Published online: 8-30-2019

Performance Assessment Practice as Professional Learning

Vanessa Svihla

University of New Mexico, vsvihla@unm.edu

Tim Kubik

Project ARC, Tim@proj-arc.com

Tori Stevens-Shauger

ACE Leadership High School, tori@aceleadership.org

IJPBL is Published in Open Access Format through the Generous Support of the [Teaching Academy at Purdue University](#), the [School of Education at Indiana University](#), and the [Jeannine Rainbolt College of Education at the University of Oklahoma](#).

Recommended Citation

Svihla, V. , Kubik, T. , & Stevens-Shauger, T. (2019). Performance Assessment Practice as Professional Learning. *Interdisciplinary Journal of Problem-Based Learning*, 13(2).

Available at: <https://doi.org/10.7771/1541-5015.1812>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

This is an Open Access journal. This means that it uses a funding model that does not charge readers or their institutions for access. Readers may freely read, download, copy, distribute, print, search, or link to the full texts of articles. This journal is covered under the [CC BY-NC-ND license](#).

THE INTERDISCIPLINARY JOURNAL OF PROBLEM-BASED LEARNING

SPECIAL ISSUE: UNPACKING THE ROLE OF ASSESSMENT
IN PROBLEM- AND PROJECT-BASED LEARNING

Performance Assessment Practice as Professional Learning

Vanessa Svihla (University of New Mexico), Tim Kubik (Project ARC),
and Tori Stephens-Shaugar (ACE Leadership High School)

ABSTRACT

While performance assessment (PA) is well aligned to project-based learning (PjBL), teachers find it challenging to design and implement PA that is faithful to the authentic context of their projects and viewed externally as rigorous. In contrast to standardizing PA tasks—thereby diminishing authenticity—we formed a research-practice partnership (Coburn, Penuel, & Geil, 2013) that developed and used a “shell” to guide teachers in planning, implementing, and engaging in rigorous dialogues that evaluate and elevate PA practice across four PjBL schools. Drawing from analysis of artifacts and audio-recorded professional development sessions, we highlight how the effort to standardize PA practice while maintaining fidelity to authentic context provided rich opportunities for teacher learning and fostered higher levels of teacher responsibility for assessment.

Keywords: performance assessment; project-based learning; teacher professional development

Introduction and Research Purpose

In both practice and research, there is increasing consensus that project-based learning (PjBL) benefits students (e.g., Geier et al., 2008), especially when learning targets include application of content and skills. There is also acknowledgment of alignment between PjBL and performance assessment (PA) (Lenz, Wells, & Kingston, 2015). In the United States, under the 2015 Every Student Succeeds Act (ESSA), states are encouraged to incorporate PA to measure complex learning, creating more opportunities for PA use (Darling-Hammond, 2017); however, these efforts, anchored to accountability structures, have focused on scaling. While scaling PA offers a promising alternative to traditional standardized testing, we argue such approaches remain limited in meeting students' learning needs for authentic and relevant PjBL (e.g., Newmann, King, & Carmichael, 2007). Opportunity youth remain the most vulnerable and mismeasured when assessments fail to be relevant and authentic, as students do not invest effort in them in a way that demonstrates their complex understanding of material.

In this paper, we report on a research-practice partnership (RPP) that developed authentic PA practice at ACE Leadership High School (ACE), a school with a social justice mission to reach opportunity youth—not through training to vocational standards alone—but to graduate leaders in the construction profession who will be collaborative, client-driven, design thinkers.

Conceptual Framework

We draw on our perspectives as researcher, school leader, and education consultant and situate our research at the nexus of authentic assessment, performance assessment, and teacher professional learning. To develop a contextualized understanding of these, we consider concerns over validity and scalability, and how these have shaped PA. We frame these considerations through the lens of teacher responsibility. However, because responsibility can manifest in multiple ways and lacks a commonly agreed upon definition (Helker & Wosnitza, 2014; Holdorf & Greenwald, 2018), we build on past research to consider three levels of teacher responsibility, all of which include a sense of responsibility to someone or

something beyond themselves: (1) responsibility as compliance-based accountability, primarily to federal, state, and/or local agencies; (2) responsibility as commitment or dedication to completing a task in a way that will be perceived as dependable and trustworthy by stakeholders—including federal, state, and/or local agencies, but especially students, peers, families, communities, and so on; and (3) in a forward-looking, purposeful manner, responsibility as taking initiative for or being receptive to additional related tasks.

