

University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Faculty Publications: Department of Teaching,
Learning and Teacher Education

Department of Teaching, Learning and Teacher
Education

2013

Developing A System of Program Assessment Within Teacher Education: Lessons Learned

Kara Viesca

University of Nebraska-Lincoln, kara.viesca@unl.edu

Emilie Mitescu Reagan

Teachers College

Sarah Enterline

Boston College

Ann Marie Gleeson

Primary Source

Follow this and additional works at: <http://digitalcommons.unl.edu/teachlearnfacpub>

Viesca, Kara; Mitescu Reagan, Emilie; Enterline, Sarah; and Gleeson, Ann Marie, "Developing A System of Program Assessment Within Teacher Education: Lessons Learned" (2013). *Faculty Publications: Department of Teaching, Learning and Teacher Education*. 223.

<http://digitalcommons.unl.edu/teachlearnfacpub/223>

This Article is brought to you for free and open access by the Department of Teaching, Learning and Teacher Education at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications: Department of Teaching, Learning and Teacher Education by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Developing A System of Program Assessment Within Teacher Education: Lessons Learned

Kara Mitchell Viesca

Urban Community Teacher Education, University of Colorado–Denver

Emilie Mitescu Reagan

Teachers College

Sarah Enterline

Boston College

Ann Marie Gleeson

Primary Source

Coresponding Author: Kara Mitchell Viesca, PhD, University of Colorado–Denver, Urban Community Teacher Education, Campus Box 106, P.O. Box 173364, Denver, CO 80217, USA. E-mail: kara.viesca@ucdenver.edu

Our intention in this article is to present one institution’s efforts to take on program assessment and respond to calls for accountability. To do so, the teacher education program simultaneously sought to address the narrowly defined measures called for by policy makers and politicians, while at the same time broadening and expanding outcomes (such as student learning) and the missions, themes, and values of the teacher education program. The following describes the accountability movement in teacher education and then offers the lessons learned within one program developing a system of assessment within teacher education.

Developing a System of Program Assessment Within Teacher Education: Lessons Learned

As part of the broader conversation on accountability at the higher education institutional level, the teacher education accountability movement of the last 20 years has focused on defining teacher quality and effectiveness. Specifically, it emerged from the standards movement of the 1990s (Cochran-Smith, 2003) and is grounded in the assumption that raising educational outcomes for P–12 pupils requires higher standards and an improved teaching workforce (Wilson, Rozelle, & Mikeska, 2011). However, this movement has recently shifted from accountability of inputs (e.g., course offerings, internships) to outcomes (e.g., P–12 pupil achievement of program graduates) (Cuban, 2004). In other words, the current notions of teacher education accountability include both an emphasis on what happens during teacher education in terms of coursework and student teaching, and an increased emphasis on and examination of the demonstrated outcomes of graduates in terms of teaching practice and the success of graduates’ P–12 pupils. Although this chain of impact from teacher education to P–12 pupil achievement is viewed as logical by the general public, the media, many policy makers, and those outside of education, this logic is an overly simplistic and

necessarily problematic lens with which to view teacher education (Cochran-Smith, 2003; Earley, 2000; Engel, 2000).

British educational philosopher Gert Biesta (2009) provided a useful analysis of the current broader culture of educational accountability suggesting that, on the surface, the accountability movement that is focused on publicly offering assessment data is designed to enhance transparency, democracy, and relationships among families, teachers, and schools.

However, in practice, rather than increasing the voice of citizens, educational accountability has shifted the role of parents and pupils from citizens to economic consumers who have no control over educational accountability mechanisms. Further, current notions of education accountability are problematic for their narrow views of what is assessed and how it is assessed. This is due to the situation that has been created “where it is no longer the question what schools can do for their students, but what students can do for their schools” (Biesta, 2009, p. 656). Specifically, students are expected to score well on standardized exams to demonstrate school success.

It is not a huge leap to see this phenomenon applied to teacher education; certainly its application can be observed across the country. For example, certain teacher education programs will admit candidates into programs only after they have passed certain licensing tests required for teacher licensure and certification. However, substantial evidence suggests that these tests are not necessarily correlated with teacher effectiveness (Buddin & Zamorro, 2009; Cochran-Smith & Dudley-Marling, 2001; Goldhaber, 2007), and may create disparities based on race and ethnicity for teachers entering the profession (Cochran Smith & Dudley-Marling, 2001; Goldhaber & Hansen, 2010; Hood & Parker, 1991). It appears that rather than ensuring higher quality educational outcomes and opportunities for all teachers and pupils, the current method of accountability may be exacerbating and extending current inequities by limiting opportunities for teachers from varying race, class, gender, and linguistic backgrounds (Goldhaber & Hansen, 2010). Therefore, the push for accountability, based on the narrow definition given above, in education generally, as well as in teacher education specifically, is questionable because effective measures that accurately, validly, and reliably assess quality teaching and learning are few.

Despite such issues and critiques, it appears that current policies and initiatives regarding accountability in teacher education remain narrow in scope. Most alarming is the common belief that P–12 pupil achievement (mainly measured through standardized test scores) is the key—and often only relevant—outcome of teacher education. Examples include the Center for American Progress’s recent policy brief, *Measuring what Matters: A Stronger Accountability Model for Teacher Education* (Crowe, 2010). Among other “guiding principles,” Crowe recommended that all states have the same accountability system for teacher education and “every state’s teacher education program accountability system should include a teacher effectiveness measure that reports the extent to which program graduates help their K–12 students to learn” (p. 2). Similarly, at the federal level, teacher and teacher education accountability has shifted. Instead of calling for highly qualified teachers (No Child Left Behind, 2002), they instead call for data that demonstrate a causal link between teacher education programs and P–12 pupil achievement, as measured by standardized test scores. An example of these initiatives includes the recent introduction of the bipartisan Senate Bill 11462 (2011), *Growing Excellent Achievement Training Academies (GREAT) Teachers and Principals Act*. This act, though not yet passed and expected to be part of the larger debate over the eventual reauthorization of the Elementary and Secondary Education Act (ESEA), would support the development of non-university based teacher credentialing programs that tie graduation “to improving student academic achievement.” As stated in the bill, “programs that fail to produce great teachers or principals [as demonstrated by student achievement] will not be reauthorized.” In another recent initiative, Secretary of Education Arne Duncan described the Obama Administration’s Plan for Teacher Education Reform and Improvement (U.S. Department of Education, 2011) as holding

teacher preparation programs “to a clear standard of quality that includes but is not limited to their record of preparing and placing teachers who deliver results for P–12 students” (p. 2). Through this plan, federal funding would target programs with graduates who have pupils who either demonstrate large gains on standardized tests or perform well already (i.e., attend schools with higher socioeconomic status, lower urban populations, and limited racial and linguistic diversity). Together, these proposed policies and initiatives target pupil achievement (measured by standardized assessments) as the primary outcome of teacher education. Doing this fails to acknowledge many other relevant and complex factors impacting outcomes such as local context (Norman, 2010), issues with accurately measuring what pupils from non-native English speaking and other diverse backgrounds know and can do (Abedi & Lord, 2001; Basterra, Tumbull, & Solano-Flores, 2011; Dallmann-Jones, 2011), as well as the complexity of measuring gains on these tests (which require at least three years of testing in the same subject).

