

Seven Legitimate Apprehensions About Evaluating Teacher Education Programs and Seven “Beyond Excuses” Imperatives

AUDREY AMREIN-BEARDSLEY

Arizona State University

JOSHUA H. BARNETT

National Institute for Excellence in Teaching

TIRUPALAVANAM G. GANESH

Arizona State University

Background: *Via the reauthorization of the Higher Education Act (HEA), stronger accountability proponents are now knocking on the doors of the colleges of education that prepare teachers and, many argue, prepare teachers ineffectively. This is raising questions about how effective and necessary teacher education programs indeed are. While research continues to evidence that teachers have a large impact on student achievement, the examination of teacher education programs is a rational backward mapping of understanding how teachers impact students. Nonetheless, whether and how evaluations of teacher education programs should be conducted is yet another hotly debated issue in the profession.*

Purpose: *The purpose of this project is to describe how one of the largest teacher education programs in the nation has taken a lead position toward evaluating itself, and has begun to take responsibility for its impact on the public school system. This research also presents the process of establishing a self-evaluation initiative across the state of Arizona and provides a roadmap for how other colleges and universities might begin a similar process.*

Setting and Participants: *This work focuses on the Teacher Preparation Research and Evaluation Project (T-PREP) that spawned via the collaborative efforts among the deans and representative faculty from Arizona State University (ASU), Northern Arizona University (NAU), and the University of Arizona (UofA). The colleges of education located*

within each respective university are the colleges that train the vast majority of educators in the state of Arizona. Participants also included other key stakeholders in the state of Arizona, including the deans and representative faculty from the aforementioned colleges of education, leaders representing the Arizona Department of Education (ADE), and other key leaders and constituents involved in the state's education system (e.g., the state's union and school board leaders and representatives).

Research Design: *This serves as a case study example of how others might conduct such self-examinations at the collaborative and the institutional level, as well as more local levels.*

Conclusions: *This work resulted in a set of seven "beyond excuses" imperatives that participants involved in the T-PREP consortium developed and participants at the local level carried forward. The seven key imperatives are important for other colleges of education to consider as they too embark on pathways toward examining their teacher education programs and using evaluation results in both formative and summative ways.*

INTRODUCTION

Over the previous four decades, U.S. educational policymakers have trekked down a previously unprecedented accountability path. They started with the minimum competency movement in the 1970s (Bracey, 1995; Heubert & Hauser, 1999; Kreitzer, Madaus, & Haney, 1989) and continued with the proposition that our students may not be world class in the release of *A Nation at Risk* in the 1980s (U.S. Department of Education, 1983). The accountability trajectory seemed to have peaked with the passage of No Child Left Behind (NCLB, 2002), but with the congressional reauthorization of NCLB, Race to the Top, and Secretary of Education Arne Duncan's recent waiver system excusing states from not meeting 100% proficiency targets if they agree to attach even more consequences to educational outputs, our nation's policymakers continue their push for stronger accountability (see also R. Rothstein, 2011).

In higher education, teacher education programs have also been led down a similar pressure-by-policy path. In 2007, after the passage of NCLB, came the federal reauthorization of the Higher Education Act of 1965 (HEA, 2007). This placed increased responsibilities on teacher education programs, charging the leaders and the teacher educators within them to be held more accountable for their impact on student learning and achievement in PreK-12 public schools (Cochran-Smith, 2009; Goodwin, 2009; Ludlow et al., 2010). While this was indeed a high stakes in higher education movement that began during President George H. W. Bush's administration, and one that has been iterated via state-based policies since (Peck, Galluci, & Sloan, 2010), now more than ever many argue that teacher education in the United States is in jeopardy (Zeichner, 2010), although others note that our teacher education system is perhaps performing as well as it ever has (Glass, 2008).

Even with the legislative changes, prior to 2007, many if not most administrators and teacher educators have held themselves accountable for accreditation purposes, for example, via the National Council for Accreditation of Teacher Education (NCATE) and the Teacher Education Accreditation Council (TEAC). Like with the passage of NCLB, the 2007 reauthorization of HEA increased the pressure, mandating that teacher education programs were also to be held accountable beyond just accreditation. Teacher education programs were to be subjected to external ranking mechanisms and state report cards, where initial teacher certification (ITC) graduates' licensure test scores and their students' test scores would be used to measure teacher education program quality, or a lack thereof (Cochran-Smith, 2001, 2004, 2005, 2009; Cochran-Smith & Fries, 2001; Darling-Hammond, 2006a, 2006b; Goodwin, 2009; Hamel & Merz, 2005; Ludlow et al., 2010; Peck et al., 2010).

As an example of what was occurring during this transition, a research team (different from those discussed here) affiliated with the Value Added Assessment of Teacher Preparation Project (VAA-TPP) at Louisiana State University (LSU) constructed a longitudinal database, which connected K-12 students to their teachers in core content areas and was then connected back to the university where the teachers received their credentials (Noell & Burns, 2006; Noell, Porter, & Patt, 2007). The combination of the teachers' students' scores was then combined and used to determine how effectively the state's teacher education programs prepared teachers to improve student achievement scores. This work continues today, appropriately, at least in theory.

To those who insist that teacher education has at least something to do with teacher quality (Darling-Hammond, 2006a, 2006b; Darling-Hammond & Sykes, 2003; Shulman, 1988; Wilson, Floden, & Ferrini-Mundy, 2002), this new role for teacher education programs, to better evaluate what they do, is rational and reasonable. It is sensible as well that both the program leaders and the teacher educators who actually educate the future teachers within such programs be involved in this work, as they are indeed a vested group with a collective responsibility to determine if what they are doing is high quality, meaningful, and impactful. This is important, at least for internal purposes if nothing more. Notwithstanding, this need for empirical reflection is also externally kindled as outsiders continue to argue that "very little is known about if and how teacher education affects practice" (Good et al., 2006, p. 411; see also Cochran-Smith, Feiman-Nemser, McIntyre, & Demers, 2008; Harris & Sass, 2007). This ongoing deliberation reinforces the urgency to determine what it is teacher education programs are doing well, and where they are most in need of reform and reculturation (Cochran-Smith, 2009).

In addition, as researchers continue investigating the connection between teacher education programs and teacher performance in the classroom, educational policymakers continue to fundamentally question whether teacher education is solvent, or a broken down bureaucratic system that “needs to be turned upside down” (Education Digest, 2011, p. 9). They question whether teaching requires formal professional training, specifically in pedagogy; whether applied experiences in the classroom matter; and, most importantly, whether traditional versus alternative teacher educators in fact graduate teachers who effectively promote student learning and achievement (Chingos & Peterson, 2011; Cochran-Smith, 2001, 2009; Cochran-Smith & Fries, 2001; Darling-Hammond, 2006a, 2006b; Darling-Hammond & Sykes, 2003; Ludlow et al., 2010; Peck et al., 2010). Lacking proof based on rigorous research and empirical evidence (Cochran-Smith, 2004, 2009), policymakers continue to promote a technicist view of teacher education, and advance its replacement via nontraditional, alternative, and even for-profit pathways (Zeichner, 2010).

It is our contention in this manuscript that, given the current situation, it is vital for teacher education programs, including both teacher education leaders and faculty members, to engage and hold themselves accountable for purposes beyond accreditation. They need to prove that, via their teacher education programs, they are preparing measurably effective teachers (see also Barnett & Amrein-Beardsley, 2011). They need to do this particularly if they are to save themselves from potential elimination and replacement (Cochran-Smith, 2009; Goodwin, 2009; Harris & Sass, 2007; Ludlow et al., 2010).

In this manuscript, we first offer a review of the current state of teacher education in the United States, a review of what empirical actions have been taken thus far and some respective results, a discussion of the traditional and nontraditional methods typically used in such research, and a conversation about the conceptual frameworks often structuring this work. Second, we propose a “beyond excuses” framework for conducting teacher education evaluations. The rationale for the “beyond excuses” moniker is that federal policies have influenced and continue to influence the debate about whether teacher educators are responsible for their graduates and, more so, the extent to which their graduates impact student learning and achievement. Additionally, future legislation will likely include more sanctions or ranking mechanisms to help reduce and understand teacher education data, quality indicators, and the “value” teacher educators “add” to this production function, so being proactive now is crucial. Within this section, we also describe evaluative efforts of one consortium and one college as they, like others, have spent nearly

five years negotiating through these research processes and dilemmas, in “fashion[s] responsive to local values and concerns while also meeting state requirements” (Peck et al., 2010, p. 451; see also Cochran-Smith, 2009). Details about challenges, paradigm shifts, overcoming complexities, data use, and data informed reforms are also presented.

SEVEN LEGITIMATE APPREHENSIONS ABOUT EVALUATING TEACHER EDUCATION PROGRAMS

Approximately 40 years since Coleman and his colleagues (1966) posited that schools and teachers have little to do with what students learn in school, the educational research community has come to a consensus that teachers do in fact cause increases, and probably the most significant increases, in student achievement of all education-related variables (Berry, Fuller, & Reeves, 2007; Boyd et al., 2006; Chingos & Peterson, 2011; Cochran-Smith, 2004, 2005; Darling-Hammond & Sykes, 2003; Yinger & Hendricks-Lee, 2000). Teachers impact student achievement, and this model is illustrated in Figure 1.

