Driving Blind:
Why We Need Standardized Performance Assessment
In Teacher Education

Charles A. Peck
University of Washington

Marcy Singer-Gabella
Vanderbilt University

Tine Sloan
University of California, Santa Barbara

Susan Lin
University of Washington

Abstract

In this article we argue that standardized teaching performance assessments (TPAs) offer a uniquely valuable resource for learning and improvement of practice in teacher education. The affordances of TPAs as opportunities for learning are identified at four levels, including those for teacher candidates, individual faculty, organizational learning at the program level, and organizational networks that span program boundaries. We conclude that TPAs can provide motivation and direction for continuous program improvement efforts, contribute to the development of a common and concrete language of practice, and accelerate the professionalization of teaching.

Contemporary policy discourse in teacher education is dominated by concerns, and in many cases outright skepticism, about the efficacy of programs designed to prepare new teachers. These concerns have been reflected in a wave of state and federal program accountability policies, many of which have been accompanied by rhetoric aimed at identification and elimination of *low performing* programs (Crowe, 2010). Whatever the merits of such a policy objective (and we believe it to be highly problematic), our focus in the present article is on a second objective of these policies,

which has to do with the ways in which program outcome data might afford opportunities for learning and program improvement.

The idea of building strong feedback systems that allow teacher preparation programs to systematically use program outcome data to improve their effectiveness is hardly a breathtaking theory of action. Nevertheless, in one review of the status of research and reform in the field of teacher education, Humphrey et al. (2000) concluded:

> The evaluative frame of mind has not yet penetrated teacher education. On the basis of available research, we can describe what has been undertaken in the name of reforming teacher preparation during the past 15 years. However, it is nearly impossible to describe or summarize whether the undertakings have been effective. (p. 30)

While external pressures for evidence-based program improvement in teacher education have increased considerably since the time of this review (cf. new CAEP Standards, 2013), the field remains highly dependent on locally designed measures for evaluating candidate learning and program outcomes that are of unknown reliability and validity and of very limited value in making decisions about program change (Feuer, Floden, Chudowsky, & Ahn, 2013). One interpretation of this state of affairs is that programs of teacher education are essentially *driving blind*, operating without trustworthy measures of program outcomes to guide efforts at improvement of practice.

In this article, we address this problem by making an argument for the unique affordances of one specific type of program outcome measure as a tool for improvement of teacher education: standardized performance assessments of teaching. In doing so, we do not intend to imply that other types of outcome measures (e.g., graduate and employer satisfaction surveys, placement and retention studies, value-added measures of P-12 student achievement) cannot be used in sensible ways as tools for evaluating program quality. On the contrary, we follow others in observing that no single measure is by itself an entirely adequate means of evaluating the effectiveness of individual teachers (Cantrell & Kane, 2013), much less the quality of a teacher preparation program (Feuer et al., 2013). Our claim, however, is that standardized teaching performance assessments (TPAs[1]) are uniquely valuable with respect to the role that they can play in both motivating and guiding concrete actions aimed at program improvement (Darling-Hammond, 2010; Peck & McDonald, 2013).

Unique Contributions of Teacher Performance Assessments

Several distinguishing features of TPAs are fundamental to their value as sources of concrete and actionable feedback to program faculty, academic leaders, and teacher candidates. Perhaps most important, TPAs are by design aimed at producing

rich and concrete descriptions of teacher performance in the contexts of practical activity (Darling-Hammond & Snyder, 2000). Records of performance produced in actual classroom teaching events, such as lesson plans, video clips of teaching, and samples of P-12 student work, provide concrete and richly contextualized documentation of teaching *practice* that may be directly related to the goals and processes of instruction within programs of teacher preparation. This may be contrasted with more abstract kinds of information yielded by other program evaluation measures, such as satisfaction surveys or value-added measures based on P-12 student achievement. Data from surveys or value-added measures may signal cause for concern in specific program areas – but these kinds of data provide relatively little guidance in identifying the sources of identified problems or strategies for improvement. TPAs also differ in important ways from direct observational measures of classroom interaction (e.g., Pianta & Hamre, 2009), insofar as TPAs attempt to provide more complete accounts of teaching practice, including artifacts of curriculum planning and assessment and evaluation processes, in addition to observational records of interactions between teachers and students. This means that TPAs afford a particularly rich descriptive context for interpreting some of the antecedents (e.g., planning skills) and outcomes (e.g., samples of student work) of instructional interactions between teachers and students.

The affordances of concrete and richly-contextualized information about program outcomes related to candidate teaching practice are particularly relevant to the motivational dynamics underlying program improvement. Evidence from multiple fields of professional practice suggests practitioners are more likely to use feedback when it is referenced to the specific activities for which they have responsibility and over which they have control (Popper & Lipshutz, 1998; Ryan & Deci, 2000; Saptya, Reimer, & Bickman, 2005). For example, while program survey data may indicate that graduates in their first year of teaching do not feel prepared to meet the needs of students with disabilities, this feedback is so general that faculty might reasonably attribute the problem to a wide variety of factors, including problems with methods coursework, problems with classroom management preparation, or problems with modeling of effective practice in field placements. In contrast, lesson plans and video evidence from a TPA might reveal that many candidates are selecting developmentally inappropriate texts and relying extensively on initiate-respond-evaluate questioning routines (Mehan, 1979), leading to the disengagement of students with less developed literacy skills. These kinds of richly contextualized performance data are readily connected to specific methods courses, and consequently are more likely to engage the attention and action of the individuals teaching those courses. Studies of TPA implementation suggest that these kinds of data do indeed arrest the attention of faculty and often lead to substantive program improvement actions (Peck, Gallucci, & Sloan, 2010; Peck & McDonald, 2013; Torgerson, Macy, Beare, & Tanner, 2009).