With these challenges in mind, the current demand for teacher education accountability that focuses solely on P–12 pupil academic achievement is downright alarming and fraught with problems (Cochran-Smith, Gleeson, & Mitchell, 2010). In sum, if one were to believe that it is appropriate to grant federal funding to a teacher education program only if its graduates’ pupils excel on standardized tests, then it is implied that (a) the standardized tests reliably and validly measure changes in pupil academic achievement over time; (b) each pupil’s achievement is accurately linked to the teacher who was responsible for teaching the material on the test; (c) this teacher teaches in a tested grade level and subject area in a traditional, public school; (d) this teacher, and his/her pupil data, are accurately tracked by the district and state education departments; (e) the teacher is accurately tracked by the teacher education program that conferred their degree; and (f) the teacher education program maintains accurate records on each graduate with regard to admissions applications, classes taken, teaching experiences, and preservice performance. These assumptions are, more often than not, violated at some point. Moreover, violations occur not due to a disagreement that these connections are unimportant, but rather because the resources (e.g., staff, technology, quality assurance protocols) needed to build such systems are scarce, and the impetus to draw connections across systems is not well-supported. Specifically, teacher education programs have little (if any) control over how standardized tests are developed, how pupils are linked to teachers, how such data are managed and linked to teachers, and how the data are linked to teacher preparation programs. Therefore, it is a complex and difficult task to create a direct connection between teacher preparation work and P–12 pupil learning (Cochran-Smith et al., 2010; Cochran-Smith & Fries, 2005; Kennedy, 1999; Schalock, Schalock, & Ayres, 2006; Wasley & McDiarmid, 2004; Wiens, 2012).

Along these lines, attempts to assess teacher education programs via the performance of graduates as well as eventual student achievement opens up a logistical minefield with regard to a few key details. These details include data access and disclosure agreements; the tracking of teachers and pupils; and forced reliance on data systems that were designed, maintained, and accessed by district and state workers and for-profit testing companies instead of by higher education or faculty/staff within teacher education programs. In other words, teacher education programs are held accountable to measures and assessments for which they have limited quality control mechanisms. That is, unless they commit themselves to developing their own internal systems of assessment that can build on the assessments that are under their control (i.e., course offerings, performance assessments, and other candidate, graduate, and pupil learning indicators), establishing this link is futile.

Furthermore, to add to the complexity, research that addresses programmatic accountability may not be well known to, or valued by, those who work in other fields. Additionally, there are often methodological or political differences that lead to disagreements about what the evidence actually means and what its implications are for policy and practice (Cochran-Smith et al., 2011). However, by addressing the narrow conceptions of accountability, including assessing teacher can-

didate, graduate, and P–12 pupil learning, and expanding on these notions to address programmatic values and beliefs and institutional mission, institutions can address internal and external accountability demands.

One Program’s Approach to Accountability

The teacher education program highlighted in this article is located in a mid-sized, private, liberal arts Jesuit Catholic University in a metropolitan area. The teacher education program prepares approximately 250 teachers annually in early childhood, elementary, and secondary education. It also prepares specialists in reading and special education at both the undergraduate and graduate levels, and has been consistently ranked by U.S. News and World Report as one of the top 20 schools of education in the country. The university’s philosophy aligns with the school of education’s dedication to working towards “the goals of social justice,” and the teacher education program’s overarching theme of promoting socially just teaching.

In 2003, the university was chosen by the Carnegie Corporation of New York as one of 11 institutions to receive a five-year grant as part of the Teachers for a New Era initiative. Specifically, the institution was tasked with improving teacher education in the university setting. As part of this endeavor, an “Evidence Team”¹ was established to examine the teacher education program and its impact on relevant outcomes. After reviewing research on teacher education, the Evidence Team developed a conceptual framework that describes the complex relationship among teacher candidates’ prior experiences and beliefs, their learning experiences, teaching practices, and beliefs while in and at the completion of the teacher preparation program, and their beliefs and practices once they become classroom teachers of record. Within this framework, the team situated these inter-related pieces within larger classroom, school, community, university, and accountability contexts. Ultimately, the framework suggested that these factors contribute in different ways to pupil learning and development, teacher retention, and teaching for social justice (Cochran- Smith, Reagan, & Shakman, 2009).

Working from this conceptual framework, the Evidence Team employed a “dialectical mixed methods” approach (Greene & Caracelli, 2003) to develop a portfolio of quantitative, qualitative, and mixed methods studies. Following Greene and Caracelli, Cochran- Smith, & The Boston College Evidence Team (2009) elaborated on this approach by stating, “[D]ifferent research designs and approaches are regarded as providing valuable, but always partial, perspectives on the topic under investigation, and the tensions created by studies’ differing assumptions and ways of knowing are regarded as generative of richer understandings rather than incompatible approaches” (p. 461). Specifically, the studies in the portfolio included a survey study; a set of longitudinal qualitative case studies investigating teacher candidates’ entering characteristics, their learning experiences in the teacher education program, teaching beliefs and practices, and their pupils’ learning; a comparison study of teaching practices and student outcomes of two pathways to teaching; an inquiry study examining teacher candidates’ approaches to inquiry-into-practice; an investigation of teacher candidates’ performance on a capstone performance assessment; and a cross-sectional, value-added assessment of standardized assessment scores from students of teachers who were graduates of the teacher education program.

This portfolio became the foundation of the assessment and accountability measures used by the teacher education program and allowed for faculty and staff to engage in critical evaluation of the program, thereby providing entry into the larger national accountability conversation as previously discussed. Integral to the work of the Evidence Team was the incorporation of programmatic documentation (mission, themes, curricula, etc.) and faculty input into the development of the empirical research studies, as well as in analysis of data and interpretation of the findings. For

example, survey and interview questions were vetted by program faculty prior to administration to ensure that the results would be valuable and relevant for the program. Additionally, the Evidence Team research focused on the teacher education program's overarching theme of promoting social justice. The program, therefore, had access to an overwhelming amount of useful evidence about the programs' candidates, graduates, their experiences and their teaching practices.

This data-rich environment allowed faculty to easily pursue national accreditation from the Teacher Education Accreditation Council (TEAC) because the program could make clear claims about the knowledge and skills of teacher candidates and graduates. From this research, the program faculty were able to support the claims that the programs' candidates and graduates (a) believed in and were committed to teaching for social justice, which is defined as improving the learning of all pupils and enhancing their life chances; (b) possessed subject matter knowledge, pedagogical knowledge, and pedagogical content knowledge, and were able to demonstrate this knowledge in practice; (c) were knowledgeable about and understood the relationships among culture, language, learning, and schooling; (d) developed and demonstrated in practice social justice orientations, commitments, and interpretive frameworks; (e) demonstrated commitment to learning across their professional lifespan and possessed knowledge of technology tools to do so; and (f) assessed and promoted all pupils' learning. In pursuing national accreditation, the teacher education program made accountability claims around and provided evidence to support teacher candidate, graduate and pupil learning, but also expanded its claims to assess key themes, missions and goals of the program.