Figure 1. Causality model illustrating teacher effects on student learning and achievement

$$Y_{\text{Teacher}} \rightarrow Z_{\text{Student1}}, Z_{\text{Student2}}, Z_{\text{Student3}}, \dots$$

But, the current accountability debate is no longer limited to whether and what impact teachers have on students in classrooms. The discussion has moved to how teacher education programs influence student performance. There is now an additional variable in the aforementioned trajectory—the teacher education program. Teacher educators and leaders must now investigate how well their programs prepare teachers and how well their ITC graduates promote student learning and achievement in schools. This model is illustrated in Figure 2.

Figure 2. Causality model illustrating purported teacher education effects on teacher effects on student learning and achievement

$$X_{\text{TeacherEducationProgram}} \rightarrow Y_{\text{Teacher}} \rightarrow Z_{\text{Student1}}, Z_{\text{Student2}}, Z_{\text{Student3}}, \dots$$

There are, at present, three units of analysis to link empirically and causally, but few if any researchers have developed compelling or appropriate (Noell & Burns, 2006; Noell et al., 2007) methods to examine how much of a teacher’s impact on student learning can be attributed to the teacher education unit. Such an empirical undertaking is reasonably and rightfully complicated. Although some teacher education researchers and evaluators are making progress, little has still been done to help

others satisfactorily explore this relationship, particularly at local levels (Boyd et al., 2006; Darling-Hammond, 2006a; Hamel & Merz, 2005; Harris & Sass, 2007; Russell & Wineburg, 2007). This is largely due to seven key apprehensions that contaminate such empirical investigations.

Apprehension #1: Model Unidimensionality

The model posed is inappropriately one-dimensional. More than 50% of college graduates attend more than one higher education institution before receiving a bachelor's degree (Ewell, Schild, & Paulson, 2003), and approximately 60% of teacher education occurs in general liberal arts and sciences, and other academic departments outside of teacher education. There are many more variables that contribute to teachers' knowledge by the time they graduate than just the teacher education program. Therefore, when evaluating teacher education, evaluators need to isolate the impact universities, or other colleges in which students are prepared, might have from the teacher education program itself (Anrig, 1986; Darling-Hammond & Sykes, 2003).

Apprehension #2: Self-Selection Into the Profession

The implied assumptions of the aforementioned linear formula are overly simplistic given the nonrandomness of the teacher candidate population. The types of students who enter teacher education programs and the personality characteristics they bring with them present another challenge. Self-selection, a traditional measurement threat to internal validity, occurs when groups of people at the focus of empirical research are distinctly different from the group(s) to which they are compared. If teacher candidates who enroll in a traditional teacher education program are arguably different from teacher candidates who enroll in an alternative program, and both groups are compared once they become teachers, one group might have a distinct and unfair advantage over the other. This difference may occur not because they are better teachers or were better prepared by either teacher education program, but because of the personal characteristics they brought with them to the profession. What cannot be overlooked, controlled for, or dismissed from these comparative investigations are teachers' enduring qualities—whether they are caring, dedicated, motivated, sensitive, respectful, etc., as these characteristics are positively related to teacher effectiveness (Boyd et al., 2006; Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2007; Harris & Sass, 2007; Shulman, 1988; Wenglinsky, 2002).

Apprehension #3: Nonrandom Distribution of Teachers

Teachers are nonrandomly distributed into schools after graduation as well. The type of teacher education program from which a student graduates is highly correlated with the type and location of the school in which the teacher enters the profession (Good et al., 2006; Harris & Sass, 2007; Rivkin, 2007; Wineburg, 2006), especially given the geographic proximity of the teacher education program to surrounding school districts and the types of schools in which student teachers are placed. This presents another challenge. If a certain teacher education program is located in a relatively affluent area, and if ITC graduates become teachers in its surrounding schools, they will have a distinct and unfair advantage over ITC graduates from the same or other programs who teach elsewhere, possibly in high-needs schools. Because of the nonrandom distribution of teachers, teachers who choose to teach in less challenging schools are sometimes falsely given credit for having more success with their students than teachers in more challenging schools, simply because of the type of students enrolled in the schools in which teachers take positions (Newton, Darling-Hammond, Haertel, & Thomas, 2010). Without randomly distributing teachers across schools, comparison groups will never be adequately equivalent, as implied in this model, to warrant valid assertions about teacher education quality (Boyd et al., 2006; Good et al., 2006). It should be noted, however, that whether the use of students' pretest scores and other covariates can account or control for such inter- and intra-classroom variations is still being debated and remains highly uncertain (Ballou, Sanders, & Wright, 2004; Capitol Hill Briefing, 2011; Koedel & Betts, 2010; Kupermintz, 2003; McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004; J. Rothstein, 2009; Tekwe et al., 2004).

Apprehension #4: Nonrandom Placement of Students Into Classrooms

Students are also not randomly placed into classrooms. Sometimes the oft-praised "best" teachers are more likely to have some of the brightest students in their classes because of students who self-select into these classes, parents who assertively request certain teachers for their children, and other local or ability-tracking placement policies and procedures (Monk, 1987). On the flip side, sometimes the presumed "best" teachers are assigned some of the most difficult-to-teach students because school administrators believe that high-quality teachers will have the greatest impact on the students who need them most (Clotfelter, Ladd, & Vigdor, 2007; Rivkin, 2007). Students' innate abilities and motivation levels bias even the most basic examinations in which researchers attempt to link

teachers with student learning (Newton et al., 2010; Harris & Sass, 2007; Rivkin, 2007). Without randomly assigning students to classes, teachers' classes will also never be adequately equivalent, again as assumed in this model. However, the degree to which such systematic errors, often considered measurement biases, impact value-added output is yet highly unsettled (Ballou et al., 2004; Capitol Hill Briefing, 2011; Koedel & Betts, 2010; Kupermintz, 2003; McCaffrey et al., 2004; J. Rothstein, 2009; Tekwe et al., 2004).

Apprehension #5: Post-Graduation Impact Variables

A student's performance is also empirically compounded by what teachers learn "on the job" post-graduation via professional development (see, for example, Greenleaf et al., 2011). If researchers are to measure the impact of a teacher education program using student achievement, and ITC graduates have received professional development, mentoring, and enrichment opportunities post-graduation, researchers might deliberate whether it is feasible to disentangle the impact that professional development, versus teacher education, has on teacher quality and students' learning over time. ITC graduates' opportunities to learn on the job, and the extent to which they take advantage of such opportunities, introduces yet another source of construct irrelevant variance (CIV) into, what seemed to be, the conceptually simple relational formula presented earlier (Good et al., 2006; Harris & Sass, 2007; Rivkin, 2007; Yinger, Daniel, & Lawton, 2007). CIV is generally prevalent when a test measures too many variables, including extraneous and uncontrolled variables that ultimately impact test outcomes and test-based inferences (Haladyna & Downing, 2004).

For instance, a report by the Education Commission of the States (Kaufman, 2007) found that teacher education programs across the United States have been found to be inconsistent. Around 30 states and territories were identified as having teacher education programs as defined by the state in statute or code, or by the state's department of education. Yet many school districts may also implement their own programs without state-level approval. Furthermore, there are qualitative differences in these teacher induction programs that make this issue more complex, particularly when trying to categorize programs for analyses that require reductionistic classifications.

Apprehension #6: Construct-Irrelevant Variance

Other sources of CIV need to be considered as well. These include whether teacher effectiveness can be appropriately assessed if teachers

(a) teach in multigrade classrooms, (b) team teach with other more or less effective teachers, (c) teach smaller classes as correlated with student achievement (Clotfelter et al., 2007), and (d) have access to different resources and technologies. For students, difficulties with such inquiries might happen because they (a) are English Language Learners (ELLs), have Individualized Education Programs (IEP), or are supported by special education teachers or aides whose competencies may vary; (b) switch schools or teachers mid-year; or (c) take more than one class in a certain subject area simultaneously or within the same school year. These are the issues currently plaguing the value-added analyses being conducted across the United States (Au, 2010; Haertel, 2011; Harris, 2011; Hill, Kapitula, & Umland, 2011; Newton et al., 2010; Papay, 2010; J. Rothstein, 2009). Ultimately, adding the examination of teacher education effects (see Figure 2 above) in addition to just teacher effects (see Figure 1 above) will, without a doubt, exacerbate these problems further.

Apprehension #7: Overreliance on Students' Large-Scale Standardized Test Scores

This model is also problematic because it is built almost entirely on students' standardized test scores as indicators of teacher, and now teacher education, quality (Baker et al., 2010). Students' standardized test scores, usually aggregated at the classroom, school, district, and state levels, are being used as the main, and too often only, measure of student learning and achievement (Noell & Burns, 2006; Noell et al., 2007). While gauging the quantifiable effects of nearly everything measurable is becoming the norm (Cochran-Smith, 2001, 2004, 2005, 2009; Zeichner, 2010), such practices contradict what all professional organizations on educational and psychological measurement recommend (AERA, APA, NCME, 1999).

There are some who ignore these challenges, however. Predominantly, value-added researchers, who continue to promote and advance their proclaimed, more sophisticated, primarily econometric models, too often minimize the problems and issues with conducting the research presented herein (Chingos & Peterson, 2011; Harris, 2011; LeClaire, 2011; Sparks, 2011). Their models are based on extraordinary assumptions (Amrein-Beardsley, 2012; Scherrer, 2011) that do not adequately address the aforementioned threats to validity, sources of CIV, or all of the other complexities inherent in quasi-experimental studies. Even the most sophisticated model will never hold up if valid inferences are to be made in the ways theorized, and never will this research be done without accepting these assumptions unless random sets of college students are forced to become teachers, ITC graduates are randomly assigned

to randomly selected schools, and students are randomly assigned to classrooms within these schools (Corcoran, 2010; Harris, 2009; Ishii & Rivkin, 2009; Linn, 2008; Nelson, 2011; J. Rothstein, 2009).