The importance of concrete and contextualized representations of practice, such as those generated by TPAs, has also been documented in the context of P-12 teacher learning and collaboration work aimed at improvement of practice. For example, Horn and her colleagues (Horn, 2010; Horn & Little, 2010) have carried out a series of detailed examinations of individual and collective learning processes situated in P-12 teacher work groups. Findings from these studies underscore the importance of what Horn and her colleagues refer to as *representational adequacy* in establishing shared and concrete understanding of practice. Hall and Horn (2013) observe that this shared understanding is crucial to the negotiation of change

> because conceptual change relies on inter-subjectivity of participants, representational adequacy becomes an important resource for this work...[O]nce representations are negotiated, they support the reification of new concepts and related practices… [R]eifications have the possibility of circulating into larger networks of practice. (p. 251)

A similar finding related to the function of candidate planning documents, video clips, and samples of P-12 student work as rich and concrete representations of teaching practice has been reported in the context of teacher educators' implementation of TPAs (Peck & McDonald, 2013). These reports suggest, specifically, that joint faculty experiences with training, scoring, and interpretation of candidate work samples facilitate the development of a common and concrete language of practice, which is essential to the negotiation of collective and coherent programmatic change (Peck et al., 2010).

In the following sections, we present several examples of what Pecheone and Chung (2006) have referred to as the *educative* affordances of TPAs, at four levels of practice. First, we describe the ways in which TPAs may afford teacher candidates useful opportunities for analysis and improvement of their own teaching practice (Chung, 2008; Chung & Whittaker, 2007). Second, we identify opportunities for learning that TPAs afford for teacher education program faculty and staff – course instructors, practicum supervisors/coaches, and cooperating teachers – at the level of individual practice. Third, we describe opportunities for collective or *organizational* learning and program-level change that have been reported in accounts of TPA implementation (Peck et al., 2010). Fourth, we illustrate some of the affordances for cross-program and cross-institutional learning that emerge when common TPAs are used across programmatic contexts (Stillman et al., 2013).

Figure 1 depicts the ways in which we conceptualize these opportunities for learning and improvement of practice to be nested in one another. For example, we suggest that evidence of teacher candidate learning documented through a TPA constitutes a context for faculty learning and improvement of practice. Similarly,

individual faculty learning functions as a context which affords faculty opportunities to learn from one another and to negotiate program change.
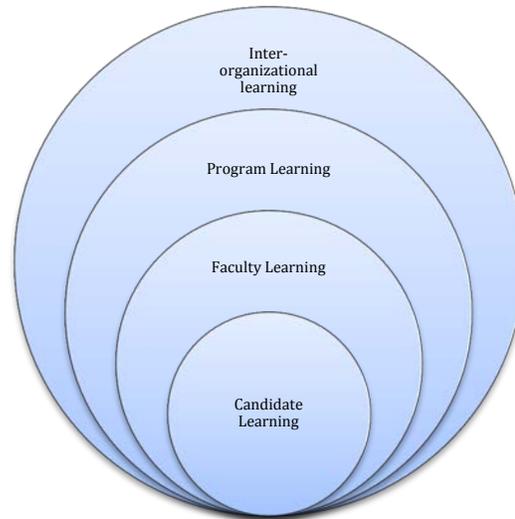

Figure 1. Nested opportunities for learning and improvement of practice afforded through teaching performance assessments.

We conclude by offering some comments regarding the potential of a national teacher performance assessment initiative for building a common and concrete language of practice and argue for the importance of establishing such a language for the professionalization of teaching.

TPAs and In-service and Pre-service Teacher Learning

Studies from several fields of professional education have found that both novice and experienced professionals can learn from participating in performance assessments. For example, in the field of medical education, professional performance assessment tools such as the Virtual Patient Simulation (VPS) and the Objective Structured Clinical Examination (OSCE) have been used to evaluate the clinical competence of novice physicians. In one study on the VPS, the medical students participating in a focus group reported that knowledge from a simulated patient exam was more easily retained than knowledge from textbooks (Botezatu, Halt, & Furs, 2010). In a related study with third year medical students, the OSCE was initially used as a study tool and then as a hands-on assessment of clinical competence (Brazeau, Boyd, & Crosson, 2002). Students reported that the specific and direct feedback they received through the performance assessment process improved their understanding of their strengths and weaknesses in clinical encounters.

In education, studies of practicing teachers who participate in the development of teaching portfolios suggest that substantial learning takes place for them as they analyze and reflect on artifacts of their classroom practices. In one early and influential study, Athanases (1994) investigated the learning that practicing teachers experienced in the context of their preparation of teaching performance portfolios. These teachers noted improvement in their instruction as they expanded strategies for assessing student learning and enhanced reflection about teaching. Athananses' findings about learning processes associated with teaching performance assessments have been replicated at the pre-service level, where a variety of portfolio-based assessment tools have been used to evaluate novice teacher performance. One of these, the Teacher Work Sample (TWS; Schalock, 1998) has been used for several years by a variety of teacher education programs in the United States as both a formative and summative performance assessment for beginning teachers (Devlin-Scherer, Daly, Burroughs, & McCartan, 2007; Kohler, Henning, & Usma-Wilches, 2008). In one study of TWS implementation, Kohler et al. (2008) reported that after developing the Teacher Work Sample portfolio, novice teachers were able to make more sophisticated modifications to their teaching and attributed this to the opportunity the TWS provided them to reflect on their teaching experience.

In a related study of pre-service teacher candidates, Chung (2008) used case studies and focus groups to examine teacher candidate learning in the context of their participation in the Performance Assessment for California Teachers (PACT). Evidence from the case studies indicated that the assessment requirement that candidates independently plan and teach two multi-lesson learning segments in literacy and math contributed to candidates' development of new knowledge about teaching (Chung 2008). The data reported from the focus group echoed the findings from the case study participants. Candidates reported learning from elements of the PACT related to: (1) greater emphasis on assessment of student learning, (2) the opportunity to plan inter-disciplinary lesson units, and (3) intentional reflection on their teaching based on student learning outcomes.

An important dimension of Chung's (2008) study was her attempt to distinguish what candidates reported about learning specifically from their experiences with PACT as differentiated from other sources of learning in their credential program. The new learning that candidates reported that built upon and extended previous program experiences included planning an extended learning segment, learning about individual students, modifying lessons based on assessment of student learning, increasing attention to English Learner students, integrating content areas, attending to content standards, and aligning assessment with instructional plans. Chung (2008) interpreted her findings to suggest that performance assessments such as the PACT "can be useful learning tools to strengthen the professional preparation of new teachers in ways that lead to more learner-centered, assessment-driven teaching" (p. 23).