Over time, the Evidence Team established a teacher education system of assessment from a rich, mixed methods approach that was relevant to both local (internal) interests and those of the accountability movement at the national level (external), including the pursuit of national accreditation (Enterline, 2011). Through this work, the team used multiple study assessment to more completely ascertain the impact and outcomes of the program. As a result, team members learned several lessons that would be valuable to other programs seeking to design and create similar systems of assessment.

Therefore, in this article, we, four members from the larger Evidence Team, will first describe three studies that were conducted from the same system of assessment, but offered varying perspectives on the same issue of relevance to the teacher education program. Using multiple studies from varying perspectives provided a more complete picture of the impact and outcomes of the teacher preparation program. These studies also illustrate overarching lessons learned from building and working within this system of teacher education assessment and accountability. Specifically, we argue that in program assessment and accountability, one type of evidence (e.g., quantitative) will never be as powerful as multiple forms of evidence. We also assert that teacher education programs should validly and reliably assess processes and outcomes of their programs. And finally, we argue that in doing what we assert is important in teacher preparation, specifically assessing what is important to both internal and external stakeholders from varying perspectives and approaches, unintended findings may arise. However, unflattering results from teacher education assessment can and should be embraced as well as viewed as important for programmatic improvement at multiple levels.

The following section first describes three studies that emanated from Evidence Team research, their results, and how these results speak to each other. Then, we will discuss in more detail our lessons learned, drawing on examples from these three studies. To be clear, these three studies were not the only three studies from the teacher education assessment system described above. However, together these three studies provide valuable information on the teacher education program's ability to meet some of its expressed goals. In addition, these three studies provide important insights into the aspects of a quality system of teacher education assessment.

A Multiple-Study Approach to Understanding Impacts and Outcomes

Building on the commitment toward social justice expressed by both the institution and the teacher education program, the following three studies examined the impact and outcomes of learning to teach for social justice through varied approaches and perspectives. This multiple-study approach offers many insights relevant to the program itself, teacher education generally, and approaches to teacher education assessment and accountability. In this section we will briefly discuss definitions of teaching for social justice. Next, each study is described individually. Lastly, we discuss what they collectively offer for understanding program impact and outcomes regarding learning to teach for social justice in this era of increased accountability that is narrowly directed at pupil test scores. These three studies will also be referenced in the next section of our paper, which is focused on what lessons we have learned through our efforts in teacher education assessment and accountability.

Defining Teaching for Social Justice

As Hytten and Bettez (2011) argued, it is difficult to be against the abstract idea of social justice as “we learn to pledge allegiance to a country that supposedly stands for ‘liberty and justice for all’ ” (p. 8). Yet, this idea of social justice as well as what it means to teach for it is not always clear or consistently defined (Kapustka, Howell, Clayton, & Thomas, 2009; Zeichner, 2006) and can “relate to a range of different practices and beliefs” (Agarwal, Epstein, Oppenheim, Ovler, & Sonu, 2010, p. 238). Therefore, even within the teacher education program’s system of accountability and assessment under investigation, there were multiple perspectives on what it means to teach for social justice. However, fundamentally, faculty and researchers agreed that teaching for social justice means improving the life chances of P–12 students through high quality educational opportunities. This perspective provides a baseline agreement, but does not necessarily encompass all notions of teaching for social justice, particularly those requiring students and teachers to critique and challenge dominant social structures and the myths that perpetuate inequity (Mitchell, 2012a, 2012b). It also does not necessarily encompass the perspective that social justice educators should affirm students’ cultural differences as assets and design instruction that builds on students’ experiential knowledge while also challenging societal inequities through leadership, advocacy, and organizing (Cochran-Smith, Reagan, Shakman, & The BC TNE Evidence Team, 2009; Cook-Sather, Cohen, & Alter, 2010; Zeichner, 2009). We recognize the challenge in coming to consensus on a definition of teaching for social justice, but we also see the accomplishment in at least creating a foundation of agreement through which this collaborative work was accomplished. The following sections illustrate some particular examples of how teaching for social justice within one teacher preparation program has been examined from various perspectives and approaches.

Beliefs About Teaching for Social Justice

Reagan (2011) examined beliefs about teaching for social justice, as measured by an established, psychometrically sound instrument, the Learning to Teach for Social Justice-Beliefs scale (Enterline, Cochran-Smith, Ludlow, & Mitescu, 2008; Ludlow, Enterline, & Cochran-Smith, 2008). In addition, this study analyzed reported experiences and perceptions that may be related to and influence beliefs about teaching for social justice, as assessed by open-ended, qualitatively oriented questions collected through program-specific entry and exit surveys (Ludlow et al., 2008).

Using the Rasch rating scale and multiple regression, this research analyzed longitudinal survey data from two cohorts of undergraduate teacher candidates (n = 134) who completed the teacher education program. This study investigated individuals in the aggregate, variability within and across cohorts, and change across time. In addition, this research sought to untangle and iden-

tify how reported experiences and perceptions before as well as during formal teacher education were related to beliefs about and commitment to teaching for social justice.

Findings suggested that teacher candidates' beliefs about teaching for social justice changed from the time they entered the teacher education program to the time they graduated. Specifically, most teacher candidates began the teacher preparation program with uncertainty and limited commitment to teaching for social justice. When these same teacher candidates graduated 4 years later, on average, their beliefs about teaching for social justice were more complex and nuanced. Reagan's work will be discussed further in the section regarding methodologically rigorous, valid, and reliable research.

Longitudinal Case Study of Learning to Teach for Social Justice

Through a 4-year longitudinal qualitative case study that investigated the impact and outcomes of teacher education on one teacher candidate/novice teacher, Mitchell (2011) critically assessed the explicit goals of the social justice-focused teacher education program. Additionally, institutionalized racism and linguisticism, or language-based discrimination (García, 2009; Phillipson, 1992), were interrogated for their role in the education of secondary bilingual learners and their teachers.

Mitchell (2011) utilized critical race theory (CRT) (Crenshaw, 1988; Ladson-Billings & Tate, 1995), a theory developed in response to the stalled advances of the civil rights era during the mid-1970s. As a theory, CRT was developed to expose the "ways that so-called race-neutral laws and policies perpetuate racial and/or ethnic and gender subordination" (Bernal, 2002, p. 108). Therefore, it was a powerful theoretical tool for assessing the outcomes and impacts of the teacher education program striving to prepare socially just teachers.

This research utilized longitudinal qualitative data that were collected over several years for a larger qualitative case study project. Stake (2000) called the type of analysis conducted in this study an "instrumental case study," in which a case is "examined mainly to provide insight into an issue or to redraw a generalization" (p. 437). Here the experiences of a teacher candidate/novice teacher over time were instrumental in illustrating various successes and failures of the systems of education that prepare teachers, support teachers, and within which teachers work and how those systems either perpetuate or disrupt institutional racism and linguisticism.