Consider one of the most widely respected econometric models being advanced through New York's Teacher Pathways Project that involves 75 teacher education programs at 20 major teacher education institutions. Thus far, this projects' econometricians have evidenced that the gap between the qualifications of New York City (NYC) teachers in high- and low-poverty schools narrowed substantially between 2000 and 2005 (Boyd, Lankford, Loeb, Rockoff, & Wyckoff, 2008), largely because of new teachers hired through the NYC Teaching Fellows program and Teach for America (TFA). They have also found that classroom-based and applied learning opportunities facilitate more effective, first-year teachers (Boyd et al., 2008). But whether these new teachers are actually more effective in the classroom is still speculative due to the incredible amount of variation in actual classrooms, the lack of randomization for those within the certification programs, the schools in which they work, and the respective placement of students within classrooms, not to mention the available resources within those schools.

Notwithstanding, while these are all legitimate concerns that need to be considered before connecting student achievement data to teacher education programs, progress is still being made. Several teacher education programs and program consortiums are making advances toward examining these complex relationships. For example, California State University's Center for Teacher Quality (CTQ) built a mosaic to help them examine the impact of ITC graduates on their students' learning and achievement from 23 different university systems. Value-added models using standardized test scores are being supplemented with alternative measures of student learning across core and noncore subjects. Surveys administered to ITC graduates and their employers are being used, as well as teaching performance assessments (Center for Teacher Quality, 2007; Russell & Wineburg, 2007). Stanford University and 29 other California universities are currently implementing their Performance Assessment for California Teachers (PACT) project (Darling-Hammond, 2006a; Pecheone & Chung, 2006; Toch & Rothman, 2008). They are using survey and interview research methods to assess what candidates report they have learned in their programs and are assessing student learning using pre- and post-tests, work samples, employer surveys, clinical observations, and a validated teacher performance assessment largely modeled after the National Board for Professional Teaching Standards (NBPTS) Five Core Propositions for Teaching (Darling-Hammond, 2006a; Rubenstein, 2007). This project stands in stark contrast to those

in which researchers are utilizing and basing a teacher education program's value on either a snapshot of, or the gains derived from large-scale, standardized tests. Similarly, Ohio's Teacher Quality Partnership (TQP) involves all 50 colleges and universities within the state and, while using value-added data, it relies on qualitative methods to assess the impact of its teacher education programs on student learning and achievement as well (Berry et al., 2007).

Also, as discussed previously, the team affiliated with VAA-TPP at LSU has constructed a longitudinal database connecting students to teachers in core content areas, which are then used to indicate teacher education effectiveness. While these statistical models continue to be under construction, they are gaining prominence within the education reform debate and policy discussions (Noell & Burns, 2006; Noell et al., 2007). However, akin to the Education Value-Added Assessment System (EVAAS®) being used primarily for teacher evaluation and measurement purposes, this model presently relies mostly on test scores, overlooking some of the extraordinary assumptions previously noted and addressed elsewhere (Amrein-Beardsley, 2012; Scherrer, 2011). This model, like many of the others, still also falls short of being able to provide substantive feedback to the education programs regarding how they might change and improve. Even when the model works perfectly, it will only be able to indicate which teacher education programs are more highly correlated with graduates' students' achievement, which, as previously articulated, is riddled with problems (see also Noell, Gansle, Patt, & Schafer, 2009).

The methods used in the aforementioned states, and the others not discussed, are worth noting, however, as the spectrum of approaches may help others conceptualize their local education evaluation endeavors better, especially given the limitations and assumptions inherent within this research. In addition, if methods and measures from traditional and nontraditional teacher education programs are used collectively, this might help those involved get at teacher education quality more accurately, validly, and holistically as studies progress.

SEVEN "BEYOND EXCUSES" IMPERATIVES FOR EVALUATING TEACHER EDUCATION

Amid the noise, teacher education programs are still increasingly reconfiguring (Cochran-Smith, 2009) and connecting around various sets of state and national teaching standards to help them clarify their goals and drive what they do, from curriculum and instruction through research and evaluation projects (Cochran-Smith, 2001; Cochran-Smith

& Fries, 2001; Darling-Hammond, 2006a, 2006b; Russell & Wineburg, 2007; Yinger & Hendricks-Lee, 2000). At the national level, the frameworks most often used are the Interstate New Teacher Assessment and Support Consortium (INTASC) guidelines for entering/novice teachers, the NCATE standards for general teachers and their professional preparation units, and the National Board for Professional Teaching Standards (NBPTS) for expert or highly accomplished teachers. They all value content knowledge and pedagogy as equally important criteria for what teachers should know and be able to do across varied content areas.

National and state standards present teacher educators with a context to help them conceptualize, define, and assess what it means to be an effective teacher (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010), and will likely help to define what it means to prepare one. But, as teacher education programs coalesce and move toward building local consensus in terms of how they can most effectively use the aforementioned methodologies to evaluate themselves, they also, as we argue next, need to address a set of seven imperatives to ensure their efforts will be beneficial. These imperatives include (1) conceptualizing the purposes and reasons for evaluating teacher education programs; (2) defining effective teacher education programs and ideal ITC graduates; (3) building valid evaluation models; (4) resolving whether and which standards might be used to structure these models; (5) choosing appropriate data collection and analytical methods, and using or developing proper assessments sustaining these methods; (6) deciding who should be involved in decision making and at what levels (and whether nontraditional teacher educators should participate in such evaluations); and (7) determining how such program evaluations might be financed, supported, and sustained.

These imperatives were developed five years ago when a consortium of deans and faculty representing each of the state of Arizona's public colleges of education (i.e., Arizona State University [ASU], Northern Arizona University [NAU], and the University of Arizona [UofA]), along with approximately 30 educational leaders and stakeholders from throughout the state (e.g., representing the Arizona Department of Education [ADE], the Arizona Education Association [AEA], Arizona School Boards Association [ASBA]) collectively decided it was time to begin these evaluation studies. They created a collaborative tri-university initiative called the Teacher Preparation Research and Evaluation Project (T-PREP) (for more information, please see <http://education.asu.edu/projects/t-prep>).

The colleges involved, each residing within and representing the aforementioned universities and yielding a combined enrollment of approximately 10,000 education students, graduate the vast majority of

teachers every year. They also operate within a Republican-dominated, fiscally conservative state while they work to educate the second fastest growing population of PreK-12 students (U.S. Census Bureau, 2010); students who also consistently rank among the worst (bottom quintile) in the nation across grades and subject areas on the National Assessment of Educational Progress (NAEP; NCES, 2011). These teacher educators are also increasingly competing with and functioning alongside a growing number of alternative certification programs in the state.

For the first several years the T-PREP consortium met twice per year to steer and move this work forward, while a smaller working group that consisted of one to three faculty representatives from each college met more often. Through the many meetings and discussions, conversations at national conferences and meetings, and knowledge of additional legislative mandates (e.g. Common Core State Standards Initiative, 2010; HEA, 2007) their seven imperatives developed. These imperatives helped them create a framework for not only how they could become more reflective and hold themselves more accountable for the excellence of their graduates, but also internally evaluate and improve upon their programs and make more valid and evidenced-based claims that they, indeed, were still relevant.

Details within each of the seven imperatives devised follow. Also included are cases in point regarding what the particular college at focus, the Mary Lou Fulton Teachers College (MLFTC), has done since per each T-PREP imperative. MLFTC, the college now graduating more teachers than any other in the nation (Sawchuk, 2011) is situated within ASU, the university now spearheading the T-PREP initiative.

Imperative #1: Conceptualizing Purpose

Teacher educators and administrators need to analyze the reasons for evaluating their teacher education programs and, for purposes beyond accreditation (Cochran-Smith, 2009; Peck et al., 2010), ask why it is important to be held and to hold themselves accountable. Instead of resisting or dismissing this research as too complicated and complex (for example because of the uncontrollable variables involved), or instead, outright rejecting this research altogether (even if rejection is appropriate, given a lack of financial or human resources or capacities), teacher educators might work together to begin developing and defining more valid and innovative ways to examine their programs' strengths and weaknesses for both summative and formative purposes.

In this case, while there was initial resistance for all of the reasons above, all T-PREP project members decided that this was essential research

that they needed to inaugurate in order to respond to the aforementioned growing concerns about the antiquated and futile role of teacher education. High levels of commitment were also obtained due to the ever-present knowledge that teacher education programs were going to have to engage sooner or later (i.e., HEA, 2007). As the initial meetings took place, consortium members were also surprised that the relatively straightforward research questions initially asked were not, at that time, answerable (e.g., How many of a college's graduates remained in the field after one, three, or five years? Is there a difference in retention by certification type?). This too added fervor, as embarrassing as it was to not know the answers to even the simplest of questions. Thereafter, most everyone was on board because they realized that what they knew about the impact of their programs was virtually nil, particularly at any systemic level. As well, they agreed, there were important policy questions that needed to be answered, especially about the consistent and persistent need for strong, committed teachers in schools with high poverty and low achievement. They were unsure how their teacher education programs were satisfying these needs as well.

Those leading the T-PREP evaluation efforts at ASU took this one step further. They partnered with 25 high-needs districts, accounting for 230 schools, almost 12,000 teachers, and nearly 200,000 students, to facilitate these evaluations, largely by gaining wide-scale access to the schools and districts in which their graduates were teaching, or were not teaching for comparative purposes. In return for accessible data, the college began providing data tracking, other research, and program evaluation services to district personnel, and continuous professional development for teachers and administrators. Via the district–university partnerships that ensued and a resultant level of increased access, teacher education evaluators are becoming better equipped to evaluate their program's impact beyond graduation.