Background

Despite myriad references to tests being “valid,” validity is not a property of any test instrument, but rather how it is interpreted—a measurement concern—and how it is used—a prediction concern (Messick, 1989). With ever-growing prediction concerns over how high-stakes, standardized assessments have been used, researchers and practitioners alike have sought to scale PA feasibly. Stanford’s Center for Assessment, Learning, and Equity (SCALE) led much of this work, joined by organizations like the Deeper Learning Network and Asia Society’s International Studies Schools Network (AS/ISSN). Initially, this effort focused on developing PA shells—blueprints to provide conceptual guidance on PA design (Solano Flores, Shavelson, & Schneider, 2001). After many attempts to develop a generic shell, SCALE gradually shifted toward designing a bank of tasks with teachers (SCALE, 2018). More flexible than SCALE’s bank of tasks, AS/ISSN created 16 general shells accompanied by rubrics and educative materials (Davis & Krajcik, 2005)—resources to inform teachers about specific global issues that are the focus of their PA tasks (Asia Society, 2016).

In such efforts, PA is seen as a collection of structured and standardized tasks; teachers score students’ responses—their process or products—using specific criteria (Stecher, 2010). Teachers calibrate by looking at “common pieces of student work” (Research for Action, 2014, para. 3). Calibration is central to this process, because scoring is challenging and effortful. By standardizing the tasks and looking at common pieces of work, it is much easier to achieve reliability, provided teachers are knowledgeable about the skills and content being measured, have clarity about levels of performance and a scoring guide, and participate in training on scoring (Stecher, 2010). This process places responsibility on teachers to evaluate student work with fidelity, making them responsible for the accuracy of their scores and accountable to those external to the school. In such settings, teachers are less likely to display responsibility as dedication or initiative (Christophersen, Elstad, & Turmo, 2014).

When teachers have greater autonomy and opportunities for professional learning—a common characteristic of PjBL schools—they are likely to display responsibility as

dedication to supporting learning and responsibility as sense of purpose (Matteucci, Guglielmi, & Lauermann, 2017). Research suggests that involving teachers in the PA design process and providing them with high-quality professional development can help them improve their understanding of learning standards (Finch, 2016), foster a sense of ownership over assessment, and enhance the chance that they will use the assessments as intended (Palermo & Thomson, 2018). For instance, a study of large-scale PA involved teachers from Tennessee in writing and reviewing cognitively complex, constructed response items and then identifying student responses that could serve as exemplars on rubrics (Palermo & Thomson, 2018). Most teachers reportedly found the professional development useful and planned to use more cognitively complex, constructed response items in their formative assessment practice. We see this as an example of how assessment can shape instruction. In this case, professional development shifted teachers toward a form of test preparation that represented an improvement over what would be seen with multiple-choice exams.

However, increased emphasis on scaling PA and concerns that teacher-created PA tasks can vary in quality together have tended to lead toward the creation and use of standardized tasks (Wei & Cor, 2015) or of standardizing professional learning communities (Darling-Hammond & Falk, 2013). For instance, a policy analysis of 12 states that implemented PA revealed the centrality of investing in teacher capacity (Stosich, Snyder, & Wilczak, 2018) through communities of practice that support teacher learning about task design, data analysis, and improved instruction (Darling-Hammond & Falk, 2013). While we would not disagree that such communities of practice are important in PA at scale, the focus on standardization means that, as with standardized tests, teachers once again have limited opportunities to address equity in communities of opportunity youth.

Building on this, we consider PA along a continuum (Table 1), anchoring to characterizations of assessment of, for, and as learning (Earl, 2012) and key considerations for teacher professional learning, including task design, data analysis, and aligning assessment with instruction (Brown & Mevs, 2012). Who designs PA tasks and how they are scaffolded to undertake this design has implications for teacher learning, responsibility, and validity. Brown and Mevs (2012) note that while it may be tempting to use externally designed PAs, it is a “profound mistake” to do so, even though engaging teachers in designing PA necessitates significant professional learning about PA design and implementation, data literacy, and translation of data into instructional decisions (p. 24). While standardized PA tasks present a feasible alternative to traditional assessments, they risk losing their authenticity for students (Darling-Hammond & Adamson, 2010); this

in turn may reduce the validity of the results. When making an argument that results are valid for a particular use (Mislevy, Steinberg, & Almond, 2003), how students engage with the test should be considered (APA, AERA, & NCME, 2014). When students perceive an assessment as authentic, they invest more effort, and the results provide a more valid account of what students know and can do (Gulikers, Kester, Kirschner, & Bastiaens, 2008). Thus, we argue here that when working with opportunity youth in PjBL settings, authenticity is paramount to student learning.