This study found weaknesses within the teacher education program due to a lack of cohesion across coursework and fieldwork. The findings challenged the program's notions of success: the case study participant was widely perceived as a successful graduate from the nationally recognized teacher education program under investigation; however, this in-depth analysis uncovered the ways racism and linguisticism were perpetuated through her teacher education experiences and into her classroom policy and practice. These difficult findings stand in contrast to the program's desired outcomes and demonstrate how such critical, in-depth assessments of teacher education can become powerful tools for program improvement. Mitchell's work will be discussed in greater depth below in the section regarding asking challenging questions and accepting unintended findings.

Social Justice in K–12 Pupil Learning

Gleeson (2011) conducted a third study to examine how one social justice-oriented teacher education program used qualitative and quantitative methods to assess teacher candidate and pupil learning in a rigorous and evidence-based manner that also reflected core values of the program. This assessment draws on the framework of "authentic intellectual work" (Newmann & Associates, 1996), which emphasizes the construction of knowledge, disciplined inquiry, and value beyond school, as a way to measure learning outcomes consistent with programmatic goals. This study explored the extent to which teacher candidates/program graduates engaged their pupils in

authentic intellectual work during their pre-service and first 2 years of teaching and their pupils' response to those learning opportunities.

Drawing on data gathered during a longitudinal qualitative case studies project on learning to teach, this study utilized a qualitatively driven, concurrent mixed methods approach (Creswell, Plano Clark, Gutmann, & Hanson, 2003; Morse & Niehaus, 2009) and cross-case study methodology (Stake, 2000) to examine the authentic intellectual work of 11 teacher candidates/program graduates and their pupils, during their pre-service and first 2 years of teaching. As part of the Teacher Assessment/Pupil Learning Protocol (TAPL), participants were interviewed about assessment and pupil learning in relation to specific samples of assessment tasks/assignments that they had used in their classroom and pupils' work on those assessments. Researchers then scored the assessment tasks and pupil work in accordance with a rubric based on the framework of "authentic intellectual work" (Newmann & Associates, 1996). Over 3 years, participants took part in five TAPL interviews ($n = 55$). Researchers scored one assessment from each interview ($n = 55$) and a random sample of pupil work ($n = 550$). Interpretive qualitative and statistical techniques were used to analyze the interview data and rubric scores, respectively.

These analyses were integrated to construct a nuanced description and understanding of the extent to which program graduates and their pupils engaged in authentic intellectual work and the conditions that influenced this work over time. The degree to which teachers enacted authentic intellectual work during the pre-service period and first two years of teaching varied and depended heavily on school context. This research addressed how factors such as mandated curricula and school culture constrained or supported this type of work. Further, this study demonstrated the value of mixed methods in program assessment and underscores the importance of using qualitative data to fully understand quantitative measures of teacher and pupil learning. By using the TAPL Protocol, this program gathered rigorous evidence to assess learning outcomes in a manner consistent with the program's core values and drew on this data to both support program accreditation and inform institutional changes. Gleeson's study will be discussed in further detail in the section focused on the value of mixed-methods approaches and versatile forms of evidence.

A Multiple-Study Overall View of Learning to Teach for Social Justice

In combination, these three studies offer a nuanced perspective of the impact and outcomes of the teacher education program and its expressed goal of preparing teachers to teach for social justice. Reagan's (2011) study demonstrated that over time the expressed commitment of undergraduates towards teaching for social justice does improve. This is promising and suggests some success on behalf of the teacher preparation program, particularly for undergraduates (the students under investigation). Gleeson's (2011) study also offered some promising results in terms of the ability of teacher candidates/novice teachers to provide authentic intellectual activities in their classrooms in support of high levels of K–12 pupil learning. Her work also provided insight into some of the structural and policy-related challenges teacher candidates/novice teachers face in creating and engaging in high quality assessment practices such as using mandated tests even when students are not yet prepared to succeed on them. Despite the promising results of Reagan's (2011) and Gleeson's (2011) work, Mitchell's (2011) study suggested there is likely more to do to accomplish the program goals of preparing teachers to teach for social justice. Although focused on just one teacher candidate/novice teacher, this work uncovered some serious issues in terms of program coherence that should be further investigated and addressed in order to truly prepare teachers to teach for social justice.

The results of these studies have impacted program improvement in the teacher preparation program in important ways. For instance, these results have influenced the maintenance and institutionalization of the survey system. In addition, fieldwork experiences have been enhanced for teacher candidates to systematically include their evaluations of student work, similar to the

research done with TAPL. Finally, consistent efforts are underway to improve opportunities around preparing teachers to work effectively with bilingual learners. The program and institution has taken the results of these studies as opportunities to learn and improve—efforts that should be at the heart of any teacher education system of assessment and accountability.

Lessons Learned

After years of engaging in teacher education assessment and accountability, the work of the Evidence Team offers some valuable lessons that should inform other programs and researchers engaged in similar work. The following sections will describe these lessons and refer back to the three studies described above as examples.

The Value of Mixed-Methods Approaches and Versatile Forms of Evidence

A teacher education program should assess what it values by incorporating diverse methodological and disciplinary perspectives. Successful accountability measures should thereby operate under the often complex assumption that no single outcome and no single research design can capture the full picture of a teacher education program. This assumption is essential to building a culture of evidence (Cochran-Smith et al., 2009) that acknowledges faculty and staff values and beliefs, and can respond to internal and external stakeholders. These assessments allow external stakeholders to better understand and use various methods to measure and track programmatic change within their own programs.

The three studies (Gleeson, 2011; Mitchell, 2011; Reagan, 2011) described in this article used different methods to assess the same teacher education program in ways that were consistent with program values, thereby reflecting the local context. Each study measured effectiveness in relation to teaching for social justice, a guiding principle for the program. These program assessments attempted to account for such things as critical thinking, authentic learning, sociocultural influences on teaching and learning, multiple evaluative measures, and learning over time.

Gleeson's (2011) study offered a particularly informative example of how multiple methods can be used to provide a more complex and nuanced understanding of teacher and student learning. Using the TAPL protocol, teachers in this study submitted assessments that they used in their classroom and their pupils' work on those assessments. They were then asked about the following topics in a semi-structured interview: the creation and implementation of the assessment; their learning goals for pupils and evaluation of pupil learning; how well the assessment worked and might be modified in the future; and what they considered "high," "medium," and "low" examples of pupil performance. Qualitative methods were used to examine teachers' thoughts about pupil learning. In the second part of the TAPL protocol, researchers used a rubric (RISER, 2001) based on Newmann and Associate's (1996) framework of "authentic intellectual work" to assign scores to assessments and student work.² Descriptive and inferential statistical analyses were conducted to describe the extent to which participants demonstrated authentic intellectual work. These quantitative data provided a measure to compare authentic intellectual work among and between participants and helped inform other qualitative data.

The use of both quantitative and qualitative data in the TAPL protocol was essential for understanding how program graduates considered and enacted authentic intellectual work in their classrooms. In many cases, the scores that teachers' received based on assessments and pupil work did not accurately reflect a teachers' understanding of or commitment to student learning. The qualitative analysis of the interview data illustrated that there was a complex set of factors that influenced teachers' abilities to engage their pupils in this type of work, many of which, such as high student absenteeism or standardized curriculum, were beyond teachers' direct control in their classroom. On

the contrary, there were also cases in which teachers who had relatively high pupil performance on the quantitative rubric exhibited questionable practices through the qualitative analysis. Using two measures provided a more accurate representation of teacher practice.