In a more recent study conducted with teacher candidates at the end of their teacher certification program, Lin (2012) used semi-structured clinical interviews to capture reports of candidate learning related to their completion of the edTPA. Lin used an oral think aloud protocol administered one month after the teacher candidates submitted their performance assessment to capture candidates' description and evaluation of their thinking and learning as they completed planning, instruction, and assessment tasks for the edTPA. Like Chung's (2008) study, the teacher candidates reported learning experiences with the edTPA that integrated and extended many of those they had undergone in their teacher preparation program. For example, the candidates described writing lengthy and detailed lesson plans and commentaries–beyond what any single coursework or practicum assignment in the program had previously required. Candidates reported that this experience deepened their reasoning about instructional decisions, and also deepened their knowledge about their students. One participant offered, "[The] edTPA really pushed me to seek out the special education teacher and find out a lot about individual students and how I could cater to them," and another candidate described her experience this way:

> It made me really think about the idea of building toward one idea and how each lesson is kind of like a sunburst. The main idea is right here in the circle and every lesson is one of the aspects that will eventually build toward the whole. It just really helped me see the benefit of having the lessons be interconnected in a way that the students might not see but then build toward an assessment that integrates all of those concepts. Making sure everything connects towards a finite end.

Taken together, these studies suggest that standardized teaching performance assessments such as the TWS, PACT, and the edTPA, can provide candidates an important opportunity to integrate and extend their knowledge about teaching, deepen their understanding of the needs of individual students, and improve their practice. Both the Chung (2008) and Lin (2012) studies also suggest, however, that these outcomes are not necessarily achieved for all candidates. An important research agenda for teacher educators has to do with identifying the programmatic conditions under which candidates do learn from their experiences with TPAs and ensuring that relevant supports are provided to candidates.

Faculty Learning: From the Individual to the Collective

In addition to pre-service teacher candidate learning, TPAs afford significant opportunities for program faculty and staff learning related to improvement of both individual and collective practice. We offer several examples of this drawn from the Peck et al. (2010) study of PACT implementation. It should be noted that the opportunities for faculty learning and program improvement we identify from this relatively detailed study are consistent with those reported in other studies of TPA

implementation (Bunch, Aguirre, & Tellez, 2009; Peck & McDonald, 2013; Torgerson et al., 2009).

As with teacher candidate learning, many faculty opportunities for learning and improvement of practice are related to the fact that *holistic* assessments such as the TWS, PACT,or edTPA, require candidates to assemble and analyze a record of their actual practice in the classroom. This means that candidates must *integrate and enact* their knowledge of students, curriculum, and instruction in the context of the kinds of complex and dynamic conditions that resemble those they will be expected to manage once they are licensed. The results of these assessments often surprise faculty, whose evaluations of candidate learning seldom reach beyond coursework assignments and projects that focus on relatively narrow areas of candidate knowledge and skill (Peck & McDonald, 2013). The way that performance assessment data create new learning demands for program faculty is reflected in the comments from a field supervisor who participated in the Peck et al. (2010) study:

> I was still at the stage where the special education teacher takes care of (students with IEPs) and (the classroom teacher) doesn't have to. But in my observations of student teachers since PACT, I have tried to focus on the classrooms and identifying kids and their (special) needs. So, in that sense the TPAs have helped me focus on these issues.

Moreover, *collaborative analysis* of TPA data in this study pushed individual faculty to learn more about other pieces of the program and to integrate the work they were doing in specific courses or supervision settings with a larger and more comprehensive understanding of what candidates needed. A foundations instructor articulated the importance of the collaborative analysis process in this way:

> [T]he other change would be the conversations between everyone. Not only the content of the conversations, which revolved around looking at student work, we're all looking at the *same* student work as opposed to supervisors looking at lesson plans and course instructors looking at assignments in courses… before those types of conversations didn't really take place with all of those people— pockets of people, but not everyone.

These kinds of analyses made the whole group realize where candidates were not meeting expectations, and this not only made individuals feel more accountable for their individual practice, but it also served to change the way faculty held themselves responsible for their own learning:

> [W]e looked at the PACT data [during a faculty retreat] and realized that academic language sucks. It was so bad. I stood up and apologized to everybody. And Chris said, "Why should it be your responsibility? It should be all

of our responsibility." So the next year our professional development focused on academic language.

Another instructor later noted the impact of this kind of professional development on her own practice:

> And so when I teach my social studies strategies, I teach the way I have always taught them except that now I use sentence frames; I make sure that they can create the kind of scaffolding necessary you know to support English learners. Then certain kinds of critical terms in history are really pointed out to them. That is the academic language and these are the concepts we want to teach the students.

Learning at the Program Level

Opportunities for learning that are situated in the process of adopting and using a standardized teaching performance assessment extend beyond those we have identified for individual teacher candidates and faculty. In fact, some of the most significant affordances of contemporary TPA instruments such as TWS, PACT, and edTPA may actually emerge at the level of collective or *organizational* learning (Peck, Gallucci, Sloan, & Lippencott, 2009). Existing studies of TPA implementation (Bunch et al., 2009; Peck et al., 2010; Peck & McDonald, 2013; Ruesser, Butler, Symonds, Vetter, & Wall, 2007) are consistent in identifying increased collaboration and learning among faculty, field supervisors, and cooperating teachers as one of the most significant affordances of standardized teacher performance assessment.

