). “Re-culturing” teacher education: Inquiry, evidence, and action. *Journal of Teacher Education*, 60, 458–468.
- Council for the Accreditation of Educator Preparation. (2018). *Introduction: The CAEP Standards*. Retrieved February 12, 2018, from <http://caepnet.org/standards/introduction>
- Cuban, L. (2004). Looking through the rearview mirror at school accountability. In K. Sirotnik (Ed.), *Holding accountability accountable* (pp. 18–34). New York, NY: Teachers College Press.
- Darling-Hammond, L., Amrien-Bearsley, A., Haertel, E., & Rothstein, J. (2011, September 14). Getting teacher evaluation right: A background paper for policy makers. Retrieved November 21, 2016, from <http://www.aera.net/Portals/38/docs/New%20Logo%20Research%20on%20Teacher%20Evaluation%20AERA-NAE%20Briefing.pdf>
- Data Quality Campaign. (2014, November 1). Paving the path to success. Retrieved August 14, 2016, from <http://dataqualitycampaign.org/resource/data-action-2014-paving-path-success/>
- David, J., Coe, P., Kannapel, P., McDiarmid, G., & Mazur, J. (2003). *Improving low-performing schools: A study of Kentucky’s highly skilled educators program*. Lexington, KY: Partnership for Kentucky Schools.
- Deans for Impact. (n.d.). *From Chaos to Coherence*. Retrieved August 14, 2016, from [http://deansforimpact.org/policy\\_brief.html](http://deansforimpact.org/policy_brief.html)
- Dilworth, M. (2012). Historically black colleges and universities in teacher education reform: Where are we? *The Journal of Negro Education*, 81, 121–135.
- Earley, P. (2001). Title II requirements for schools, colleges, and departments of education. Retrieved August 12, 2016, from <http://www.ericdigests.org/2002-3/title.htm>
- Education Week. (2015). *Teacher education group airs criticism of new accreditor*. Retrieved February 12, 2018, from <https://www.edweek.org/ew/articles/2015/03/18/teacher-education-group-airs-criticism-of-new.html>
- Eurich, N., & Wade, B. (1986). *Corporate classrooms: The learning business (A Carnegie Foundation special report)*. San Francisco, CA: Jossey-Bass.
- Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record*, 103, 1013–1055.
- Feuer, M., Floden, R., Chudowsky, N., & Ahn, J. (2013). *Evaluation of teacher preparation programs: Purposes, methods, and policy Options*. Washington, DC: National Academy of Education.
- Finnigan, K., & Gross, B. (2007). Do accountability policy sanctions influence teacher motivation? Lessons from Chicago’s low-performing schools. *American Educational Research Journal*, 44, 594–629.
- Fleener, J., & Exner, P. (2011). Dimensions of teacher education accountability: A Louisiana perspective of value-added. In P. D. Imig & N. Michelli Earley (Eds.), *Teacher education policy in the United States* (pp. 26–43). New York, NY: Routledge.
- Fraser, J. (2007). *Preparing America’s teachers*. New York, NY: Teachers College Press.
- Fullan, M., Cutfress, C., & Kilcher, A. (2009). Chapter 2: 8 forces for leaders of change. In M. Fullan (Ed.), *The challenge of change: Start school improvement now!* (pp. 9–20). Thousand Oaks, CA: Sage.
- Gatlin, D. (2008). *Thinking outside the university: Innovation in alternative teacher certification*. Chicago, IL: Center for American Progress.

- Green, P., Baker, B., & Oluwole, J. (2012). Legal implications of dismissing teachers on the basis of value-added measures based on student test scores. *BYU Education and Law Journal*, 1, 1–29.
- Greenberg, J., & Walsh, K. (2010). Ed school essentials: Evaluating the fundamentals of teacher training programs in Texas. Retrieved September 5, 2016, from [http://www.nctq.org/dmsView/Ed\\_School\\_Essentials\\_Training\\_Programs\\_Texas\\_Executive\\_Summary\\_pdf](http://www.nctq.org/dmsView/Ed_School_Essentials_Training_Programs_Texas_Executive_Summary_pdf)
- Haertel, E. (2013). *Reliability and validity of inferences about teachers based on student test scores*. Princeton, NJ: ETS.
- Henry, G., Purtell, K., Bastian, K., Fortner, C., Thompson, C., Campbell, S., & Patterson, K. (2014). The effects of teacher entry portals on student achievement. *Journal of Teacher Education*, 65, 7–23.
- Henry, G., Thompson, C., Bastian, K., Fortner, K., Kershaw, D., Purtell, P., & Zulli, R. (2010). Portal report: Teacher preparation and student test scores in North Carolina. Retrieved August 17, 2016, from [http://publicpolicy.web.unc.edu/files/2014/02/Portal\\_TeachPrep-TestScore\\_June2010\\_Final.pdf](http://publicpolicy.web.unc.edu/files/2014/02/Portal_TeachPrep-TestScore_June2010_Final.pdf)
- Imig, D. (2011). The future of teacher education. In P. Earley, D. Imig, & N. Michelli (Eds.), *Teacher education policy in the United States* (pp. 215–219). New York, NY: Routledge.