Imperative #2: Defining an Effective Teacher Education Program

Teacher educators, administrators, and other stakeholders need to define what an effective teacher education program and the ideal ITC graduate looks like. What are the characteristics of a good teacher? Is this different from what one of the aforementioned organizations might suggest (e.g., NBPTS)? What about the diversification of the teaching profession? How might teacher education programs recruit these teachers at higher rates into preprofession training? Should ideal ITC graduates choose to enroll at increasing rates count in terms of evaluating teacher education program quality? What do teacher candidates need during

their professional training to become successful teachers, particularly in terms of their content knowledge and pedagogical skills? And, how can teacher educators better prepare teacher candidates toward these ends? These were all questions of interest given this consortium.

In this case, college leaders worked with university administrators and faculty in ASU's college of arts and sciences to reform all of their teacher education programs to be more responsive to the needs of the state and districts. College leaders contend that continuously challenging conventionality and putting into check their own assumptions about what it is they are and should be doing to contribute to student learning and achievement in the state's PreK-12 public schools is obligatory. College leaders have also redefined the types of teachers they should be inviting into and graduating from their programs, defining them as teachers who have a deep understanding of subject-area content, create an environment of achievement for all PreK-12 students, and are much more exposed to and experienced in working in the actual conditions of real classrooms.

This decision followed a candid conversation, during which T-PREP consortium leaders questioned whether everyone could or should become a teacher, after which they had to acknowledge and accept that all teachers are not equal, and probably could never be equal in skills and potential effectiveness, regardless of their preservice training. Paying more attention to the Colorado Index Score for applicants and raising the percentage of students who have strong, positively predictive scores is and continues to be a core goal. Related, as other colleges of education move forward in considering who might be qualified, and theoretically privileged, to become a teacher, they must help to defy the common, and likely accurate, perceptions that colleges of education are unwilling and unable to develop the capacity to change. Thoroughly rejecting this prior notion, college leaders also contend that raising their own standards will create a shift to entice more of the best and brightest into the profession. It should be noted here, however, that the best and brightest should not just mean, as is typical, the best and whitest (Sleeter, 2008; Villegas & Irvine, 2010; Villegas & Lucas, 2004). In addition, now that these characteristics have been defined, program leaders are becoming better equipped to investigate whether, indeed, they are preparing the teachers they desire now by definition.

It is also important to build partnerships and relationships between education and other disciplines. Often, this kind of collaboration has to begin at the president or provost level, as it was in the case at ASU. There has to be a nonnegotiable stance that "It takes an entire university to prepare a teacher." Those conversations have to turn into action and

work plans so that content as well as pedagogy specialists have an infrastructure that allows them to constantly improve teacher education. See discussions forthcoming about the college's efforts to reform the preservice curricula at both the university and community college levels.

External partners must be actively recruited as well. For example, the Sanford Inspire Program (for more information, please see <http://sanfordinspireprogram.org/>) is an \$18 million, donor-funded initiative devised to integrate the best practices of TFA within the college and, if data are supportive, traditional teacher education programs, practices, and paradigms altogether. Following this, the college has revamped its traditional programs to include substantially more field-based instruction, and practical and applied learning opportunities (see also Sawchuck, 2011).

This is not without problems, however. While the goal of the Sanford Inspire Program is to prepare highly effective teachers by integrating best practices from both TFA and the college, nobody believes that all practices from TFA can, or should, transfer directly into the college. For example, TFA uses a highly selective model for accepting corps members that they believe is predictive of success in the classroom. That is not the college's intent, necessarily. The college's goal is to provide access to all prospective students who meet program requirements. To that end, the college is working on a broader agenda that, while ideologically different, is still working to attract students who are committed to the idea that each child deserves an excellent teacher. The college is just going about this in a different way and testing new strategies to recruit more students and more talent to the teaching profession along the way.

Imperative #3: Valid Models

Teacher educators and leaders need to decide how they might reasonably go about measuring program quality, as stated. Within the college in focus, leaders set out to create the aforementioned research and evaluation services to help them build and foster partnerships with districts, but also to facilitate access to data and the construction of mutual data systems to help them measure teacher quality as a reflection of teacher education quality. Leaders also set out to not only prepare caring, competent, and capable teachers, but teachers who are measurably effective (Barnett & Amrein-Beardsley, 2011). In as much as these teachers are still being defined, this work is being used to inform the state as it too moves toward measuring teacher and teacher education quality.

Thus far, at ASU, measurably effective teachers are not being defined using sole indicators of student achievement. Rather, effective teachers are being distinguished as those who are retained; positively impact

student learning, achievement, and growth on multiple measures of student learning and achievement; collaborate with school districts to implement meaningful reform; help to turn around the lowest-performing schools; and challenge historical norms. These indicators were identified through discussions with various stakeholders involved in the T-PREP consortium's efforts, again including the deans and representative faculty from the colleges, representatives from the state department and the union, and others. And, while the indicators include student achievement outcomes (i.e., derived via value-added measures), they also include performance indicators captured during student teaching and beyond the teacher education programs (to see a model of the evaluation structure, please see the Appendix).

For example, to supplement value-added analyses and to take a more mixed-methods approach to this research in line with the educational measurement standards of the profession (AERA, APA, & NCME, 1999), the college adopted the National Institute for Excellence in Teaching's (NIET) validated observational instrument at the core of the System for Teacher and Student Advancement (TAP™). See further details forthcoming, but ultimately the time and effort put into this politicized and somewhat controversial approach (Mathematica, 2010; Sawchuck, 2010a) most common in PreK-12 schools (TAP, 2012) should provide a better, more holistic, and intuitively more valid picture of what it means to be an effective teacher and, in this case, effective teacher education program (see also Capitol Hill Briefing, 2011; Darling-Hammond, 2006a), as many of the college's graduates move into university-partner schools who are also using the same TAP system. While, of course, this facilitates the longitudinal analyses needed to evaluate teacher education effectiveness, these efforts are just now being examined as cumulative numbers of graduates increasingly contribute to longitudinal records. Most notably, these longitudinal records now include observational data along with student outcome data, and this will ultimately facilitate a more valid approach to inquiry, one based on mixed methods.

Imperative #4: Standard Setting

Teacher educators need to decide whether those involved should adopt a set of national or state standards to help frame these evaluations and the instruments and assessments developed and deployed to conduct them. If so, which standards are most appropriate for local programs, and at what level should these decisions be made?

In this case, when T-PREP leaders addressed the details of data collection and analysis, they agreed all endeavors and instruments would be aligned with the Five Core Propositions of the NBPTS. The propositions

exist to help others frame what it means to be an exemplary teacher in terms of a rich fusion of knowledge, skills, dispositions, and beliefs (NBPTS, 2011). The NBPTS propositions continue to provide a unified vision of exemplary teaching, but because of practical limitations (namely, a lack of validated instruments that are aligned with the core propositions), the college, as mentioned, began supplementing efforts using the TAP observational instrument instead. Teacher educators are using TAP to help them better conceptualize and capture what it means to be a measurably effective teacher (TAP, 2012). Like the state, they are also incorporating the INTASC standards.

At the applied level, the integration of the TAP observational instrument was needed to more concretely help teacher education program leaders restructure teaching programs and initiatives to (a) recruit more high-quality teachers, (b) provide teachers with a career continuum, (c) implement teacher-led professional development, (d) establish a more rigorous teacher accountability system, and (e) grant proportionate compensation (via grant projects) based on teachers' positions, skills, knowledge, and performance (Schacter & Thum, 2005). Using this framework and videos of exemplary practice, the goal was to build teacher candidates' theoretical and research perspectives on how and why "best practices" are best. These efforts make the college one of the first to use TAP in higher education as a program-defining, intensive, and completely integrated formative teaching/learning tool, versus a tool traditionally used for summative purposes only.

Specifically, teacher candidates engage in a rigorous performance assessment process conducted twice each semester during a full year of student teaching. The process involves a planning protocol (to promote the investigation of content using Common Core Standards), a formal observation (that is videotaped), a self-evaluation using the TAP instructional rubric (along with the videotape and K-12 student outcome data), an evaluation using the TAP rubric (conducted by a faculty member who is housed in the school district full time), and a post-conference/coaching session (conducted by the same faculty member within 24 hours) to define areas of reinforcement and refinement.

In addition, because the TAP system is an in-service model, the college worked with the NIET Best Practices Center to create a preservice model focused primarily on two of TAP's four key elements: "instructionally focused accountability" and "ongoing applied professional growth" (TAP, 2012). While teacher candidates are still expected to be proficient on all TAP indicators by graduation, in the preservice model, teacher candidates are also required to engage in professional development, given their ongoing and embedded coursework, alongside their mentors within schools prior to the start of the school year.

It should be noted here, again, that the use of this rubric here is primarily for formative purposes. While the TAP system has been evidenced to be one of the most objective rubric systems available, particularly as it has undergone a number of validity studies, it is not being used as a hard-and-fast tool for either individual or teacher education accountability, particularly because this would stretch far beyond its research-based utilities (Glazerman & Seifullah, 2010; Sawchuck, 2010b; Schacter & Thun, 2005; Solomon, White, Cohen, & Woo, 2007). Its value lies more in its capacity as a descriptive tool that yields signals about what a preservice teacher might be doing well, must be doing prior to graduation, or what the teacher education program might be or might not be doing well overall.