However, the term *authentic assessment* has been defined in myriad ways. We adapt a definition based on a metasynthesis (Frey, Schmitt, & Allen, 2012), and foreground the importance of validity in terms of use, including from the

point of view of students and community partners (Moss, Girard, & Haniford, 2006; Newmann et al., 2007): authentic PA is jointly relevant to the student and recognizable to an authentic public audience; it provides formative feedback from experts in and outside of school en route to mastery of contextual and cognitively complex skills and content that have value beyond school. As such, we agree that PA quality can be assessed by considering task authenticity, cognitive complexity, relevance, fairness, transparency, educational consequences, directness, reproducibility, and comparability (Baartman, Bastiaens, Kirschner, & Van der Vleuten, 2006).

This stance means teachers face challenges in designing and scoring PAs, especially given that there is a general agreement that they have much learning to do when it comes

Objective	Assessment as Learning	Assessment for Learning	Assessment of Learning
Who designs the PA?	Classroom teachers	Classroom teachers designing with external tools (including performance tasks)	Consultants, experts, and/or teams representing external organizations
Intended end-user(s)	Students of designing teachers	Students, often those part of a particular initiative	Teachers, policy makers
Authenticity of learning experiences	Alignment to real-world practices embedded in school culture	Alignment to external standards reflective of aspirational school culture	Alignment to external standards irrespective of school culture; enhances likelihood that measurements are objectively valid for comparisons.
Relevance to teacher professional learning	Learn to design PA as embedded within PjBL teaching practice	Learn to select/adapt PA from a bank of tasks with a clear idea of what constitutes poor and good performance	Bank of tasks provides clear idea of what constitutes poor and good performance
Relevance to student learning experiences	Identifying students' assets and opportunities for growth	Identifying deficits as opportunities for teacher-led instruction	Identifying deficits as part of corrective program evaluation
Assuring validity among complex possibilities	Selecting and documenting evidence of student learning enhances engagement and likelihood that results are ecologically valid	Selecting and training around exemplars of evidence enhances likelihood that measurements are valid for comparisons and adjusting instruction	Selecting and training raters with sufficient knowledge of the learning targets being measured assures the rating criteria are applied objectively

Table 1. Continuum of authenticity and professional learning opportunities from performance assessments.

to the question of assessment and PA in particular (Gerber, 2018; Greenberg, 2012; Learning Forward, 2011). We therefore consider the kinds of collaborative teacher professional learning that are possible when using such assessment practices, in contrast to traditional notions of training and fidelity to an external model (e.g., Davis & Krajcik, 2005).

Teachers have the capacity to make valid and reliable judgments about their students' work on assessments, provided the assessments either include clear specification or directly measure the content or skills intended, when teachers are knowledgeable about the content and skills being measured (Perry & Meisels, 1996; Südkamp, Kaiser, & Möller, 2012), and when they are asked to evaluate student work in terms of learning (Harlen, 2005). In fact, equitable use—and therefore, validity—of PA depends on involving teachers in the design, implementation, and evaluation of assessments as a means to support student learning (Darling-Hammond, 1994).

Providing opportunities to reflect on PA practice with other teachers in the same school can support the development of a school-wide culture of assessment as learning (Grob, Holmeier, & Labudde, 2017). As part of this, teachers need opportunities to try out, adapt, and reflect on PA practices and tools (Shepard, 1997). Aligning professional learning opportunities to the desired teaching approach—in our case, turning PA practice into PjBL for teachers—can deepen teacher understanding of the approach (Salinitri, Wilhelm, & Crabtree, 2015).

Research Design and Questions

We report on the design, refinement, use of, and learning related to the PA shell, developed in part to protect PjBL practice in schools that serve opportunity youth. The research was conducted as a research-practice partnership (RPP) (Coburn et al., 2013; Penuel, Coburn, & Gallagher, 2013) between a university researcher with expertise in PjBL and assessment, an education consultant with expertise in PjBL and professional learning, and a school leader with applied expertise in PjBL design. As an RPP, the goal was to address a persistent problem of practice across a small network of schools through collaborative commitment to iterative design and reflection (Coburn et al., 2013; Penuel et al., 2013)—in this case, to develop a rich, robust and rigorous approach to PA that could be used to counter dominant narratives about school failure, teacher ineffectiveness, and student deficits. Amidst external efforts to create standardized PA tasks, the network school principals felt it was important to design PAs that students would care about; they feared that without that care, students would engage much as they did with traditional standardized tests, treating them as a foregone failure.