For example, Peck et al. (2010) described how the candidate work samples from early pilot testing of PACT were used as a focal point for faculty discussion and collaboration. In one event described in this study, a group of faculty, field supervisors, and cooperating teachers jointly examined work samples collected from a small group of teacher candidates who were participating in the pilot test. Consistent with the observation of Saptya et al. (2005) that practitioners tend to interpret the outcomes of their work through a positive filter, the initial faculty response to the pilot work samples was quite positive, even celebratory: "In initial discussions, both faculty and students expressed a sense that TEP students had been well prepared for the performance assessments" (Peck et al., 2010, p. 457). However, after closer examination of the work samples, one group member reported that "the interesting thing that came up is that people were surprised at in some cases how shallow the answers were" (p. 457). Another commented, "We really do get the sense that their assessment skills are not up to the level I think needs to be in order to respond to the PACT. We agreed we are going to push this more" (p. 457).

Three aspects of this event appeared to be significant with respect to the way close examination of the teacher performance assessment work samples functioned as a context for collective learning and program improvement. First, the analysis of the work samples served to contradict faculty assumptions and beliefs about the adequacy of the preparation candidates had received for doing the performance assessment. Engestrom (2001) and others have observed that the emergence of these kinds of disruptions and contradictions can function as pivotal events in the process of institutional change. Second, the observation that candidate performance did not meet faculty expectations was *shared* across multiple faculty, field supervisors, and cooperating teachers who participated in the review of the work samples—moving this experience into public and collective view within the program. Third, the *qualitative* analysis and evaluation of the direct artifacts of candidates' classroom practice in the work samples led to identification of very concrete implications for action. As one instructor put it, "We are finally looking at student work and really pinpointing some of the areas that need to be dealt with" (Peck et al., 2010, p. 457).

While performance assessment results led faculty and staff in this program to identify specific needs for program improvement, it is important to note that this in itself did not constitute program change. In a separate longitudinal analysis of the data from this project, Peck et al. (2009) described how collective insights arising from the analysis of the performance assessment data were taken up for further action by individuals and small groups of faculty. The most significant aspect of this phenomenon for the present discussion has to do with the ways in which the products of individual learning and proposals for program change, which originated in the analysis of candidate work samples, were then deliberated and acted upon by the larger collective of the program.

These kinds of program level changes did not spring simply from the analysis of the TPA data. Rather, the affordances of the performance data as a resource for program improvement were actualized through the strategic use of organizational policies and practices that supported this kind of program level data analysis, learning, and change (Peck & McDonald, in press). One important example of this was the careful and strategic design of faculty meetings where relevant data were put "on the table" for open analysis and discussion. This required considerable pre-meeting planning and organization of data sets in ways that allowed faculty to focus precious and scarce meeting time on specific high priority issues of concern. Second, program leaders invested considerable authority in small mixed working groups of program members (faculty, field supervisors, and cooperating teachers) that took up specific issues of concern that emerged through the data analysis process. These ad hoc working groups then developed proposals for action that were subsequently presented to and deliberated by the larger group of program faculty and staff. Third, concrete and visible administrative supports for newly emerging forms of collective work were developed and implemented in ways that made visible the value of faculty-led data

analysis, decision making, and program change work (Peck & McDonald, in press). For example, regular coursework and field supervision activities were suspended annually for one week to support program-wide participation in local evaluation of the TPA work samples, and quarterly program-wide retreats were held to provide a venue for small work groups to present their work and recommendations to the larger collective of program members.

In examining the emerging literature on the implementation of standardized teacher performance assessments in programs of pre-service teacher education, we are struck by the recurring reference to the ways in which faculty and staff participation in the training and scoring activities related to evaluation of candidate work samples can foster development of a "common language" among program participants (Peck et al., 2010; Peck & McDonald, 2013; Peck, Muzzo, & Sexton, 2012). This appears to us to be a pivotal aspect of the relationship between teacher performance assessment and programmatic change processes. In their study of interactional processes in P-12 teacher work groups, Hall and Horn (2013) make a similar argument. Their analysis suggests that teachers' negotiation of a shared understanding and interpretation of concrete *representations of practice* is essential if collaborative work is to lead to conceptual change and changes in practice. The relatively rich and concrete artifacts of teacher candidate practice, including examples of work from candidates' own P-12 students that are generated through standardized teacher performance assessments, are an example of the kinds of representations of teaching and learning around which teachers can negotiate a shared understanding of program outcomes and a shared agenda for change.

Learning Across Institutional Boundaries

The learning opportunities afforded by TPAs for teacher candidates, faculty, and programs described above are all situated *within* institutions. In this section, we turn to affordances for learning *across* institutional boundaries. To ground the discussion, we highlight two different cases. In the first, a TPA functioned as an object of shared inquiry among individuals located in different higher education institutions, but who were identified within a single community of practice. In the second, the TPA was taken up by stakeholders from distinct communities of practice who were working toward different ends. For both, learning hinged on supports for collective sense-making about the nature of good teaching as represented in a given TPA support that included the suspension of routine operations to consider alternative ways of working (Hall & Horn, 2012). Unique, however, were the substance of and affordances for this negotiation of practices and purposes.

Case 1: Different institutions, shared identities. In their article, "Putting PACT in Context and Context in PACT," Stillman et al. (2013) position the California PACT assessment as an occasion for collaboration and critical inquiry. Their work was

motivated by a shared concern that PACT lacked a metric to assess candidates' abilities to facilitate learning among diverse populations. In response, they launched a dialogue and ultimately an investigation of how such practice might be represented and assessed, within the data generated by the PACT.

The construct of *contextualizing practice*, which emerged from the authors' initial conversations and analyses of candidate submissions, expresses a commitment to forms of pedagogy that leverage learners' cultural assets and everyday practices as resources for learning. Based on a review of scholarship on teaching diverse groups and extensive negotiation of the views of quality teaching represented, the authors teased out two dimensions within the construct and conjectured a rough developmental progression for each. The first, *ideological clarity*, focuses on teacher candidates' conceptions of their own agency and responsibility for serving all learners well. The second, *pedagogical clarity*, spotlights candidates' enactment of pedagogical practices that capitalize on the assets that learners bring. The authors crossed these progressions to form a matrix and then applied the matrix to analyze a set of PACT submissions from across their campuses that varied by subject area and general score. Thus, the authors grounded their theoretical conceptions and conjectures in rich examples provided by the PACT assessment and also identified troubling patterns across campuses about candidates' tendency to neglect "students' lived experience – as members of families and communities – as providing rich funds of knowledge for school-based learning" (Stillman et al., 2013, p. 17).