Imperative #5: Data Collection and Analyses

Related, teacher educators and leaders need to decide what data collection tools (e.g., the TAP observational system) and analytical methods might be constructed and used to help evaluate program impact, how often data should be collected, and for what purposes. The goal here was to conduct disciplined evaluations of individual courses and entire programs, and to ultimately use data to inform positive change. While the concept of collecting data on teacher education programs is certainly not novel, the utility of this information is where improvement is needed. Currently, most teacher education programs likely collect voluminous amounts of data about their students; however, these data are not well used or purposefully collected. Rather, data collection needs to be conducted and connected to action.

With the focus on action, T-PREP consortium leaders created a concept map (please see the appendix) that would help them address the details of instrumentation, data collection, analysis, and dissemination. Evaluators at ASU specifically are collecting the following data: preprogram data from the application materials of incoming students, program data collected from instruments created by the evaluation team and the professional field experience offices, and postprogram data from the state department of education. All data are housed in each institution's data system, its field experience offices, in larger university systems, and via the state's data warehouse.

College leaders at ASU were also instrumental in the conceptualization and development of the state's data warehouse system that tracks teacher graduates for research purposes. Leaders also helped to create an electronic Institutional Recommendation (IR) system in which all program graduates' information is automatically uploaded into the state system.

From there, it can be combined with employment records that facilitate tracking each institution's graduates per year, making tracking and comparing teachers throughout the state more feasible. The focus of this system was to reduce paperwork and confusion between the state department of education and the colleges of education; therefore, the state department of education funded this project as part of its own internal improvements. These records will eventually be combined with state and local student test scores and other achievement indicators, facilitating such analyses further.

This increased access to data triggered a joint discussion about what indicators should be used to measure teacher education quality and, primarily, the role that standardized tests should play in these investigations (see also Baker et al., 2010; Cochran-Smith, 2001, 2004, 2005, 2009; Zeichner, 2010). Should ITC graduates' students' learning be measured using standardized tests? Should standardized test scores be used at all? What if they are used to measure value added or growth? How else might researchers go about measuring student learning? In the same vein, should ITC graduates' scores on required licensure tests be used to measure program quality?

In this case, those involved decided to take a holistic approach. They decided standardized test scores might be used to evaluate ITC graduates' impact or positive influence on student achievement via value-added measures, with mutual understandings about what standardized tests can and cannot do and about the limitations of value-added analyses (Au, 2010; Haertel, 2011; Harris, 2011; Hill et al., 2011; Newton et al., 2010; Papay, 2010; J. Rothstein, 2009). And, while results would be contextualized and limitations defined, ultimately those involved agreed that using standardized tests, at the student or teacher level, as the only indicator of program quality would be negligent and would violate the standards of the profession on the appropriate uses of tests (AERA, APA, & NCME, 1999).

The primary reason to support adopting a value-added model, however, stemmed from the decision of the state. Once the state decided it was going to use a value-added measure, the project leaders already working in this area helped inform and contribute to the conversation. The oft-cited Student Growth Percentiles (SGP) model (Betebenner & Linn, 2010) was selected to determine graduates' influence on students, although legislation is still being finalized a propos implementation and use. This model was selected because it is open access, open to further manipulation and investigation, and appropriately descriptive (vs. causative), that is, compared to other commonly used value-added models.

Imperative #6: Decision Making

Teacher educators and leaders need to decide who should be involved in decision making and at what levels they should be engaged. Rather than relying only on what traditional teacher educators might have to say about their programs, they should also ask other preparers of teachers to contribute to the construction of reality and understanding. This would assist leaders' efforts to become stronger democratic players in educational policy, and help to legitimize the field (Hamel & Merz, 2005; Russell & Wineburg, 2007; Stake, 1967; Wineburg, 2006; Yinger & Hendricks-Lee, 2000). As well, this would help those who evidently prepare teachers in more particular ways or areas better, at least theoretically, help teacher education as a whole.

For example, alternative teacher education programs (e.g., TFA, other for-profits, and the community colleges or high schools in which teachers are also more frequently being prepared) are often not being included in discussions about teacher education quality. As one dean within the project noted, "It is easy enough to compete [with each other], far more difficult to collaborate." If the ultimate goal is to determine what components of teacher education programs impact teacher quality and student learning most, teacher educators should be willing to learn from each other to improve programs across the board. In addition, regardless of the impact of different approaches, there is no question that teacher candidates continue to be trained through both traditional and alternative certification routes. As such, not including alternative programs in this research might not be politically prudent or practically wise, and perhaps shortsighted, defeatist, and self-promoting (Yinger & Hendricks-Lee, 2000). Nontraditional programs might have something to offer traditional teacher education programs regarding direct experience (Raymond, Fletcher, & Luque, 2001), high school programming (Good et al., 2006), school and classroom-based assessment (Farr, 2010), recruiting teacher candidates from diverse backgrounds (Wilson et al., 2002), and teaching in the most difficult to teach schools (Boyd et al., 2006).

In this case, college leaders are working with the state department of education that is in charge of the accreditation process. But also, and more importantly, they have the aligned goal of evaluating all of the state's teacher education programs for the betterment of the state's entire educational system. This partnership makes the most sense.

At the same time, however, college leaders are still overcoming the reluctance to learn from and be influenced by their competitors, even given the aforementioned Sanford Inspire Program. For example, in the area of curriculum, some might argue that there is not really anything

that the college can learn from an organization like TFA. But, college researchers analyzed the research behind TFA's core curriculum alongside the college's curriculum, and have no major differences in terms of content or pedagogy. On the contrary, they found that much of the TFA curriculum is based on the same research, theory, and practice that were already in place across the college's programs. Yet tensions still exist, again particularly in the area of curriculum, but this might be more because of the highly politicized differences between traditional teacher education programs and alternative paths like TFA versus true, observable differences between the two programs that are now partners.

Externally, the college is also working more deliberately with other private and public competitors. For example, college leaders have traditionally viewed the community colleges as a way to transfer students of more diverse backgrounds into teacher education programs. The college has facilitated this through carefully articulated agreements that are constantly maintained to make it possible for the recruitment of diverse populations that do not always enter a four-year university. This has not been the most progressive approach to diversifying the teaching force (Sleeter, 2008; Villegas & Irvine, 2010; Villegas & Lucas, 2004), but this has duplicitously been in place for years.

Now, the college is viewing its community college partners as partners, as leaders from all institutions involved are working together to reform preservice curricula. Specifically, the college is working more deliberately with the university president's office, as well as accepting the university's colleges of liberal arts and sciences, engineering, and arts and letters as full partners in educating teachers, chiefly in terms of content knowledge and expertise. The college has begun to have content courses delivered by faculty from these external colleges to better prepare students with relevant subject-area expertise, and to develop their leadership, critical thinking, communication, and organizational skills in these areas. This interdisciplinary work has been supported via the development or reform of 40 lower division subject area courses as part of the college's Teaching Foundations Project (for more information, please see <http://orc.teach.asu.edu/tfpdev>).

A lesson learned here, again, is that such efforts are not easily implemented. Candid and often extensive discussions are necessary for content program faculty new to teacher education (e.g., faculty members from university biology, history, and physics departments), those now being charged with teaching undergraduate courses, to understand the differences between their traditional and ITC undergraduate students. At the university level, for example, ongoing collaborations have resulted in the design of forty new courses for those entering the college's

teacher education program. This transition is ever-beckoning a balance in understanding and approach, that is, help for faculty relatively new to teacher education both to understand the needs of teacher education students and to recognize the very real differences between those potentially going on to teach PreK-12 students and others. Notwithstanding, it is precisely these conversations that are happening that should help teacher educators approach teacher education in more innovative and new ways—where content requirements are increased without being unduly burdensome, and where purpose and impact are at the forefront of course design.

As well, the college has recognized the need to reform its own methods coursework to also prepare teacher candidates for a rigorous clinical experience. Specifically, each syllabus was reformed to include both formative and summative performance assessments based on indicators from the TAP instructional rubric. The formative assessments provide teacher candidates explicit opportunities to practice new learning and to get feedback from peers, mentor teachers, and faculty instructors. In the summative assessments, each teacher candidate is expected to demonstrate proficiency on the TAP instructional rubric. These course-embedded assessments using the TAP instructional rubric are aimed at preparing teacher candidates for the performance assessment process (which now represents over 50% of their student teaching grade). The long-term plan is to foster students' actual implementation of content knowledge and exemplary practice in school sites almost exclusively.

Imperative #7: Sustainable Funding

Teacher educators and leaders need to determine how these program evaluations might be sufficiently financed, supported, and sustained (see also Cochran-Smith, 2009). Without substantial financial and human resource support, teacher education evaluations of this scale are not feasible. At a larger scale, the Carnegie Corporation (Darling-Hammond, 2006a; Russell & Wineburg, 2007), the Milken Family Foundation (Goldrick, 2002), and local and private foundations, state departments of education, universities, and teacher education programs are pitching in to help those involved develop these models before bringing them to scale. This issue will likely present more substantial financial challenges in the long run, however, especially if these investigations are to be continued indefinitely.

For this project, a community foundation within the state provided the seed funds to get T-PREP's work started. Otherwise, faculty and program evaluators within have borne the research and service costs of planning, conducting, and managing these evaluations. As they proceed, however,

they are reminded that even with a clear vision, without such an investment, these types of long-range projects may vanish. Notwithstanding, the faculty members involved continue to apply for grants to support the overall project. These efforts were also acknowledged through over \$75 million in grants awarded to the college by the U.S. Department of Education (Des Georges, 2010; Parker, 2010). These resources will continue to help implement the aforementioned plans, and they continue to facilitate the connection among the college and its partnering districts.