As a result, the principals felt that such assessments did not provide a valid measure of what their students actually knew and could do.

We examine how the development of a PA shell and its implementation fostered higher levels of teacher responsibility for assessment, at a point when accountability, evaluation, and standardized testing had contributed to teaching having an increasingly de-professionalized status. We sought to standardize a process of PA design, implementation, documentation, and evaluation that fit with the contextual PjBL practice we sought to protect. To support teachers, we provided opportunities for them to use the PA shell and reflect on that use. Our research was guided by the following questions:

- Given the focus on standardizing PA tasks at the time, how did the RPP shape and maintain the vision and practice of PA in authentic contexts?
- How did organizing professional practice around the PA shell foster higher levels of teacher responsibility over assessment?
- How did external accountability efforts, including those related to standardized PA tasks, influence teachers' understanding of PA and their PA practice?

Methods

Detailing the evolution of the RPP and the process of PA shell development and its impact is beyond the scope of a single article. Yet, understanding the overall context is critical to making sense of how the RPP led to sustained and expanded PA practice, as depicted in Figure 1.

Setting, Participants, and Project-based Approach

In this study, we focus in particular on ACE Leadership High School, a not-for-profit charter high school formed to jointly serve opportunity youth and address industry partners' anticipated need for employees. Dr. Kubik's professional coaching with the Buck Institute for Education (BIE) shaped ACE's initial standards-based PjBL approach, but guidance from industry and community partners encouraged more authentic qualities in their projects. Industry partners recognized that traditional vocational schooling was not sufficient—they needed graduates with stronger design skills who could transform the industry. This led the principal (Ms. Stephens-Shauger, an author of this paper) to work with Dr. Kubik (an education consultant, also an author) to modify BIE's approach by placing standards in the real-world contexts of potential clients, rather than real-world skills in the academic context of a core content area course.

Figure 1. Timeline of ACE and PA shell development, use, and expansion.

Ms. Stephens-Shauger led most of the weekly PD sessions and organized four weeks of annual PD, while consultants like Dr. Kubik provided professional learning to meet the needs of teacher inquiry. A key focus of PD was the design and tuning of projects to align with industry and community partners. Ms. Stephens-Shauger oversaw curriculum through the lens of PD, supported by two other school leaders responsible for student support and community partnerships. This team approach was repeated in projects, which were team-taught by two or three teachers. During the initial period of data collection, the staff included 15 teachers and 8 staff related to support and engagement for approximately 200 students, most of whom were off-track to graduation and reengaging after dropping out or being habitually truant.

Data Collection, Selection, and Analysis

We documented the five-year process of PA development, use, and refinement through versioning, interviews, field notes, artifacts, and audio recordings. Dr. Svihla was embedded in the school for nine months across two years, during which she conducted participant observation (DeWalt &

DeWalt, 2010). Dr. Kubik made many trips to the school to provide inquiry-based workshops he codesigned with Ms. Stephens-Shauger to ensure they aligned with the data teachers were collecting, including serving as a critical friend during the PA development process.

We created a data corpus by searching our work and research records. The corpus included data from 55 unique events, such as visits to the school and PD sessions. We analyzed emails, field notes, agendas, and other written artifactual data using content analysis (Saldaña, 2015). We created a detailed 33-page timeline to identify salient and critical moments. From these, we selected data to transcribe, particularly events that served as opportunities for teachers to consider the purpose of PA and reflect on their use of PA. We analyzed these using tenets of interaction analysis, especially participation structures (Jordan & Henderson, 1995) and markers of ownership and references to external accountability. We conducted analysis iteratively, gradually refining early insights into themes related to authenticity, responsibility, and accountability.

Results and Discussion

We highlight the potential of PA to serve as professional learning amidst external accountability pressures. We draw key inferences about the process that led to higher levels of teacher responsibility for assessment.

Shaping and Maintaining the Vision and Practice of Authentic Performance Assessment

In 2012, realizing teachers needed to begin rethinking assessment, Ms. Stephens-Shaugar asked Dr. Kubik to provide coaching on identifying evidence of learning. Together, they designed a simple template that helped teachers link evidence in student work to specific outcomes. Around this time, another consultant introduced Ms. Stephens-Shaugar to the New York Performance Standards Consortium. After much researching, reading, and reflecting on the practice in her own school, Ms. Stephens-Shaugar sent out a call to teachers and schools she thought might be interested in forming a performance assessment network (PAN), noting that the Consortium had “very good results” despite that “every school in the Consortium is different.” After mentioning that the state education department approved a pilot project, she urged others to join: “We can’t do it alone.”