The process described by Stillman et al. (2013) parallels what Engestrom (2001) has termed "expansive learning" – learning in which the endpoint is not known in advance and which requires the creation of new tools and/or ways of working. As posited by Engestrom, such learning is driven by contradictions in the activity itself and is supported by a set of strategic moves: questioning, analysis of data, and negotiation of meaning, modeling, and testing the model. The work pursued by Stillman and her colleagues (2013) features these characteristics. Motivated by a perceived contradiction between program values and a high stakes assessment intended to be comprehensive, faculty asked, "does the PACT miss core programmatic values?" Together they analyzed local data and a broad range of scholarship to clarify the construct of contextualizing practice, modeled the construct in the form of a matrix, and tested out the matrix by applying it to PACT submissions.

While expansive learning emphasizes the creative and context dependent work of learning at the boundary, the existence of a standardized tool proved productive to this work in several ways. First, as described in previous cases, the PACT supported the negotiation of a common language and set of images for representing quality teaching; however, in this example, PACT also provided a launching point for describing a construct perceived as *absent* from the PACT rubrics. In testing out the contextualization matrix, faculty participants drew not only on the rich data set

generated through the standardized assessment model, but also on the practice of scoring against a common rubric. This re-appropriation ultimately led to insights that would guide programmatic change. Second, standardization afforded the opportunities for participants from different institutions to expand the area of focus beyond individual programs, to consider features of novice performance—as interpreted through the PACT and contextualization matrix—across multiple contexts. By conducting inquiry across multiple programs, faculty made visible and attended to patterns in performance that transcended individual campuses. Presumably, such a process might also reveal patterns that were rooted in specific institutional contexts.

Case 2: Different institutions, collaborative dissonance. In the preceding case, while participants worked in different institutional contexts, they claimed a common identity as teacher educators. Thus they entered the collaboration with a shared set of practices and purposes including, but not limited to, the use of the PACT as an assessment tool. Our second example spotlights opportunities for learning among stakeholders who work in distinct institutions and whose practices and purposes differ in kind.

In 2008-2009, teacher educators from eight Tennessee higher education institutions – seven public and one independent – moved to adopt the PACT as a strategy for teacher preparation improvement. By the end of that year, these institutions, along with the State Department of Education and the State Board of Education, formed a partnership and committed to the piloting and later field testing of the successor to the PACT, the TPA (ultimately, edTPA). From 2009-2013, university faculty and Board and Department leadership communicated frequently to share progress and troubleshoot challenges as institutions began working with the assessment. As in other cases described in this article, as faculty members and policy makers mapped the alignment with the state professional framework and evaluation tools, they took up the language of the TPA to develop a shared understanding of *effective practice*.

During this interval, two other dynamics became apparent. First, buy-in was increasing across stakeholder groups. Second, stakeholder purposes were becoming more obviously distinct. Faculty teams leading adoption within their university programs positioned the TPA as a lever for program inquiry and redesign. They valued the assessment's alignment with locally valued practices, its sensitivity to differences in context and content, and its potential to provide a rich snapshot of candidates' teaching. The interest of the state policy makers ultimately, however, lay in potential of the TPA as a tool for gatekeeping (i.e., licensure) and/or program evaluation. Board of Education and Department of Education partners ultimately sought a standardized assessment that would effectively discriminate between more and less effective novice teaching, would allow for comparisons across programmatic contexts, and would be predictive of future performance. Hence, the TPA came to function as a *boundary object* (Star &

Griesemer, 1989; Wenger, 2000) which was "both plastic enough to adapt to local needs and constraints of the several parties employing it, yet robust enough to maintain a common identity across sites" (Star & Griesemer, 1989, p. 393). In this context, the edTPA offered a reifying vision of good teaching that was taken up by teacher education faculty and state policy makers as a tool in the pursuit of related but different aspects of the development and certification of teacher quality.

While there is a compelling literature on the hazards of working across boundaries (c.f. Banner, Donnelly, & Ryder, 2012; Kellog, Orlikowski, & Yates, 2006; Oswick & Robinson, 2009), research also suggests that given the right organizational conditions, intersections at the boundary provide the opportunity for collective sense-making and learning (Cobb & McClain, 2006; Engestrom & Sannino, 2010; Wenger, 2000). Indeed some claim that inter-organizational learning is "energized" by the contradictions (Engestrom, 2001, p. 140). So how does risk turn into opportunity?

In Tennessee, two kinds of practices emerged as critical. First, recognizing their different – and conflicting – motivations, partners were vigilant in their efforts to be inclusive and transparent and to cultivate productive dissent. In monthly conference calls that were open to university and policy representatives, the norm evolved that concerns, objections, and disagreements would be voiced, discussed, and acted upon. Second, stakeholders capitalized on the diversity of positional identities and strategic expertise in relation to the TPA to advance learning of the group. As one example, recognizing the substantial overtime that participating campuses were investing in the TPA pilot, the Deputy Director of the State Board of Education (who possesses a deep grasp of the structure and nuances of state policy) successfully advocated for an exception to licensure requirements, allowing teacher education candidates to substitute the TPA for the Praxis II Principles of Learning and Teaching exam. This move not only asserted support for the partnership work, but also advanced the goals of both stakeholder groups. On campuses that took advantage of the exception, candidates and faculty could concentrate on an assessment perceived as more clearly tied to valued learning outcomes. Meanwhile, if more IHEs (institutions of higher education) piloted the assessment, the state could generate sufficient data to determine whether the TPA would be a productive measure for program improvement and evaluation or for the assessment of candidates' readiness to teach.