Each assessment provided valuable, yet partial, perspectives on the topic under investigation, and the tensions created by the results should be regarded as generative of a richer understanding rather than a result of incompatible approaches (Greene & Caracelli, 2003). Bridging the differences and weaknesses in terms of various methodologies adds to the challenge of trying to measure complex constructs, such as P–12 pupil achievement as it links to teacher education for social justice. However, teacher educators and researchers can capitalize on a variety of different research methodologies (quantitative, qualitative, and mixed-methods) to fulfill their assessment and accountability requirements. By thoughtfully doing so, a more comprehensive, nuanced, and complex picture may emerge regarding teacher education processes and outcomes. Further, when psychometrically strong assessments are utilized, teacher education assessment systems and practices can meet the demands of external stakeholders while not divorcing them from their internal values and beliefs.

Methodologically Rigorous, Valid, and Reliable Assessments

As described above, the teacher education program maintains a substantial commitment to preparing teachers to teach for social justice, which is ultimately defined as enhancing the life chances of pupils. Therefore, many of the program's accountability instruments and studies focus on teaching for social justice as an outcome (Cochran-Smith et al., 2009). Following Greene and Caracelli (2003) and the prior work of the Evidence Team, the teacher education program assumed that no one instrument or assessment could accurately assess teaching for social justice (Cochran-Smith et al., 2009). Therefore, it attempted to use a variety of methodologically rigorous and diverse studies to support its accreditation claims regarding preparing teachers to promote social justice.

A particular example of the development and use of a methodologically rigorous, valid, and reliable approach to measuring teacher candidates' beliefs regarding social justice was conducted by Reagan (2011) and builds off of the work of the Evidence Team. Reagan's study used a psychometrically sound instrument, the Learning to Teach for Social Justice-Beliefs (LTSJ-B) scale (Ludlow et al., 2008), to assess candidates' beliefs about teaching for social justice. Conclusions were based on their responses to program wide entry and exit surveys. The LTSJ-B scale was designed according to the principles of Rasch measurement (Ludlow et al., 2008), with which the research team hypothesized that individuals would differ in terms of their commitment to and beliefs about teaching for social justice. Therefore beliefs could be measured along a continuum of varying commitment to teaching for social justice. Furthermore, the team conceptualized specific practices of teaching for social justice that would be more or less difficult to endorse and would also fall along a continuum of statements (items). After multiple administrations, the resulting classical test theory (e.g., factor and reliability analyses) and Rasch item response theory analyses confirmed that the 12-item LTSJ-B scale was internally consistent, one-dimensional, invariant across administrations, and defined a conceptually coherent continuum of statements from easier to endorse to more difficult to endorse (Enterline, et al., 2008; Ludlow et al., 2008). Moreover, there is evidence to suggest that the teaching practices described in the LTSJ-B scale are understood in a similar manner across international settings (Ludlow, Enterline, et al., 2010).

Framed by the quantitative criticalist perspective (Stage, 2007), Reagan's (2011) study employed sophisticated statistical techniques to assess a conceptually complex construct, such as teacher candidates' beliefs about teaching for social justice (Ludlow et al., 2008). The study thus allowed for valid and reliable inferences about the teacher education program's graduates. Specifically, as described above, Reagan used the Rasch rating scale and multiple regression analyses to

examine longitudinal survey data from two intact cohorts of undergraduate teacher candidates (n = 134) who completed the same social justice-oriented teacher education program. In doing so, this study sought to untangle and identify whether reported experiences and perceptions before and during formal teacher education were related to beliefs about and commitment to teaching for social justice. Reagan found that teacher candidates' beliefs about teaching for social justice changed and varied from the time they entered the teacher education program to the time they graduated. Reagan's findings provide some evidence that this formal teacher education was not a "weak intervention" (Lortie, 1975). Rather, this study demonstrated that candidates' experiences develop and change over time over the course of the program.

The power of Reagan's (2011) study is that it added to the findings of other thoughtful and rigorous studies that examine the phenomenon of learning to teach for social justice in the same context using multiple methods and drawing on multiple perspectives and theoretical frames (i.e., Cochran-Smith et al., 2009; Gleeson, 2011; Mitchell, 2011; Reagan, 2011). Further, the study attempted to expand on the limited understanding of "outcomes" of teacher education by framing teacher candidates' beliefs—or interpretive frameworks through which individuals filter and mediate practice—as a legitimate and measurable outcome of teacher education. In this way, Reagan's work maintained substantial internal value for the teacher preparation program as well as contributed to larger, external conversations concerning program accountability.

A strong system of assessment is one that utilizes individual assessments that provide reliable and valid data. The results of each assessment, though they provide a partial view of the program in isolation, must be accurate. Reliable and valid assessments allow for dissemination and collaboration within the field outside of individual teacher education programs. This is because the scoring and coding processes are transparent and can be applied in different contexts. If a program has access to an array of assessments (such as the LTSJ-B scale), it can address specific and nuanced components of the curriculum, as well as candidates' and graduates' perspectives and experiences. However, the fundamental concept of teacher quality is highly political with no agreed-on definition (Wang, Lin, Spalding, Klecka, & Odell, 2011). What one program or policymaker may call success, another may not agree with. Accordingly, local perspectives, contexts, and ideologies should inform the accountability measures used by teacher education programs.

Openly Asking Challenging Questions and Accepting Unintended Findings

In our current political climate in which, "schools across the country are under attack by a broader neoliberal agenda that is severely limiting opportunities for equity and justice in education" (Pikower, 2011, p. 1106), this internal interest in teacher education for social justice, although inconsistent and questionable in some cases, may not be reflected in external accountability policies and practices. Despite the potential disconnect between the values and priorities of teacher education programs and the accountability measures being imposed externally, teacher education assessment work can and should focus on evaluating what is important internally to teacher education program as well as what is important to external stakeholders. However, when teacher preparation programs engage in this kind of assessment for dual and potentially multiple purposes, they will likely be asking ethical questions that may result in unpleasant and undesirable findings.

Juárez, Smith, and Hayes (2008) discussed the potential source of such undesirable findings, emphasizing that despite a focus on teaching for social justice and a claim to care about it in the majority of programs across the United States, many teachers are not prepared with the skills to adequately challenge societal inequities as well as work with diverse populations. During a time of such close scrutiny across the education establishment (Wilson et al., 2011), teacher educators engaged in program assessment may not be inclined to design and implement research that could lead to findings of programmatic inefficiencies. However, without such research important improvement

to programs, policies, and practices cannot be made and thus the cycle of perpetuating the inequitable status quo is likely to continue.

Grant and Agosto (2008) argued that, “theoretical perspectives allow us to ask different questions about social injustices and contribute to a deeper understanding of justice or injustice” (p. 179). Therefore, accountability and assessment in teacher education should include varying theoretical perspectives in order to dig into deeper issues of justice and injustice. Yet, utilizing critical perspectives in research and assessment may uncover unfavorable issues in teacher education and beyond. Despite the need to prove our value as teacher educators during a political context in which we face constant attack, such work from critical angles can insightfully illustrate programmatic issues. This improves the practice of preparing teachers to teach for social justice and ultimately will lead to a better understanding of and impact on pupils.