Representatives of nine schools attended the first meeting, where they discussed three articles about authentic assessment and validity (Newmann et al., 2007; Pierce, 2012; Wehlage, Newmann, & Secada, 1996). Ms. Stephens-Shaugar continued to highlight the success of the Consortium while referencing literature on authentic assessment (Darling-Hammond, Ancess, & Falk, 1995). We draw attention to this as a contrast. While the Consortium’s tasks offer an alternative to traditional testing, they use standardized tasks. In contrast, the PAN members considered questions such as “Are we involving professionals as an authenticity filter?” Thus, involving teachers in PA design can keep a focus on authenticity, as others have suggested (Brown & Mevs, 2012).

With this authenticity frame in mind, Dr. Svihla drew inspiration from PA quality and validity-as-argument definitions (Baartman et al., 2006). She created a draft PA shell, which included guidance about what a PA is, a metadata section for school context information, and three sections: Section 1 is completed by teachers to guide PA design; Section 2 is completed by teachers during/after the PA to report details of its use; Section 3 is completed by someone external to the project. Section 1 included a timeline for the PA, a request for attachments (description of the PA, learning objectives) and checkboxes related to audience, format, feedback and transparency, and reliability. Section 2 included a request for attachments (documents, quizzes, rubrics, examples of student work, and a description of

modifications/changes), and checkboxes related to public presentation, authenticity, and cognitive complexity. Section 3 included specific assessments of the evidence included for these, along with global assessments of authenticity, public quality, mastery, and school context.

During this initial development, Ms. Stephens-Shaugar and Dr. Svihla also met with an official from the state’s education department who was developing standardized PA tasks. This official invited us to the two teacher workshops intended to develop PA tasks. The graduate students who documented that process reported that although a number of ideas from our view of PA were presented, when teachers—predominantly from traditional schools—were asked to develop potential PA tasks, concern about the effort in scoring such tasks led many to create multiple-choice assessments.

By contrast, when guiding teachers to review one another’s work by completing Section 3 of the PA shells at ACE in 2015, we oriented them by emphasizing questions like, “Is this meaningful work? Is this something that is a real-world practice? Are these things that we would expect students to do in the workplace? Are they intellectually authentic tasks?”

Members of the RPP maintained their focus on authenticity by referencing publications on authentic assessment and considering a diversity of school contexts linked to specific industries. While standardized PA tasks could be created for simplified professional practices (e.g., calculating the area of a room to know how much floor tile would be needed), Ms. Stephens-Shaugar had established a vision that such tasks would fall short of demonstrating student understanding of professional practices in the context of real-world clients. Both teachers and students were held accountable to this vision at end-of-project public exhibitions, attended by industry partners who wanted to know if students could think about the challenges these partners faced in their fields.

Fostering Teacher Responsibility Over Assessment

The first draft of the PA shell was introduced to teachers in late summer 2013. Dr. Svihla was apprehensive that the teachers would be resistant, because so much about assessment was prescriptive, external, and punitive. She feared that the focus on responsibility as compliance would make them defensive. Overall, teachers responded positively, noting only minor comments for revision. When a few teachers raised concerns that the PA shell implied they were “doing PA wrong,” other teachers responded that the tool was flexible and “you just make choices.”

We invited teachers to use the PA shell to guide their planning, implementation, documentation, and evaluation of their own PA practice. We provided no specific training and very little guidance other than the language of the PA shell itself. In this way, we jointly anticipated the first use

Figure 2. Teachers' commitments to improved documentation.

would result in underdocumentation and viewed this as a chance to foster higher levels of responsibility. Indeed, when the teachers reviewed one another's PA shells in early 2014, they quickly realized they had not documented enough and committed to documenting more (Figure 2). They owned the need to document various forms of data that could show whether students were making progress. In this way, we see responsibility displayed as a commitment to a form of PA practice that could be viewed by stakeholders as trustworthy.