Additionally, the president of the university allocated substantial funds, even though funds are still scarce, to support the T-PREP project specifically, indicating to the university and greater community that training effective teachers is a top university priority. This contribution was initially triggered when community members who met with the president posed some of the questions previously described. When the questions went unanswered, this was surprising to the president as well in that he also assumed teacher education personnel knew the answers. With this, the conversations echoing across colleges of education, and the continued policy changes from the federal administration, even the university president determined it was time to begin holding not just the college of education, but the university, accountable for all of the teachers it was graduating.

CALL TO ACTION

Three decades ago, David Berliner noted that those who evaluate teacher education programs too often suffer from “ostrichism” (1976, p. 5), a disease afflicting those who, when study results are unexpected or expose blemishes, stick their heads in the sand, hoping problems will pass. Teacher educators teach students to be reflective practitioners and, likewise, should have no issues with being thoughtful and critical of their own programs, practices, and paradigms. While it is true that many flaws can be explained away, for example, when distinctly different samples of students respond in significantly dissimilar ways about program quality, it is also true that nothing will change unless, after imperfections are revealed and understood in context, the flaws and failings inform change.

Notwithstanding the risks associated with conducting such evaluations in the current climate of accountability, consensus does not yet exist about how to commence or conduct these large-scale teacher education evaluations (Cochran-Smith, 2009; Ludlow et al., 2010; Peck et al., 2010; Wineburg, 2006; Zeichner, 2010). Although, many researchers across the country are working to help teacher educators take part in how they will be evaluated (Cochran-Smith, 2009; Russell & Wineburg, 2007; Wineburg, 2006). The American Association of State Colleges and

Universities (AASCU), for example, released a policy paper in which authors argued, “It is time to develop a national framework for the collection of evidence of the effectiveness of teacher education programs” (Russell & Wineburg, 2007, p. 3).

It is argued herein that teacher educators (including administrators and teacher educators as they are indeed a vested group with a collective responsibility to examine if what they are doing is of high quality, meaningful, and impactful) should commence this work if they have not done so yet. This is true, especially given all of the aforementioned concerns and uncertainties, leaving teacher educators as probably the best to conduct such research internally and as situated within their local contexts. There are too many details that make states too different and teacher education programs too unique for large-scale evaluations like those that the federal government might recommend.

Additionally, as previously noted, several current evaluation systems (e.g., VAA-TPP) rely too heavily, or in some cases solely, on student test scores, which fail to capture the complexity of teacher education programs. Further, to rank teacher education programs based on achievement values alone is inconsistent with any other educational evaluation, at least any evaluation that is wisely conducted. That said, it is critical to the profession that its members signal to the public and policymakers that the profession has established cognitive jurisdiction (Yinger & Hendricks-Lee, 2000) and has begun to evaluate its teacher education programs as separate but public entities. Failure to recognize and actively conduct such work will lead to a further condemnation of teacher education programs.

Throughout the previous sections, an evaluation framework built on seven “beyond excuses” imperatives was proposed. Beyond this framework, the experience and “progress” being made at one college involved in one project was described, as was the considerable work of those engaged in the T-PREP consortium overall. Participants have come to realize that because this system is fluid, and given the very public, high-stakes nature of this work, there exists a dire need for constant communication and collaboration with all stakeholders, from the statehouse, through the university systems, and onto the schools and districts they serve. They understand they need to collectively work toward building local evaluation models in democratic, inclusive ways, and they acknowledge their obligatory roles to collectively legitimize, publicly shape, and make transparent teacher educators’ and teacher education leaders’ points of view about such evaluative investigations.

References

- American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Amrein-Beardsley, A. (2012). Value-added measures in education: The best of the alternatives is simply not good enough. *Teachers College Record*. Retrieved from <http://www.tcrecord.org/Content.asp?ContentId=16648>
- Anrig, G. R. (1986). Teacher education and teacher testing: The rush to mandate. *Phi Delta Kappan*, 67, 447–451. doi:10.2307/2295102
- Au, W. (2010, Winter). Neither fair nor accurate: Research-based reasons why high-stakes tests should not be used to evaluate teachers. *Rethinking Schools*, 25(2). Retrieved from http://www.rethinkingschools.org/archive/25_02/25_02_au.shtml
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., . . . & Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute. Retrieved from <http://www.epi.org/publications/entry/bp278>
- Ballou, D., Sanders, W. L., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, 29(1), 37–66. doi:10.3102/10769986029001037
- Barnett, J. H. & Amrein-Beardsley, A. (2011). Actions over credentials: Moving from highly qualified to measurably effective [Commentary]. *Teachers College Record*. Retrieved from <http://www.tcrecord.org/Content.asp?ContentID=16517>
- Berliner, D. (1976). Impediments to the study of teacher effectiveness. *Journal of Teacher Education*, 27, 5–13. doi:10.1177/002248717602700103
- Berry, B., Fuller, E., & Reeves, C. (2007, March). *Linking teacher and student data to improve teacher and teaching quality*. Center for Teaching Quality, National Center for Educational Accountability, and Data Quality Campaign. Retrieved from <http://www.dataqualitycampaign.org/resources/details/122>
- Betebenner, D. W. & Linn, R. L. (2010). *Growth in student achievement: Issues of measurement, longitudinal data analysis, and accountability*. Center for K-12 Assessment and Performance Management. Retrieved from www.k12center.org/rsc/pdf/BetebennerandLinnPolicyBrief.pdf
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., Michelli, N., & Wyckoff, J. (2006). Complex by design: Investigating pathways into teaching in New York City schools. *Journal of Teacher Education*, 57(2), 102–119. doi:10.1177/0022487105285943
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2007). *Examining teacher preparation: Does the pathway make a difference?* Teacher Policy Research. Retrieved from <http://www.teacherpolicyresearch.org>
- Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management*, 27(4), 793–818. doi:10.1002/pam.20377
- Bracey, G. W. (1995). Variance happens: Get over it! *Technos*, 4(3), 22–29.
- Capitol Hill Briefing. (2011, September 14). *Getting teacher evaluation right: A challenge for policy makers*. Washington DC: Dirksen Senate Office Building (research in brief). Retrieved from <http://www.aera.net/Default.aspx?id=12856>
- Center for Teacher Quality. (2007). *Teacher preparation program evaluation based on K-12 student learning and performance assessments by school principals*. California State University. Retrieved from <http://www.calstate.edu>

- Chingos, M. M., & Peterson, P. E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review*, 30(3), 449–465.
- Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007, October). *Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects*. Cambridge, MA: National Bureau of Economic Research.
- Cochran-Smith, M. (2001). The outcomes question in teacher education. *Teaching and Teacher Education*, 17(5), 527–546. doi:10.1016/S0742-051X(01)00012-9
- Cochran-Smith, M. (2004). Teacher education in dangerous times. *Journal of Teacher Education*, 55(1), 3–7. doi:10.1177/0022487103261227
- Cochran-Smith, M. (2005). The new teacher education for better or worse? *Educational Researcher*, 34(7), 3–17. doi:10.3102/0013189X034007003
- Cochran-Smith, M. (2009). “Re-culturing” teacher education: Inquiry, evidence, and action. *Journal of Teacher Education*, 60(5), 458–468. doi:10.1177/0022487109347206
- Cochran-Smith, M., Feiman-Nemser, S., McIntyre, J., & Demers, K. E. (2008). *Handbook of research on teacher education: Enduring questions in changing contexts*. (3rd ed.). New York: Routledge.
- Cochran-Smith, M., & Fries, M. K. (2001). Sticks, stones, and ideology: The discourse of reform in teacher education. *Educational Researcher*, 30(8), 3–15. doi:10.3102/0013189X030008003
- Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfield, F., & York, R. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.
- Common Core State Standards Initiative. (2010). About the standards. Retrieved from <http://www.corestandards.org/about-the-standards>
- Corcoran, S. P. (2010). *Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice*. Providence, RI: Annenberg Institute for School Reform. Retrieved from <http://www.annenberginstitute.org/products/Corcoran.php>
- Darling-Hammond, L. (2006a). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. *Journal of Teacher Education*, 57(2), 120–138. doi:10.1177/0022487105283796
- Darling-Hammond, L. (2006b). Constructing 21st-century teacher education. *Journal of Teacher Education*, 57(3), 300–314. doi:10.1177/0022487105285962
- Darling-Hammond, L., & Sykes, G. (2003). Wanted: A national teacher supply policy for education: The right way to meet the “highly qualified teacher” challenge. *Educational Policy Analysis Archives*, 11(33). Retrieved from <http://epaa.asu.edu/epaa/v11n33/>
- Des Georges, S. (2010). Teachers College awarded \$43M grant to help reform Ariz. Schools. *ASU News*. Retrieved from http://asunews.asu.edu/20101011_education_grant
- Education Digest. (2011, March). Transforming teacher education through clinical practice: A national strategy to prepare effective teachers. *Education Digest*, 76(7), 9–13.
- Ewell, P. T., Schild, P. R., & Paulson, K. (2003, April). *Following the mobile student: Can we develop the capacity for a comprehensive database to assess student progression?* Indianapolis, IN: Lumina Foundation for Education: Research Report. Retrieved from <http://www.luminafoundation.org>
- Farr, S. (2010). *Teaching as leadership: The highly effective teacher's guide to closing the achievement gap*. San Francisco, CA: Jossey-Bass.
- Glass, G. V. (2008). *Fertilizers, pills, and magnetic strips: The fate of public education in America*. Charlotte, NC: Information Age Publishing.