Mitchell’s (2011) study, conducted as an investigation on the effects of a teacher education program that attempted to prepare socially just teachers capable of providing high quality learning opportunities for multilingual learners, provides a particularly salient example. Using CRT to examine the teacher education experiences and outcomes of one teacher candidate/novice teacher over time, the findings of this study highlighted inconsistent coursework and a practicum placement that promoted linguisticist (García, 2009; Phillipson, 1992) and racist practices. Despite these issues, the teacher candidate/new teacher in the study critiqued various classroom concerns, individually challenged the issues of inequity she witnessed, and developed her skills as a teacher. However, even with her hard work, she did not leave her teacher education program equipped with the knowledge, skills, allies and experiences necessary to work with others to challenge systemic inequity, especially in terms of institutionalized racism and linguisticism. Instead, her experiences in the teacher education program were disjointed and did not offer sufficient support for the development of teaching practices that would have consistently and powerfully disrupted the inequitable status quo. She was forced to navigate within complex systems of power on her own, which primed her for individualistic work. She learned the skills to assist individual pupils rather than to be an agent and ally of systemic change.

Despite a focus on just one teacher candidate/new teacher, this longitudinal, in-depth research suggested that when instruction within teacher education programs is inconsistent and disjointed from the actual practice teacher candidate’s witness and emulate in their fieldwork, teachers will struggle with their capability to challenge the inequitable status quo (Sleeter, 2008). A coherent teacher education program is necessary to prepare teachers committed to social justice and capable of effectively teaching multilingual learners. Teacher candidates need to be equipped with strong models, partners, and contexts in which transformative work can actually happen that disrupts institutional practices perpetuating racism and linguisticism. Smith-Maddox and Solórzano (2002) suggested that, “Prospective teachers are provided with limited educational experiences that can help them understand the central role of race and racism in education” (pp. 66–67). Several other researchers and educators have demonstrated the complexity and difficulty in addressing race and racism in teacher education (e.g., Cochran-Smith, 2000; Sleeter, 2001). Additionally Ryan and Dixson (2006) asserted that, “teacher education must provide teacher candidates with opportunities for critical reflection on their racialized beliefs and experiences” (p. 176). They also argued that, “since whiteness has long enjoyed privilege in this society, work using CRT will likely bring new critiques on procedures that have heretofore been considered necessary or even best practice” (p. 181).

Utilizing CRT as a frame to investigate the outcomes and impacts of teacher education exposed that this teacher education program may need to reconsider some of its policies and practices, even those that have “heretofore been considered necessary or even best practice” (Ryan & Dixson, 2006, p. 181). Further, the in-depth, multi-year study of one teacher candidate/new teacher proved

incredibly valuable in terms of learning about the outcomes and effects of the program over time. This assessment of the actual outcome of this program's efforts to prepare teachers for social justice provided substantial evidence that continued critical examinations from varying perspectives are necessary to truly assess a teacher education program. In such assessments, troubling evidence may be collected regarding the shortfalls, problems, and serious issues in a program. However, such evidence is valuable and should be both taken seriously and utilized for substantial efforts in program improvement, particularly in programs that truly strive to prepare teachers to teach for social justice.

Conclusion

Despite living in a context in which teacher education assessment and accountability can be very challenging, we believe there are substantial opportunities for teacher educators to powerfully shape the teacher education accountability movement. Programs can do this by designing rigorous local systems of assessment that expand limited notions of accountability, utilize multiple measures, and provide valuable insights into internal and external conversations. Through the research and associated processes described above, as well as the comprehensive system of teacher education assessment that was created, we have learned three major lessons regarding such engagement with teacher education assessment and accountability. First, no one method of data collection or analysis will be as powerful or illustrative as multiple methods and perspectives. Second, in order to engage in meaningful teacher education program assessment, valid and reliable assessment tools and instruments must be utilized. Further, it is possible to create and validate instruments that measure complex constructs, such as teacher candidates' beliefs about teaching for social justice. It is important to do this because it is the only way to assess some of the more complex and nuanced aspects of teaching and learning. Finally, we learned that it is important to conduct and engage in teacher education research from multiple perspectives, particularly critical perspectives because they illustrate potential issues and problems. The findings of such research can provide essential insights for program improvement, particularly around the issues of inequity and social justice. Additionally, unflattering findings do not need to be ignored nor utilized only for internal program improvement. The findings of one teacher education program assessment can deeply inform the work of other programs around the nation.

In addition to the lessons we illustrated in depth above, there are further lessons learned from the development and implementation of this system of assessment for teacher education. First, although successful, mixed method systems of assessment take significant amounts of time and resources and are fraught with data management and attrition concerns (White & Arzi, 2005), they are powerful tools for measuring reliable and valid development over time (Shadish, Cook, & Campbell, 2002). Therefore, it is important for those interested in seeing strong teacher education accountability to also provide resources that ensure the development of effective, rigorous, and reliable systems of assessment.

Second, by working to ensure that all stakeholders (e.g., university, school and department administrators, program faculty, pupils, and accreditors) understand the complexity of the assessment system, decision-making can be based on empirical investigations of complex core values. Examples of said values include social justice, teacher and pupil learning, and challenging racism and linguisticism. Therefore, collaboration between multiple educational stakeholders and policy makers is important to ensure that the findings of various studies are both understood and helpfully utilized to improve teacher education policies and practices.

Third, the form of program assessment described in this article is possible only if faculty, staff, and researchers are committed to defining what is important to assess, developing and revis-

ing their instruments, gathering data and conducting the analyses, determining what should/can be changed within a curriculum/program of studies, devising and implementing those changes, and asking tough questions. The current contexts within the accountability movement, as well as developments within national accreditation, are by no means stagnant. Those involved with teacher education assessments must be actively shifting and engaging with relevant contexts and communities. However, as previously mentioned, this type of work requires substantial resources, which may not be easily available to all teacher education programs. Therefore, the commitment of teacher educators, national and local funders to creating rigorous and reliable systems of teacher education assessment is crucial.

The Evidence Team's work over the years has been used internally for program improvement and continues to inform local conversations regarding effective teacher preparation at the institution where this system of assessment exists. The Evidence Team's work has also been shared nationally through multiple publications and conference presentations, particularly as an example of how valuable mixed methods studies are in programmatic assessment and accountability (e.g., Cochran-Smith et al., 2009; Ludlow, Mitescu, et al., 2010; Ludlow et al., 2011). All of the lessons learned over the years can be incorporated into a single message to all who hold stake in the assessment and accountability movement within teacher education: The most effective system of national accountability for teacher education will allow for local programs to create similar systems of assessment that hold value both internally and externally. Within such a system, each program will have a different approach to teacher education assessment; yet, when effectively created, these approaches will substantially inform and speak to the work across teacher preparation contexts nationally. We urge all involved in teacher education assessment and accountability to embrace the lessons learned from this work in order to strive for a system that supports improvement within varying local contexts while also expanding the scope and notion of teacher education accountability nationally.