Akkerman and Bakker (2011) wrote that the transformation of practice requires "dialogue and collaboration between 'flesh and blood partners' at either side of the boundary" (p. 149). In Tennessee, a state in the throes of rewriting policy on teacher licensure and accreditation, the TPA as a boundary object has provided an occasion for such dialogue and collaboration. Critical to this gathering has been its provision of evidence of valued practice (seen by faculty as a resource for program improvement) and the standardized scoring of that evidence (seen by policy makers as a resource for licensure and accreditation policy). The practices forged in this collaboration of openly

addressing conflict and mobilizing expertise distributed across boundaries have enabled stakeholders to establish a hybrid space where accountability purposes remain in play with, but have not yet overwhelmed, program improvement values. Thus, while the ultimate role of the TPA in state licensure policy has not yet been settled, partners have entered into a transformative negotiation about how teacher licensure policy change itself can proceed.

Summary Observations and Recommendations

In this article, we have reviewed evidence that suggests the unique value of standardized teacher performance assessment as a tool for improvement of teacher preparation. We have illustrated the affordances of TPAs in terms of the opportunities for learning that they can offer candidates, faculty, programs, and the field of teacher education. A critical feature of these tools lies in their *standardization* by which we refer to the process through which scorers achieve consistent ratings of candidate teaching performance. We are not naïve about the dilemmas and paradoxes of power, voice, and resistance that inevitably accompany any process of standardization. And we are respectful of thoughtful critiques of standardization grounded in these dilemmas (e.g., Au, 2013). However, we are also not naïve about the extent to which the absence of a common and concrete language of practice operates as a profound barrier to substantive collaboration and coherence within individual programs of teacher education contributes to the ongoing failure of the field to effectively engage perennial problems of connections between courses and fieldwork and inhibits the development of a useful professional knowledge-base for the field. Developing consistent (that is, *standardized*) definitions and interpretive frameworks that can be used to evaluate concrete examples of teaching practice is what allows TPAs to function as a common language of practice and as a tool for communication, collaboration, and improvement of the work of teacher preparation. It is worth noting that such a language may itself be critiqued and amended as needed to support valued outcomes and emergent practices (e.g., Stillman et al., 2013). A common language developed through a TPA need not be a dead language.

Participation, Commercialization, and TPA

An essential process related to the development of any language (professional or otherwise) has to do with the negotiation of agreement on issues of reference. Moreover, languages are not *given,* much less *received,* through a process of simple transmission; rather, they are learned through participation in concrete social activities in which the common meanings of things are established (Kaye, 1982; Vygotsky, 1978). In this context, the direct participation of program faculty and staff in training and scoring activities related to adoption and implementation of a TPA may be recognized to be isomorphic with the process of developing a common and concrete language of practice. Neither national teaching standards (e.g., NTASC) nor surveys of graduate

and employer satisfaction nor *value-added* measures of P-12 student achievement afford opportunities for participation in any kind of process whereby this kind of common and concrete understanding and interpretation of teaching practice might be achieved. This feature of TPAs is pivotal to their affordances as tools for collective learning and program improvement.

The costs of supporting faculty and staff participation in training and scoring for a TPA are not trivial, and contemporary budget pressures have led many program administrators and policy makers to turn to an outsourcing strategy for containing these costs through cooperative agreements with commercial assessment vendors. Whatever the short term advantages of this approach to addressing the costs of implementing a TPA, we believe such a decision must be weighed carefully in the context of its potential impacts on opportunities to learn for faculty, for programs, and for the field. Fortunately, there are emerging examples in which some programs have created organizational supports for faculty and staff participation in TPA training and scoring activities in ways that reduce dependence on external vendors (Peck & McDonald, in press; Peck & McDonald, 2013; Sloan, 2013). A key feature of many of these supports is the suspension of normal working procedures for a period of time to allow new kinds of work to be undertaken (Hall & Horn, 2012). Unfortunately, as Hall and Horn have noted in the context of their study of P-12 schools, these kinds of organizational supports for teacher collaboration and learning are the exception and not the rule in higher education (Peck et al., 2012).

Finally, we return here to our metaphor about *driving blind*. In doing so, we are struck by the relevance of the following observation Sapyta et al. (2005) make about the importance of feedback to practitioner learning:

> Suppose that you wanted to learn archery… If you are learning by yourself or with an instructor and are blindfolded, you have no information about where the arrow lands. In such a situation, you may begin to feel you are performing well. Intuitively, hiring a coach seems to be a better method than depending on trial and error to learn any number of different tasks. However, if neither the coach nor the student can see the target, improvement is limited because of the lack of feedback. (p. 147)

We have argued that analyzing artifacts of teaching practice collected through TPAs can provide exactly the kind of concrete feedback about specific program outcomes that has been shown to have a powerful influence on practice in other fields of human service (Kelley & Bickman, 2009; Popper & Lipshutz, 1998). We have further argued that developing valid and reliable processes for undertaking these analyses entails development of a common and concrete language of practice and that such a language is itself a critical tool for achieving deeper levels of communication, collaboration, and coherence, both within and across programs of teacher education.

Moreover, the recent emergence of a nationally available TPA (SCALE; Stanford Center for Assessment, Learning and Equity, 2013) affords the field of teacher education an unprecedented opportunity to build a shared language and a shared agenda for evaluation and improvement of practice that is responsive to many years of critique from policy makers, scholars… and program graduates. While opportunity is inevitably accompanied by risk, we believe it is time to take the blindfold off and pay closer attention to where we have been and where are going in teacher education.

References

Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research, 81*(2), 132-169. doi: 10.3102/0034654311404435

Athanases, S. Z. (1994). Teachers' reports of the effects of preparing portfolios of literacy instruction. *The Elementary School Journal, 4*, 421-439. doi:10.1086/461776

Au, W. (2013). What's a nice test like you doing in a place like this? The edTPA and corporate education reform. *Rethinking Schools*, *27*(4), 22-27.

Banner, I., Donnelly, J., & Ryder, J. (2012). Policy networks and boundary objects: Enacting curriculum reform in the absence of consensus. *Journal of Curriculum Studies, 44*(5), 577-598. doi:10.1080/00220272.2012.699558

Botezatu, M., Hult, H., & Fors, U. G. (2010). Virtual patient simulation: What do students make of it? A focus group study. *BMC Medical Education*, *10*(1), 91. doi: 10.1186/1472-6920-10-91

Brazeau, C., Boyd, L., & Crosson, J. (2002). Changing an existing OSCE to a teaching tool: The making of a teaching OSCE. *Academic Medicine*, *77*(9), 932. doi: 10.1097/00001888-200209000-00036

Bunch, G., Aguirre, J., & Tellez, K. (2009). Beyond the scores: Using candidate responses on high stakes performance assessment to inform teacher preparation for English learners. *Issues in Teacher Education*, *18*(1), 103-128.