Rather than confront underdocumentation as an issue that could be resolved by training, Ms. Stephens-Shauger approached it as an opportunity for sustained professional learning inquiries in collaboration with Dr. Kubik and other consultants. For example, when we asked teachers to assess using evidence, they wondered, "What is adequate evidence?" "Does it have to look a certain way?" In order to help them answer such questions, we needed to give them practice with various types of evidence. In gaining that experience, they developed confidence and autonomy as PjBL professionals. To further build this capacity, Dr. Kubik worked with Ms. Stephens-Shauger in the spring of

2014 to design an inquiry-based workshop on evidence they could collect, guided by the question, "What should assessment look like when student needs go beyond our rubrics?" Teachers worked collaboratively to compare qualitative and quantitative evidence and to distinguish direct from indirect evidence. They reported that "communicating about the evidence we capture" was important in order to "learn practices from one another." Teachers, recognizing that their questions had been heard, displayed responsibility as initiative in their own inquiry-based professional learning. This, in turn, brought them to an increased—if not yet perfect—understanding of how to work with the challenges of documentation presented by the PA shell. Ms. Stephens-Shauger noted that teachers developed understanding that PA provided more choice of how to demonstrate student growth within day-to-day assessment practices, and this led to deeper conversations about mastery, including considerations of the kinds of learning experiences that support its development.

During a PD week in 2015, Dr. Svihla and Ms. Stephens-Shauger cofacilitated a workshop, with Dr. Kubik invited by Ms. Stephens-Shauger to "provoke us" in the process of

completing Section 3 of the PA shell. All staff, not just teachers, participated in the workshop, which began with a brief review of the purpose of the PA shell, and Dr. Svihla discussed the growth she observed in their PA practice:

And that's what we sort of expected would happen. It would be hard to tell you, "Here's what you need to document" before you'd kind of gone through it once. So for those who are new, that's an important thing to know, that part of this documentation has come from experience. And so don't be afraid to ask people who've done it before, like, "What are some tips for actually documenting a project well?"

This explanation situated PA practice as peer learning, rather than learning directed by expert consultants providing instruction on "best practices." It also aligned with Ms. Stephens-Shauger's approach to professional learning, based on a belief that it is best to let teachers "get their hands dirty" and then help them gain clarity as they work through it. She sees this as the route to buy-in, but also to enhanced professionalism. Rather than training teachers, she sought the thinking they put into it. Ms. Stephens-Shauger framed the work of completing Section 3 as, "You're not looking at whether or not the students' work that might be included is quality, right? This is all about our own work as assessment development and curriculum development." In this, she invited her staff to take responsibility for assessment design as part of their PjBL practice. As they gained higher levels of responsibility—commitment to and taking initiative in PA—teachers also felt a sense of professional competence when their peers agreed on the ways in which student evidence met the design intentions of their projects. As a result, project documentation became a source of professional pride. There was no need to "get buy-in" on this process, because as it unfolded, the teachers chose to invest in it to further their own professional goals.

By 2016, an external review team summed up this progress as "a robust practice at the school and . . . much of that strength is due to teachers having ownership over developing, implementing, and evaluating Performance Assessments. . . . As they have gotten better at this practice, efforts have become more focused on increasing the relevance and coherence of projects." This does not mean, however, that the teachers are "trained" or the learning is over. As Ms. Stephens-Shauger noted: "When [teachers] are asked to identify outcomes and what evidence they should be archiving, they still struggle. . . . Our conversations about evidence are beginning to be more specific, teachers are becoming more skilled at identifying evidence and understanding what they could be pulling from to capture that evidence."

As we sought to extend PA practice to sister schools, we realized that much of the professional learning we had collectively gained could be unpacked, rather than experienced directly. Dr. Svihla created strictly internal documents—a lexicon and a set of frequently asked questions. These educational materials enhanced teacher ownership over terms related to validity and fostered a sense of belonging because of the internal and therefore insider quality of these documents. For instance, even teachers who had prior experience with the PA shell noted that they gained a deeper understanding of some of the technical constructs, such as what counts as cognitive complexity. They began using some of the terms as they talked to one another during these workshops. Where previously we seldom heard anyone but Dr. Svihla and Ms. Stephens-Shauger use terms like validity and reliability, with the lexicon in hand, teachers began to use these terms to describe what they were doing as they reviewed the evidence gathered by their peers. In one instance, as they discussed the role that external visitors played in projects, they realized the visitors contributed to the "ecological validity" of the PA. They made a deeper commitment to documenting times when they themselves invited community or industry partners to participate. With tools like the lexicon in hand while evaluating their peers' documentation in the PA shell, most teachers were able to draw lessons for their own instructional practice. We see this as evidence of a stance of responsibility as initiative.