- Imig, D., & Imig, S. (2008). From traditional certification to competitive certification: A twenty-five year retrospective. In M. Cochran-Smith, S. Feiman-Nemser, D. McIntyre, & K. Demers (Eds.), *Handbook of research on teacher education* (3rd ed., pp. 886–907). New York, NY: Routledge.
- Lockwood, J., McCaffrey, D., Hamilton, L., Stecher, B., Le, V., & Martinez, F. (2006). *The sensitivity of value-added teacher effect estimates to different mathematics achievement measures*. Santa Monica, CA: The RAND Corporation.
- Loeb, S., Valant, J., & Kasman, M. (2011). Increasing choice in the market for schools: Recent reforms and their effects on student achievement. *National Tax Journal*, 64, 141–164.
- McCluskey, N. (2011). Federal politics and policy: A misalignment of incentives. In P. Earley, D. Imig, & N. Michelli (Eds.), *Teacher education policy in the United States* (pp. 20–23). New York, NY: Routledge.
- McDiarmid, G., & Ball, D. (1988). “Many Moons”: Understanding teacher knowledge from a teacher education perspective. *The Journal*, 2, 23–37.
- McDiarmid, G., & Caprino, K. (2017). *Lessons from the Teachers for a New Era project: Evidence and accountability in teacher education*. New York, NY: Routledge.
- Mencken, H. (1928, December 31). The war upon intelligence. *Baltimore Evening Sun* (p. 5).
- Mintop, H., & Trujillo, T. (2007). The practical relevance of accountability systems for school improvement: A descriptive analysis of California schools. *Educational Evaluation and Policy Analysis*, 29, 319–352.
- Moss, P., & Piety, P. (2007). Evidence and decision making. In P. Moss (Ed.), *106th Yearbook of the National Society for the Study of Education* (pp. 1–14). Malden, MA: Blackwell.
- National Commission on Excellence in Education. (1983). *A Nation at Risk: The imperative for educational reform*. Washington, DC: U.S. Department of Education.
- NC Policy Watch. (2015). Altered State: How 5 years of conservative rule have redefined North Carolina. Retrieved August 19, 2016, from <http://www.ncpolicywatch.com/category/altered-state/http://www.ncpolicywatch.com/wp-content/uploads/2015/12/NC-Policy-Watch-Altered-State-How-5-years-of-conservative-rule-have-redefined-north-carolina-december-2015.pdf>
- Noell, G., & Burns, J. (2006). Value added assessment of teacher preparation: An illustration of emerging technology. *Journal of Teacher Education*, 57, 37–50.

- Peck, C., Galucci, C., & Sloan, T. (2010). Negotiating implementation of high-stakes performance assessment policies in teacher education. *Journal of Teacher Education, 61*, 451–463.
- Peck, C., & McDonald, M. (2013). Creating “Cultures of evidence” in teacher education: Context, policy, and practice in three high-data-use programs. *New Educator, 9*, 12–28.
- Peck, C., & McDonald, M. (2014). What is a culture of evidence? How do you get one? And . . . Should you want one? *Teachers College Record, 116*. Available at: <https://www.tcrecord.org>
- Raudenbush, S. (2013, August 13). What do we know about using value-added to compare teachers who work in different schools? *Carnegie Knowledge Network*. Retrieved August 17, 2016, from <http://www.carnegieknowledgenetwork.org/briefs/comparing-teaching/>
- Raudenbush, S., & Bryk, T. (1986). A hierarchical model for studying school effects. *Sociology of Education, 59*, 1–17.
- Ryan, R., & Deci, E. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54–67.
- Sanders, W. L., & Horn, S. P. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8*, 299–311.
- Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.
- Sawchuk, S. (2015, January 7). Despite monitoring, ed. School closures are rare. *Education Week, 1*, 13–15.
- Schochet, P., & Chiang, H. (2010). *Error rates in measuring teacher and school performance based on student test score gains*. Washington, DC: IES.
- Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration, 18*, 277–310.
- Song, H., & Xu, M. (2019). From external accountability to endogenous/potential-oriented development: The road to creation of a quality assurance system for university-based teacher preparation in China. *ECNU Review of Education, 2*, 137–165.
- Steffy, B. (1993). *The Kentucky education reform*. Lanham, MD: Rowman & Littlefield Education.
- Tennessee State Board of Education. (n.d.). *Teacher preparation report card*. Retrieved from <https://www.tn.gov/sbe/topic/teacher-preparation-report-card>
- The American Association of Colleges for Teaching Education. (2015, February 27). *AACTE Board Resolution on CAEP*. Retrieved November 21, 2016, from <https://aacte.org/news-room/press-releases-state-ments/488-aacte-board-resolution-on-caep>
- Walsh, K. (2006, March 16). *Teacher preparation: Coming up empty*. Retrieved August 16, 2016, from <https://edexcellence.net/publications/teachered.html>
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. New York, NY: Cambridge University Press.
- Whitford, B., & Jones, K. (2000). *Accountability, assessment and teacher commitment: Lessons from Kentucky's reform efforts*. Albany, NY: State University of New York Press.
- Zeichner, K. (2011). Assessing state and federal policies to evaluate the quality of teacher preparation programs. In P. Earley, D. Imig, & N. Michelli (Eds), *Teacher education policy in the United States: Issues and tensions in an era of evolving expectations* (pp. 75–105). New York, NY: Routledge.