Cantrell, S., & Kane, T. J. (2013). Ensuring fair and reliable measure of effective teaching: Culminating findings from the MET Project's three-year study. Bill & Melinda Gates Foundation. Available from http://www.metproject.org/downloads/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf

Chung, R. R. (2008). Beyond assessment: Performance assessments in teacher education. *Teacher Education Quarterly*, *35*(1), 8-28.

Chung, R. R., & Whittaker, A. (2007, April). *Preservice candidates' readiness for portfolio assessment: The influence of formative features of implementation.* Paper presented at the 2007 annual meeting of the American Educational Research Association, Chicago, IL.

Cobb, P., & McClain, K. (2006). The collective mediation of a high stakes accountability program: Communities and networks of practice. *Mind, Culture, and Activity*, *13,* 80-100. doi:10.1207/s15327884mca1302_2

Crowe, E. (2010).   asuring what matters: A stronger accountability model for teacher education. Washington, D.C.: Center for American Progress.

Darling-Hammond, L. (2010). *Evaluating teacher effectiveness: How teacher performance assessments can measure and improve teaching.* Washington, DC: Center for American Progress.

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education, 16*(5-6), 523-545. doi: 10.1016/S0742-051X(00)00015-9

Devlin-Scherer, R., Daly, J., Burroughs, G., & McCartan, W. (2007). The value of the teacher work sample for improving instruction and program. *Action in Teacher Education*, *29*(1), 51-60. doi:10.1080/01626620.2007.10463439

Engestrom, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work, 14*(1), 133-156.

Engestrom, Y., & Sannino, A. (2010). Studies of expansive learning: Foundations, findings, and future challenges. *Educational Research Review, 5,* 1-24. doi: 10.1016/j.edurev.2009.12.002

Feuer, M., Floden, R., Chudowsky, N., & Ahn, J. (2013). The evaluation of teacher preparation programs: Purposes, methods and policy options. Washington, DC: National Academy of Education.

Hall, R., & Horn, L. (2012). Talk and conceptual change at work: Adequate representation and epistemic stance in a comparative analysis of statistical consulting and teacher workgroups. *Mind, Culture, and Activity, 19*, 240-258. doi: 10.1080/10749039.2012.688233

Horn, I. S. (2010). Teaching replays, teaching rehearsals, and re-visions of practice: Learning from colleagues in a mathematics teacher community. *Teachers College Record, 112*(1), 225–259.

Horn, I. S., & Little, J. W. (2010). Attending to problems of practice: Routines and resources for professional learning in teachers' workplace interactions. *American Educational Research Journal, 47*, 181–217. doi: 10.3102/0002831209345158

Humphrey, D., Adelman, N., Esch, C., Riehl, L., Shields, P., & Tiffany, J. (2000). *Preparing and Supporting New Teachers: A Literature Review.* Washington, DC: SRI International.

Kaye, K. (1982). *The mental and social life of babies: How parents create persons.* Chicago, IL: University of Chicago Press.

Kelley, S., & Bickman, L. (2009). Beyond progress monitoring: Measurement feedback systems in child and adolescent clinical practice. *Current Opinion in Psychiatry, 22*, 263-268. doi: 10.1097/YCO.0b013e32832c9162

Kellog, K. C., Orlikowski, W. J., & Yates, J. (2006). Life in the trading zone: Structuring coordination across boundaries in post-bureaucratic organizations. *Organization Science, 17*(1), 22-44. doi:10.1287/orsc.1050.0157

Kohler, F., Henning, J. E., & Usma-Wilches, J. (2008). Preparing preservice teachers to make instructional decisions: An examination of data from the teacher work sample. *Teaching and Teacher Education, 24*(8), 2108-2117. doi: 10.1016/j.tate.2008.04.002

Lin, S. (2012). *Performance assessment as a learning tool: Teacher candidates' retrospective think aloud of their TPA.* Unpublished manuscript, University of Washington, Seattle, WA.

Mehan, H. (1979). *Learning lessons: Social organization in the classroom.* Cambridge, MA: Harvard University Press. doi: 10.4159/harvard.9780674420106

Oswick, C., & Robertson, M. (2009). Boundary objects reconsidered: From bridges and anchors to barricades and mazes. *Journal of Change Management, 9*(2), 179-193. doi: 10.1080/14697010902879137

Pecheone, R., & Chung, R. (2006). Evidence in teacher education: The Performance Assessment for California Teachers. *Journal of Teacher Education*, 57(1), 22-36. doi:10.1177/0022487105284045

Peck, C., Gallucci, C., & Sloan, T. (2010). Negotiating implementation of high-stakes performance assessment policies in teacher education: From compliance to inquiry. *Journal of Teacher Education*, *61*(5), 451-463. doi: 10.1177/0022487109354520

Peck, C., Gallucci, C., Sloan, T., & Lippincott, A. (2009). Organizational learning and program renewal in teacher education: A socio-cultural theory of learning, innovation and change. *Educational Research Review, 4*, 16-25. doi:10.1016/j.edurev.2008.06.001

Peck, C., & McDonald, M. (2013). Creating "cultures of evidence" in teacher education: Context, policy and practice in three high data use programs. *The New Educator*, *9*(1), 12-28. doi: 10.1080/1547688X.2013.751312

Peck, C., & McDonald, M. (in press). What is a culture of evidence? How do you get one… and should you want one? *Teachers College Record.*

Peck, C., Muzzo, M., & Sexton, P. (2012). Program implementation of the edTPA in Washington State. Unpublished manuscript. Seattle, WA: University of Washington.