Notes:

1. Charged with developing evidence and assessing teacher education, Boston College's Teachers for a New Era Multidisciplinary Evidence Team developed an evidence portfolio with quantitative, qualitative, and mixed methods studies designed to examine impacts and relationships among teaching, learning, learning to teach, and social justice. The Evidence Team included Boston College faculty members and administrators: Marilyn Cochran-Smith (chair), Sarah Enterline, Alan Kafka, Fran Loftus, Larry Ludlow, Patrick McQuillan, Joseph Pedulla, and Gerald Pine; Jane Carter and Jeff Gilligan; and doctoral students, Joan Barnatt, Robert Baroz, Mac Cannady, Stephanie Chappe, Lisa D'Souza, Ann Marie Gleeson, Jiefang Hu, Cindy Jong, Kara Mitchell Viesca, Emilie Mitescu Reagan, Aubrey Scheopner, Karen Shakman, Yves Solomon Fernandez, and Diana Terrell.

2. The concept of authentic intellectual work places an emphasis on higher-order, deep, meaningful, and relevant learning experiences embedded within the classroom which reflects the teacher preparation program's themes of constructivism and social justice.

References

- Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, 14(3), 219–234. doi:10.1207/S15324818AME1403_2
- Agarwal, R., Epstein, S., Oppenheim, R., Oylar, C., & Sonu, D. (2010). From ideal to practice and back again: Beginning teachers teaching for social justice. *Journal of Teacher Education*, 61(3), 237–247. doi:10.1177/0022487109354521

- Basterra, M. d. R., Tumbull, E., & Solan-Flores, G. (2011). *Cultural validity in assessment: Addressing linguistic and cultural diversity*. New York, NY: Routledge.
- Bernal, D. D. (2002). Critical race theory, Latino critical theory, and critical raced-gendered epistemologies: Recognizing students of color as holders and creators of knowledge. *Qualitative Inquiry*, 8(1), 105–122.
- Biesta, G. (2009). Education between accountability and responsibility. In M. Simons, M. Olssen, & M. Peters (Eds.), *Re-reading education policies: A handbook studying the policy agenda of the 21st century* (pp. 650–666). Rotterdam, The Netherlands: Sense Publishers.
- Buddin, R., & Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, 66(2), 103–115. doi:10.1016/j.jue.2009.05.001
- Cochran-Smith, M. (2000). Blind vision: Unlearning racism in teacher education. *Harvard Educational Review*, 70(2), 157–190.
- Cochran-Smith, M. (2003). The unforgiving complexity of teaching: Avoiding simplicity in the age of accountability. *Journal of Teacher Education*, 54(3), 3–5. doi:10.1177/0022487102238653
- Cochran-Smith, M., Cannady, M., McEachern, K. P., Piazza, P., Power, C., & Ryan, A. (2011). Teachers' education, teaching practice, and retention: A cross-genre review of recent research. *Journal of Education*, 191(2), 19–31.
- Cochran-Smith, M., & Dudley-Marling, C. (2001). The flunk heard round the world. *Teaching Education*, 12(1), 49–63. doi:10.1080/10476210123057
- Cochran-Smith, M., & Fries, K. (2005). Researching teacher education in changing times: Paradigms and politics. In M. Cochran-Smith & K. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education* (pp. 69–110). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cochran-Smith, M., & The Boston College Evidence Team. (2009). Reculturing teacher education: Inquiry, evidence and action. *Journal of Teacher Education*, 60(5), 458–468. Lessons Learned 273
- Cochran-Smith, M., Gleeson, A. M., & Mitchell, K. (2010). Teacher education for social justice: What's pupil learning got to do with it? *Berkeley Review of Education*, 1(1), 35–61. Retrieved from <http://escholarship.org/uc/item/35v7b2rv>
- Cochran-Smith, M., Reagan, E., & Shakman, K. (2009). Just measures: Social justice as a teacher education outcome. *Teacher Education and Practice*, 22(3), 237–263.
- Cook-Sather, A., Cohen, J., & Alter, Z. (2010). Students leading the way toward social justice within and beyond the classroom. *Equity & Excellence in Education*, 43(2), 155–172. doi:10.1080/106656810.03719459
- Crenshaw, K. W. (1988). Race, reform, and retrenchment: Transformation and legitimation in antidiscrimination law. *Harvard Law Review*, 101(7), 1331–1387. doi:10.2307/1341398
- Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 209–240). Thousand Oaks, CA: SAGE Publications.
- Crowe, E. (2010). Measuring what matters: A stronger accountability model for teacher education. Center for American Progress. Retrieved from http://www.americanprogress.org/issues/2010/07/teacher_accountability.html
- Cuban, L. (2004). Looking in the rearview mirror at school accountability. In K. Sirotnik (Ed.), *Holding accountability accountable: What ought to matter in public education* (pp. 18–34). New York, NY: Teachers College Record.
- Dallmann-Jones, A. (2011, August 3). The real score of high-stakes testing: Damaged students and cheating professionals. [Web log content]. Retrieved from <http://ednotesonline.blogspot.com/2011/08/real-score-of-high-stakes-testing-kids.html>
- Earley, P. (2000). Finding the culprit: Federal policy and teacher education. *Educational Policy*, 14(1), 25–39. doi:10.1177/0895904800014001004
- Engel, M. (2000). *The struggle for control of public education: Market ideology vs. democratic values*. Philadelphia, PA: Temple University Press.
- Enterline, S. (2011). Capitalizing on quantitative, qualitative, and mixed-methods research in program assessment. Paper presented at the American Educational Research Association National Conference, April, 2011, New Orleans, LA.
- Enterline, S., Cochran-Smith, M., Ludlow, L., & Mitescu, E. (2008). Learning to teach for social justice: Measuring changes in the beliefs of teacher candidates. *The New Educator*, 4, 1–24.
- García, O. (2009). *Bilingual education in the 21st century: A global perspective*. Malden, MA: Wiley-Blackwell.
- Gleeson, A. M. (2011). Using qualitative and quantitative methods to assess learning outcomes in a social justice-oriented teacher education program. Paper presented at the American Educational Research Association Annual Conference, New Orleans, LA.
- Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources*, 42(4), 765–794. Retrieved from <http://proxy.bc.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=27772640&site=ehost-live>
- Goldhaber, D., & Hansen, M. (2010). Race, gender, and teacher testing: How informative a tool is teacher licensure testing? *American Educational Research Journal*, 47(1), 218–251. doi:10.3102/0002831209348970
- Grant, C., & Agosto, V. (2008). Teacher capacity and social justice in teacher education. In M. Cochran-Smith, S. Feiman Nemser, J. McIntyre, & K. Demers (Eds.), *Handbook of research on teacher education: Enduring questions in changing contexts* (3rd ed.). Philadelphia, PA: Taylor and Francis.