External Accountability Shaped PA Practice

Prior to the spring 2015 workshop described above, the school was engaged in a tumultuous rechartering process and faced increased external scrutiny, due largely to the intensified accountability efforts anchored to standardized testing. Amidst these tensions, Ms. Stephens-Shauger prompted Dr. Svihla to explain reliability and validity to her staff during a workshop. In contrast to the efforts to train teachers to score reliably, we agreed that our goal would be to leverage the teachers' *professional vision* (Goodwin, 1994); we trusted that PAs planned, implemented, and documented by these professionals are reliably interpretable when subject to scrutiny by others in their field. As Ms. Stephens-Shauger explained at another point in the 2015 workshop, "the only way we're gonna improve our practice is if [Rich] can give me feedback that's real from his perspective right, from what he sees." By using one of the teachers in the room as an example, she highlighted that they had the expertise needed to evaluate the PA shells. After teachers spent two hours completing Section 3 for four projects, we debriefed the process. Several teachers linked the inadequacy of their documentation to external accountability. For instance, as Mr. Thomas

explained, “I looked at it through the eye of a state [education department] and if that’s what we are basing our—our next five years on—we shouldn’t, they shouldn’t even be giving us three years.” Dr. Kubik drew attention to the types of student work documented in the PA shells, noting that they painted a fairly traditional picture of teaching and learning, and that distinctive practices, like their final exhibitions and work with clients, were missing. (Transcription conventions include // = overlapping talk; capital letters indicate emphasis by speaker; punctuation indicates tone, not grammar).

Dr. Kubik: You’re trying to tell a story about this school, when the school is supposed to be different, but out of an anxiety of your inability to tell that story, the story that you’re telling us is “We’re just the same like everybody else.” Right? //In terms of the evidence//

Mr. Roth: //What’s making that anxiety though//

Dr. Kubik: // Because you’re afraid you’re going to be judged—judged by [the state education department].

Mr. Roth: //like why is it that we’re having such trouble doing it?

Dr. Kubik: Or by outside methods, like standardized tests, right?

Ms. Stephens-Shauger: But how do you put it—but that’s MY fear.

Dr. Kubik: I am supposed to be provocative so//

Ms. Stephens-Shauger: //No you’re bringing something up and—but that—and that is a fear, but that’s MY fear to carry, right.

Rather than shutting this conversation down, Ms. Stephens-Shauger used it as an opportunity for her staff to access and grapple with her own fears as a school leader. In this, we see evidence of her efforts to shape teacher responsibility as dedication and initiative, rather than as external. School leaders can mitigate the effects of external accountability by creating

an environment that fosters trust, giving autonomy to teachers to make decisions, and supporting teacher professional learning (Christophersen et al., 2014; Holdorf & Greenwald, 2018). In this exchange, we note the impact that external accountability played in framing assessment of—rather than as learning. The PA shell served as a *boundary object* between external accountability forces and internal PjBL practices (Star, 2010). Boundary objects sit at the ill-structured edges of communities that lack consensus—here, the state education department and the school—and are worked on or used differently by both communities. The need for such a boundary object was clear to Ms. Stephens-Shauger from the beginning, freeing space in their minds to do the work of designing, implementing, and analyzing student growth reflected in PA. During the three trimesters that followed in the next school year (2015–2016), the school came under increased external scrutiny. The PA shell continued to be a boundary object, used externally as evidence of student performance and internally as a tool for professional learning, supporting teachers to learn beyond what their prior preparation as educators had trained them to do. Approaches like ours position PA as a “growth opportunity for teachers to improve their craft through collaboration with other teachers, while also leading to richer learning experiences for students” (French, 2017, p. 9). By placing this responsibility in teachers’ hands, they become better prepared to meet students wherever they are and support their growth. We have repeatedly observed teachers express pride in the student growth they documented in their PA shells. This restored some of the professionalism lost to standardized assessment practices, while also developing a willingness to be held accountable because of their commitment to engaging in trustworthy PA practice.

Wanting to maintain the authenticity of PA practice while meeting the external desire for quantitative data tied to mathematics and English subject area performance, the teachers developed short-cycle standardized PA tasks that could be completed during advisory, outside of project time. This, in turn, alleviated the felt need to “be the same like everybody else” when documenting their PA practice. This again highlights that teachers displayed responsibility as initiative over assessment.