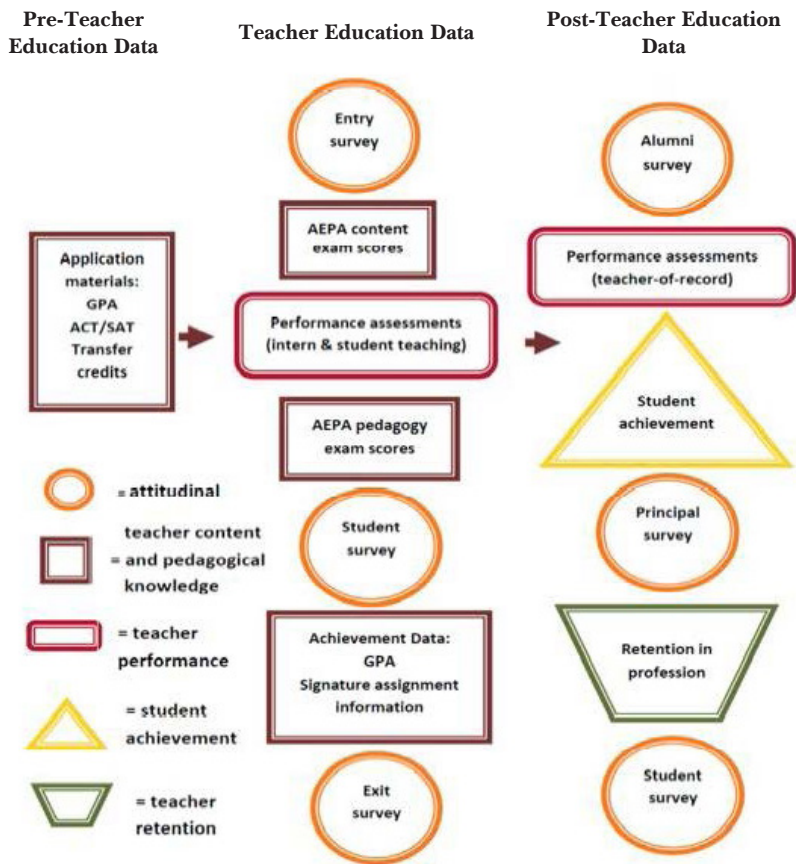
- Glazerman, S. & Seifullah, A. (2010, May). *An evaluation of the Teacher Advancement Program (TAP) in Chicago: Year two impact report*. Washington DC: Mathematica Policy Research, Inc. Retrieved from http://www.mathematica-mpr.com/newsroom/releases/2010/TAP_5_10.asp
- Goldrick, L. (2002). Improving teacher evaluation to improve teaching quality. Retrieved from <http://www.nga.org/cda/files/1202IMPROVINGTEACHEVAL.pdf>
- Good, T. L., McCaslin, M., Tsang, H. Y., Zhang, J., Wiley, C. R. H., Rabidue Bozack, A., & Hester, W. (2006). How well do 1st-year teachers teach: Does type of preparation make a difference? *Journal of Teacher Education*, 57(4), 410–430. doi:10.1177/0022487106291566
- Goodwin, L. A. (2009). Remaking our teacher education history through self-study. *Studying Teacher Education*, 5(2), 143–146. doi:10.1080/17425960903306583
- Greenleaf, C. L., Litman, C., Hanson, T. L., Rosen, R., Boscardin, C. K., Herman, J., . . . Jones, B. (2011). Integrating literacy and science in biology: Teaching and learning impacts reading apprenticeship professional development. *American Educational Research Journal*, 48(3), 647–717.
- Haertel, E. (2011). *Using student test scores to distinguish good teachers from bad*. Paper presented at the Annual Conference of the American Educational Research Association (AERA), New Orleans, LA.
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17–27.
- Hamel, F. L., & Merz, C. (2005). Reframing accountability: A preservice program wrestles with mandated reform. *Journal of Teacher Education*, 56(2), 157–167. doi:10.1177/0022487105274458
- Harris, D. N. (2009). Would accountability based on teacher value added be smart policy? An evaluation of the statistical properties and policy alternatives. *Education Finance and Policy*, 4, 319–350. doi:10.1162/edfp.2009.4.4.319
- Harris, D. N., (2011). *Value-added measures in education: What every educator needs to know*. Cambridge, MA: Harvard Education Press.
- Harris, D. N., & Sass, T. R. (2007). *Teacher training, teacher quality and student achievement*. National Center for Analysis of Longitudinal Data in Education Research. Retrieved from <http://www.caldercenter.org/>
- Heubert, J. P., & Hauser, R. M. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press. Retrieved from http://www.nap.edu/catalog.php?record_id=6336
- Higher Education Amendments (HEA). (2007, November 15). *U.S. Senate: the Committee on Health, Education, Labor, and Pensions*. Retrieved from <http://www.govtrack.us/congress/bill.xpd?bill=s110-1642>
- Hill, H. C., Kapitulka, L., & Umland, K. (2011, June). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794–831. doi:10.3102/0002831210387916
- Ishii, J., & Rivkin, S. G. (2009). Impediments to the estimation of teacher value added. *Education Finance and Policy*, 4, 520–536. doi:10.1162/edfp.2009.4.4.520
- Kaufman, J. (2007, December). *Teaching quality: Induction programs for new teachers*. Denver, CO: Education Commission of the States.
- Koedel, C., & Betts, J. R. (2010). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education Finance and Policy* 6(1), 18–42. doi:10.1162/EDFP_a_00027
- Kreitzer, A. E., Madaus, G. F., & Haney, W. (1989). Competency testing and dropouts. In L. Weis, E. Farrar, & H. G. Petrie (Eds.), *Dropouts from school: Issues, dilemmas, and solutions*. Albany, NY: State University of New York Press.

- Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee value added assessment system. *Educational Evaluation & Policy Analysis*, 25(3), 287–298. doi:10.3102/01623737025003287
- LeClaire, B. (2011, June 1). Will EVAAS make Wake schools Better? *Raleigh Public Record*. Retrieved from <http://www.raleighpublicrecord.org/featured/2011/06/01/will-evaas-make-wake-schools-better-part-ii/>
- Linn, R. L. (2008). Methodological issues in achieving school accountability. *Journal of Curriculum Studies*, 40, 699–711. doi:10.1080/00220270802105729
- Ludlow, L., Mitescu, E., Pedulla, J., Cochran-Smith, M., Cannady, M., Enterline, S., & Chappe, S. (2010). An accountability model for initial teacher education. *Journal of Education for Teaching: International Research and Pedagogy*, 36(4), 353–368. doi:10.1080/02607476.2010.513843
- Mathematica Policy Research Inc. (2010, June 1). *Early results show no impact of Teacher Advancement Program in Chicago: No measurable effect on teacher retention, student test scores in second year of rollout*. Retrieved from http://www.mathematica-mpr.com/Newsroom/Releases/2010/TAP_5_10.asp
- McCaffrey, D. F., Lockwood, J., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29(1), 67–101. doi:10.3102/10769986029001067
- National Board for Professional Teaching Standards (NBPTS). (2011). *Five core propositions*. Retrieved November 19, 2011 from http://www.nbpts.org/the_standards/the_five_core_propositio
- National Center for Education Statistics (NCES). (2011). *State profiles*. Retrieved November 15, 2011 from <http://nces.ed.gov/nationsreportcard/states/>
- National Governors Association Center for Best Practices & Council of Chief State School Officers (2010). *Common core state standards initiative: Frequently asked questions*. Retrieved from corestandards.org/frequently-asked-questions
- Nelson, F. H. (2011, April). *A guide for developing growth models for teacher development and evaluation*. Paper presented at the Annual Conference of the American Educational Research Association (AERA), New Orleans, LA.
- Newton, X., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010) Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Educational Policy Analysis Archives*, 18(23). Retrieved from <http://epaa.asu.edu/ojs/article/view/810>
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425. (2002). Retrieved from <http://www.ed.gov/legislation/ESEA02/>
- Noell, G. H., & Burns, J. L. (2006). Value-added assessment of teacher preparation: An illustration of emerging technology. *Journal of Teacher Education*, 57, 37–50. doi:10.1177/0022487105284466
- Noell, G. H., Gansle, K. A., Patt, R. M., & Schafer, M. J. (2009). *Value added assessment of teacher preparation in Louisiana: 2005–06 to 2008–09*. Retrieved from <http://www.tntp.org>
- Noell, G. H., Porter, B. A., & Patt, R. M. (2007). *Value added assessment of teacher preparation in Louisiana: 2004–2006*. Retrieved from <http://www.nctq.org/nctq/research/1196366605384.pdf>
- Papay, J. P. (2010). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, 48(1), 163–193. doi:10.3102/0002831210362589
- Parker, R. (2010). Arizona schools to receive more federal aid to improve. *Arizona Republic*. Retrieved from <http://www.azcentral.com/arizonarepublic/local/articles/2010/10/14/20101014arizona-schools-receive-federal-aid-for-improvement.html>