Pianta, R., & Hamre, B. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*, 109-119. doi:10.3102/0013189X09332374

Popper, M., & Lipshitz, R. (1998). Organizational learning mechanisms: A structural and cultural approach to organizational learning. *Journal of Applied Management Science, 43*(2), 161-178.

Ruesser, J., Butler, L., Symonds, M., Vetter, R., & Wall, T. (2007). An assessment system for teacher education program improvement. *International Journal of Educational Management*, *21*(2), 105-113. doi:10.1108/09513540710729908

Ryan, R., & Deci, E. (2000). Self determination theory and the facilitation of intrinsic motivation, social development and well-being. *American Psychologist, 55*(1), 68-78. doi:10.1037/0003-066X.55.1.68

Saptya, J., Riemer, M., & Bickman, L. (2005). Feedback to clinicians: Theory, research and practice. *Journal of Clinical Psychology/In Session*, *61*(2), 145-153. doi:10.1002/jclp.20107

Schalock, M. (1998). Accountability, student learning, and the preparation of and licensure of teachers: Oregon's Teacher Work Sample methodology. *Journal of Personnel Evaluation in Education*, *12 (3)*, 269-285. doi: 10.1023/A:1008023412335

Sloan, T. (2013). Distributed leadership and organizational change: Implementation of a teaching performance measure. *The New Educator*, *9*, 29-53. doi: 10.1080/1547688X.2013.751313

Stanford Center for Assessment, Learning and Equity (SCALE). (2013*). edTPA Field test: Summary report.* Palo Alto, CA: Author.

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology. *Social Studies of Science, 19*(3), 387-420. doi: 10.1177/030631289019003001

Stillman, J., Anderson, L., Arellano, A., Lindquist Wong, P., Berta-Avila, M., Alfaro, C., & Struthers, K. (2013). Putting PACT in context and context in PACT: Teacher educators collaborating around program-specific and shared learning goals. *Teacher Education Quarterly*, *40*(4)*,* 135-157.

Torgerson, C. W., Macy, S. R., Beare, P., & Tanner, D. E. (2009). Fresno Assessment of Student Teachers: A teacher performance assessment that informs practice. *Issues in Teacher Education*, *18*(1), 63-83.

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes.* Cambridge, MA: Harvard University Press.

Wenger, E. (2000). Communities of practice and social learning systems. *Organization, 7*(2), 225-246. doi: 10.1177/135050840072002

End Notes

[1] We use the generic term *teaching performance assessment* (TPA) to refer to any of the several standardized assessments of classroom practice currently used across multiple programs of elementary and secondary teacher education. Examples include the Performance Assessment for California Teachers (PACT), the edTPA, the CalTPA, and the Teacher Work Sample (TWS).

About the Authors
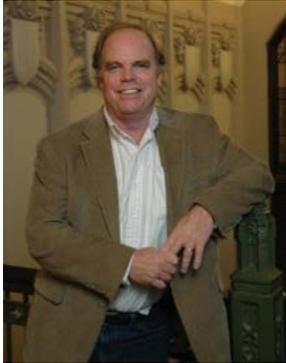
Charles (Cap) Peck, PhD, is currently professor of teacher education and special education at University of Washington. He served as Director of Teacher Education at the UW from 2003 to 2010, during which time the university underwent a major evidence-based program redesign process as part of its participation in the Teachers for A New Era Project (for which he served as Co-Director). Prior to his work at University of Washington , Cap was Director of Teacher Education at University of California, Santa Barbara, and earlier served as Director of the Southwest Washington Educational Partnership for Professional Development—a consortium of 54 urban and rural school districts. Cap's research work over the past decade has focused on policy implementation and systemic change in teacher education. He has been particularly concerned with factors affecting the extent to which programs of teacher education take up opportunities for program improvement that are afforded from new sources of outcome data. Email: capeck@u.washington.edu

Marcy Singer-Gabella, PhD, is a Professor in the Practice of Education and Associate Chair for Teacher Education in the Department of Teaching and Learning at Vanderbilt University. She teaches courses on teaching, educational foundations, and disciplinary literacies in the social studies. Her scholarship focuses on teacher learning and development, including most recently:

- ATLAS, an I3 grant project led by the National Board for Professional Teaching Standards to design video case materials that support the ongoing learning of prospective and early career teachers in grades 3-6;
- preliminary validation research on edTPA, a performance-based assessment of prospective teachers' readiness to teach; and
- NSF-funded Study of the development of elementary teachers' understandings of mathematics and science content and pedagogy, from preparation to practice.

From 1999-2007, Dr. Singer-Gabella served as Executive Director of Project GRAD/Imagine College Nashville. Before coming to Vanderbilt, Dr. Singer-Gabella taught high school social studies in New York, and then worked with the Stanford Schools Collaborative Professional Development Center in the California Bay Area. She earned her A.B. from Harvard University, her MEd from Teachers College, Columbia University, and her PhD from Stanford University. Email: m.singer-gabella@vanderbilt.edu

Tine Sloan, PhD, is currently Director of the Teacher Education Program in the Gevirtz Graduate School of Education at the University of California at Santa Barbara. She is also a Commissioner on the California Commission on Teacher Credentialing, the body responsible for setting certification and accreditation policy for the state. She served on the development team for the Performance Assessment for California Teachers (PACT), hosted the first three implementation conferences, and co-authored the implementation handbook used by the PACT consortium. She also serves as a senior consultant to the Stanford Center for Assessment, Learning and Equity (SCALE), the organization that spearheaded development of the edTPA, a teaching performance assessment used in 34 states. Her primary research interests revolve around teacher education, particularly with respect to use of performance assessment in candidate and faculty learning, and the use of candidate data for program improvement. Prior to her position at UC Santa Barbara, she was a member of faculty at the National Institute of Education, Singapore. Email: tine@education.ucsb.edu

Susan Lin is a doctoral candidate and teaching assistant at the University of Washington. Her research has focused on mobile learning, English Language Learners in rural communities, professional development for teachers, and teacher candidate learning.   Her work in the Elementary Teacher Education Program (ELTEP) is focused on the opportunities for teacher candidate learning through performance assessments. Email: susanlin@u.washington.edu