- Greene, J., & Caracelli, V. (2003). Making paradigmatic sense of mixed methods practice. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 91–110). Thousand Oaks, CA: Sage.
- Hood, S., & Parker, L. (1991). Minorities, teacher testing, and recent U.S. Supreme Court rulings: A regressive step. *Teachers College Record*, 92, 603–616.
- Hyttén, K., & Bettez, S. C. (2011). Understanding education for social justice. *Educational Foundations*, Winter-Spring, 7–24. 274 K. M. Viesca et al.
- Juárez, B. G., Smith, D. T., & Hayes, C. (2008). Social justice means just us white people: The diversity paradox in teacher education. *Democracy and Education*, 17(3), 20–25.
- Kapustka, K. M., Howell, P., Clayton, C. D., & Thomas, S. (2009). Social justice in teacher education: A qualitative content analysis of NCATE conceptual frameworks. *Equity & Excellence in Education*, 42(4), 489–505. doi:10.1080/10665680903260101
- Kennedy, M. (1999). The problem of evidence in teacher education. In R. Roth (Ed.), *The role of the university in the preparation of teachers* (pp. 87–107). Philadelphia, PA: Falmer Press.
- Ladson-Billings, G., & Tate, W. F. (1995). Toward a critical race theory of education. *Teachers College Record*, 97(1), 47–68.
- Lortie, D. (1975). *Schoolteacher: A sociological study*. Chicago: University of Chicago Press.
- Ludlow, L. H., Enterline, S., & Cochran-Smith, M. (2008). Learning to teach for social justice: An application of Rasch measurement principles. *Measurement and Evaluation in Counseling and Development*, 20, 194–214.
- Ludlow, L., Pedulla, J., Reagan, E., Enterline, S., Cannady, M., & Chappe, S. (2011). Considerations in conducting longitudinal research. *Education Policy and Analysis Archives*, 19(11), 1–19.
- Ludlow, L., Enterline, S., O'Leary, M., Ell, F., Bonilla, V., & Cochran-Smith, M. (2010). Learning to teach for social justice beliefs: An international construct invariance study. Paper presented at the American Educational Research Association National Conference, April 2010, and at the Teacher Education Forum of Aotearoa New Zealand, October 2010.
- Ludlow, L., Mitescu, E., Pedulla, J., Cochran-Smith, M., Cannady, M., Enterline, S., & Chappe, S. (2010). An accountability model of initial teacher education. *Journal of Education for Teaching*, 36(4), 353–368.
- Ludlow, L., Pedulla, J., Enterline, S., Cochran-Smith, M., Loftus, F., Salomon-Fernandez, Y., & Mitescu, E. (2008). From students to teachers: Using surveys to build a culture of evidence and inquiry. *European Journal of Teacher Education*, 31(4), 1–19.
- Mitchell, K. (2011). Uncovering troubling issues of institutionalized racism and linguicism in teacher preparation and beyond. Paper presented at the American Educational Research Association Annual Conference, New Orleans, LA.
- Mitchell, K. (2012a). Race, difference, meritocracy, and English: Majoritarian stories in the education of secondary multilingual learners. *Race Ethnicity and Education*. doi:10.1080/13613324.2011.645569
- Mitchell, K. (2012b). English is not ALL that matters in the education of secondary multilingual learners and their teachers. *International Journal of Multicultural Education*, 14(1), 1–21.
- Morse, J. M., & Niehaus, L. (2009). *Mixed method design: Principles and procedures*. Walnut Creek, CA: Left Coast Press.
- Newmann, F. M., & Associates. (1996). *Authentic achievement: Restructuring schools for intellectual quality*. San Francisco, CA: Jossey-Bass Publishers.
- No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).
- Norman, A. D. (2010). Assessing accomplished teaching: Good strides, great challenges. *Theory Into Practice*, 49(3), 203–212. doi:10.1080/00405841.2010.487755
- Phillipson, R. (1992). *Linguistic imperialism*. Oxford, UK: Oxford University Press.
- Picower, B. (2011). Resisting compliance: Learning to teach for social justice in a neoliberal context. *Teachers College Record*, 113(5), 1105–1134.
- Reagan, E. (2011). Examining the relationships among undergraduate teacher candidates' experiences, perceptions, and beliefs about teaching for social justice. Paper presented at American Educational Research Association Annual Meeting, New Orleans, LA.
- Research Institute on Secondary Education Reform (RISER). (2001). Standards and scoring criteria for assessment tasks and student performance. Madison, WI: Wisconsin Center for Education Research, University of Wisconsin.
- Ryan, C. L., & Dixon, A. D. (2006). Rethinking pedagogy to re-center race: Some reflections. *Language Arts*, 84(2), 175–183. *Lessons Learned* 275
- Schalock, D., Schalock, M., & Ayres, R. (2006). Scaling up research in teacher education: New demands on theory, measurement, and design. *Journal of Teacher Education*, 57(2), 102–119. doi: 10.1177/0022487105285615
- Senate Bill 11462. (2011). *Growing Excellent Achievement Training Academies (GREAT) Teachers and Principals Act*.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.
- Sleeter, C. E. (2001). Preparing teachers for culturally diverse schools: Research and the overwhelming presence of whiteness. *Journal of Teacher Education*, 52(2), 94–106. doi:10.1177/0022487101052002002
- Sleeter, C. E. (2008). Preparing white teachers for diverse students. In M. Cochran-Smith, S. Feiman- Nemser, D. J. McIntyre, & K. E. Demers (Eds.), *Handbook of research on teacher education: Enduring questions in changing contexts* (Vol. 3, pp. 559–582). New York, NY: Routledge/Taylor & Francis Group and The Association of Teacher Educators.
- Smith-Maddox, R., & Solórzano, D. G. (2002). Critical race theory, Paulo Freire's problem posing method, and case study

- research to confront race and racism in education. *Qualitative Inquiry*, 8(1), 66–84.
- Stage, F. K. (Ed.). (2007). Using quantitative data to answer critical questions. In: *New directions in educational research*. San Francisco: Jossey-Bass.
- Stake, R. E. (2000). Case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 435–454). Thousand Oaks, CA: Sage Publications.
- U.S. Department of Education. (2011). Our future, our teachers: The Obama administration's plan for teacher education reform and improvement. Retrieved from <http://www.ed.gov/sites/default/files/our-future-our-teachers.pdf>
- Wang, J., Lin, E., Spalding, E., Klecka, C. L., & Odell, S. J. (2011). Quality teaching and teacher education: A kaleidoscope of notions. *Journal of Teacher Education*, 62(4), 331–338. doi:10.1177/0022487111409551
- Wasley, P., & McDiarmid, G. W. (2004). Connecting the assessment of new teachers to pupil learning and to teacher preparation: National commission on teaching and America's future national summit on high quality teacher preparation. Austin, TX: National Commission on Teaching and America's Future.
- White, R. T., & Arzi, H. J. (2005). Longitudinal studies: Designs, validity, practicality, and value. *Research in Science Education*, 35, 137–149. doi:10.1007/s11165-004-3437-y
- Wiens, P. D. (2012). The missing link: Research on teacher education. *Action in Teacher Education*, 34(3), 249–261. doi:10.1080/01626620.2012.69401
- Wilson, S. M., Rozelle, J. J., & Mikeska, J. N. (2011). Cacophony or embarrassment of riches: Building a system of support for quality teaching. *Journal of Teacher Education*, 62(4), 383–394. doi:10.1177/002248711140941
- 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, 326–340.
- Zeichner, K. (2009) *Teacher education and the struggle for social justice*. New York, NY: Routledge Publications.