- Pecheone, R. L., & Chung, R. R. (2006). Evidence in teacher education: The performance assessment for California teachers (PACT). *Journal of Teacher Education*, 57(1), 22–36. doi:10.1177/0022487105284045
- Peck, C. A., Galluci, C., & Sloan, T. (2010). Negotiating implementation of high-stakes performance assessment policies in teacher education: From compliance to inquiry. *Journal of Teacher Education*, 61(5), 451–463. doi:10.1177/0022487109354520
- Raymond, M. E., Fletcher, S., & Luque, J. (2001, July). *Teach for America: An evaluation of teacher differences and student outcomes in Houston, Texas*. Retrieved from credostanford.edu/downloads/tfa.pdf
- Rivkin, S. G. (2007, November). *Value-added analysis and education policy*. National Center for Analysis of Longitudinal Data in Education Research. Urban Institute. Retrieved from http://www.urban.org/UploadedPDF/411577_value-added_analysis.pdf
- Rothstein, J. (2009, January 11). *Student sorting and bias in value-added estimation: Selection on observables and unobservables*. Cambridge, MA: The National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w14666>
- Rothstein, R. (2011, September 16). *Abet over No Child Left Behind*. *The Economic Policy Institute Blog*. Retrieved from <http://www.epi.org/blog/rothstein-ravitch-no-child-left-behind/>
- Rubenstein, G. (2007). Confronting the crisis in teacher training: Innovative schools of education invent better ways to prep educators for the classroom. *Edutopia*. Retrieved from <http://www.edutopia.org/building-a-better-teacher>
- Russell, A. & Wineburg, M. (2007). *Toward a national framework for evidence of effectiveness of teacher education programs*. American Association of State Colleges and Universities (AASCU). Retrieved from http://www.aascu.org/pdf/07_perspectives.pdf
- Sawchuck, S. (2010a, June 1). Performance-pay model shows no achievement edge. *Education Week*. Retrieved from <http://www.edweek.org/ew/articles/2010/06/01/33tap.h29.html?tkn=SLLFZe8XVYfHJJSSgGYCI87ZvETbCbN%2FXmT&cmp=clp-edweek>
- Sawchuck, S. (2010b, September 25). Merit-pay model pushed by Duncan shows no achievement edge. *Education Week*, 29(33), 1, 21.
- Sawchuk, S. (2011, November 21). ASU reforms elementary ed. content coursework. *Education Week*. Retrieved from http://blogs.edweek.org/edweek/teacherbeat/2011/11/asus_teacher_cd_college_r.html
- Schacter, J., & Thum, Y. M. (2005). TAPping into high quality teachers: Preliminary results from the teacher advancement program comprehensive school reform. *School Effectiveness and School Improvement*, 16(3), 327–353.
- Scherrer, J. (2011). Measuring teaching using value-added modeling: The imperfect panacea. *NASSP Bulletin*, 95(2), 122–140. doi:10.1177/0192636511410052
- Shulman, L. S. (1988). A union of insufficiencies: Strategies for teacher assessment in a period of educational reform. *Educational Leadership*, 46(3), 36–41.
- Sleeter, C. (2008). Equity, democracy, and neoliberal assaults on teacher education. *Teaching and Teacher Education*, 24(8), 1947–1957. doi:10.1016/j.tate.2008.04.003
- Solomon, L., White, J. T., Cohen, D., & Woo, D. (2007). *The effectiveness of the teacher advancement program*. Santa Monica, CA: National Institute for Excellence in Teaching. Retrieved from <http://www.fldoe.org/dpe/pdf/effectiveness-of-TAP.pdf>
- Sparks, S. D. (2011, November 15). “Value-added” formulas strain collaboration. *Education Week*. Retrieved from <http://www.edweek.org/ew/articles/2011/11/16/12collab-changes.h31.html?tkn=OVmFb8PQXxQi4wN6vpelNlr7%2BNhOFCbi71ml&intc=es>
- Stake, R. E. (1967). The continuance of educational evaluation. *Teachers College Record*, 68, 523–540.
- TAP. (2012). *The system for teacher and student advancement*. Santa Monica, CA: National Institute for Excellence in Teaching. Retrieved from <http://www.tapsystem.org/policyresearch/policyresearch.taf>

- Tekwe, C. D., Carter, R. L., Ma, C., Algina, J., Lucas, M. E., Roth, J., . . . Resnick, M. B. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics*, 29(1), 11–35. doi:10.3102/10769986029001011
- Toch, T., & Rothman, R. (2008). *Rush to judgment: Teacher evaluation in public education*. Education Sector. Retrieved from http://www.educationsector.org/usr_doc/RushToJudgment_ES_Jan08.pdf
- U.S. Department of Education. (1983). *A nation at risk: The imperative for educational reform*. Retrieved from <http://www.ed.gov/pubs/NatAtRisk/index.html>
- Villegas, A. M., & Irvine, J. (2010). Diversifying the teaching force: An examination of major arguments, *The Urban Review*, 42(3), 175–192. doi: 10.1007/s11256-010-0150-1
- Villegas, A. M., & Lucas, T. F. (2004). Diversifying the teacher workforce: A retrospective and prospective analysis. *Yearbook of the National Society for the Study of Education*, 103(1), 70–104. doi: 10.1111/j.1744-7984.2004.tb00031.x
- Wenglinsky, H. (2002, February 13). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives*, 10(12). Retrieved from <http://epaa.asu.edu/epaa/v10n12/>
- Wilson, S., Floden, R., & Ferrini-Mundy, J. (2002). *Teacher preparation research: Current knowledge, gaps, and recommendations*. Center for the Study of Teaching and Policy, University of Washington. Retrieved from <http://depts.washington.edu/ctpmail/PDFs/TeacherPrep-WFFM-02-2001.pdf>
- Wineburg, M. S. (2006). Evidence in teacher preparation: Establishing a framework for accountability. *Journal of Teacher Education*, 57(1), 51–64. doi:10.1177/0022487105284475
- Yinger, R. J., Daniel, K. L., & Lawton, M. (2007). *The Teacher Quality Partnership (TCP) Research Enterprise: Enabling systemic understanding and improvement*. Paper presented at European Association for Research on Learning and Instruction, Budapest, Hungary.
- Yinger, R., & Hendricks-Lee, M. (2000). The language of standards and teacher education reform. *Educational Policy*, 14(1), 94–106. doi:10.1177/0895904800014001008
- Zeichner, K. (2010). Competition, economic rationalization, increased surveillance, and attacks on diversity: Neo-liberalism and the transformation of teacher education in the U.S. *Teaching and Teacher Education*, 26(8), 1544–1552. doi:10.1016/j.tate.2010.06.004

APPENDIX



AUDREY AMREIN-BEARDSLEY is an associate professor at Arizona State University. Her research interests include educational policy, educational measurement, and research methods, and more specifically, high-stakes tests and value-added methodologies and systems. She was recently named as one of the top 121 edu-scholars in the nation, honored for being a university-based academic who is contributing most substantially to public debates about the nation's educational system. She is also the creator and host of a show, titled *Inside the Academy*, during which she interviews some of the top educational researchers in the academy (<http://insidetheacademy.asu.edu/>). Amrein-Beardsley, A. (2012). Value-added measures in education: The best of the alternatives is simply not good enough [Commentary]. *Teachers College Record*. Amrein-Beardsley, A., & Collins, C. (2012). The SAS Education Value-Added Assessment System (SAS® EVAAS®) in the Houston Independent School District (HISD): Intended and Unintended Consequences. *Education Policy Analysis Archives*, 20(12).

JOSHUA H. BARNETT is the director of research and evaluation for the National Institute for Excellence in Teaching (NIET). His primary research focus is improving the quality of education for all students through reforming how teachers and principals are evaluated and how they are compensated. His additional interests include fiscal issues around equity, adequacy, and program evaluation. He has worked as a co-principal investigator on large-scale federal projects, as well as across districts and states to help construct educator evaluation systems and professional development programs. He has developed numerous program evaluations at the school, district, and state level. Before joining NIET, he worked at Arizona State University and was a Rotary Ambassadorial Scholar at Massey University in New Zealand. His work has been recognized by invitations to be a keynote speaker to a variety of audiences and present his work through numerous national conference presentations and publications, including: Ritter, G. W. & Barnett, J. H. (2013). *A straightforward guide to teacher merit pay: Rewarding and encouraging schoolwide improvement*. Thousand Oaks, CA: Corwin. Barnett, J. H. (2012). March Madness and the Inequity Conundrum [Commentary]. *Teachers College Record*.

TIRUPALAVANAM G. GANESH is an engineer and education researcher with research interests in teacher education, learning environments, and study of K-12 and higher education systems. He has an interdisciplinary PhD with an emphasis in curriculum and instruction. He served as co-principal investigator and assistant dean for information systems (2006–2010) in the College of Education at Arizona State University when the university began the study reported in this article. He is Research Associate Professor and Assistant Dean, K-12 Education at the Ira A Fulton Schools of Engineering at Arizona State University. He is developing a non-profit organization to support educators, school systems, and families with the aim of fostering innovation in teaching and learning. In January 2012, he presented to the National Academy of Engineering's iSTEM committee on the idea of integrated Science, Technology, Engineering, and Mathematics education (see <http://www.nae.edu/File.aspx?id=55867>). His scholarship includes: Ganesh, T. G. (2011). Children-produced drawings: An interpretive and analytic tool for researchers. In E. Margolis, & L. Pauwels, (Eds.), *The Sage handbook of visual research methods*. London, UK: Sage. Ganesh, T. G. (2007, April). Commentary through visual data: A critique of the United States school accountability movement. *Visual Studies*, 22(1), 42–47. Ganesh, T. G. (2002, Fall). Held hostage by high-stakes testing: Drawing as symbolic resistance. *Teacher Education Quarterly*, 29(4), 69-72.