Concluding Thoughts

As an RPP, we leveraged our collective expertise to design PA practice authentic to the PjBL school it was intended for. At a time when most were focused on standardizing common PA tasks, our approach prioritized teacher professional learning. This reflects previous concerns raised that PA may best be suited to personalizing assessment in local systems, rather than scaling and standardizing (Tung, 2010). As an outlier,

we have contributions to make to the conversation about PA design and teacher professional practice—that uncommon professional learning may result from uncommon PA tasks.

We maintained our commitments to authenticity by referencing research on authentic assessment, posing questions about PA quality from the perspective of community partners, and developing the PA shell by considering validity as an argument (Messick, 1989) rooted in previously identified characteristics of quality PA (Baartman et al., 2006). This allowed us to focus on building PA practice that can be judged as a fidelity to context—rather than fidelity of implementation—approach. We argue that our approach aligns better to PjBL than does the creation of standardized PA tasks and scoring trainings, which, we fear, could lead teachers to feel like quality-control gatekeepers and could inspire standardization of PjBL itself. Dewey (1916, p. 127) recognized long ago that the “vice of externally supplied ends has deep roots” and that ultimately “the distrust of the teacher’s experience is then reflected in the lack of confidence in the responses of the pupils.” In focusing instead on professional learning, we found that teachers responded positively to the PA shell as a way of professionalizing their discourse around assessment of, for, and as learning for their students (Earl, 2012), and for themselves. The teachers displayed higher levels of responsibility as they consistently sought *humane* approaches to rigorous assessment without sacrificing fidelity to contextual PjBL; they took initiative in creating their own set of standardized PA tasks for use in advisory sessions to protect their ability to engage authentic PA within PjBL. In this way, their PA practice can meet the notion that validity should also be concerned with leading to improvements in educational systems (Frederiksen & Collins, 1989).

Given the less authentic nature of standardized tasks, we view these as producing less valid results, especially for our population of students, for whom engagement with the assessment must be considered as a central aspect of validity. We share concerns that standardizing PA tasks puts our students at risk of being labeled failures, when in fact they demonstrate assets in many areas not captured by the standardized approach to scoring those tasks (Zhao, 2018).

Our experience strongly suggests that teachers’ responsibility for, understanding of, and practice of assessment can benefit from engaging in a set of common professional learning practices, such as the PA shell. The same can be said for those engaged in coaching professional learning. Too often, instructional coaches appear on the scene to offer trainings on a practice and then depart to repeat the process over and over again in order to take these practices to scale. In contrast, we found we needed to engage together over several years around problems of practice arising from the PA shell in a way that challenged and enhanced our understanding of

the ways teachers engage students in meaningful PA. For us, this led to insights about the role the PA shell held in project development processes and possible professional development pathways.

Limitations and Future Work

First, we note limitations to the recent research literature that perhaps prompted the call for papers focused on assessment in PBL and PjBL. As we sought to include recently published studies, conducting a systematic literature search, we found that for K–12 settings, there was a significant focus on technology applications from design and implementation to scoring and analytics (Dimopoulos, Petropoulou, & Retalis, 2013; Redecker & Johannessen, 2013; Thomas, 2016). While these may be promising areas of research, we are skeptical that they afford the kinds of interactions we have detailed here. Though many advances have been made with regard to learning analytics approaches, teachers’ abilities to design PAs that depend on such technologies are likely to remain limited. Thus, we see a need for additional research into the impacts such assessments have on teacher learning and professionalism.

Second, our work on PA is but a single instance, carried out in a small network of schools organized around a social justice mission. While this enabled many insights about the potential of PA as professional learning, we acknowledge that many other contextual factors—that are not endemic to most schools—played a role. For instance, as a school leader, Ms. Stephens-Shaugar took seriously her responsibility to engage her teachers in sustained professional learning that reflected the PjBL model they used with students. In doing so, she built a great deal of trust with teachers who came to the school with aspirations for learning experiences similar to those they designed for their students. Such professional learning might not have occurred without the vision and trust Ms. Stephens-Shaugar invested in her teachers. Indeed, there are fruitful opportunities for understanding more about the collegial process of professional learning when the focus is on sustained inquiry into improving student learning, rather than intensive training to implement with fidelity and score reliably. If the ultimate goal for PA is improving student learning outcomes, our findings add value to a conversation that is itself worth taking to scale, while retaining fidelity to the contexts we aim to serve.

Acknowledgments

The authors would like to acknowledge funding from the Spencer Foundation/National Academy of Education. We also acknowledge collaborators Jamie Collins and Abigail Stiles, who assisted with data collection.

References

- APA, AERA, & NCME. (2014). *The standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Asia Society. (2016, September 7, 2018). ISSN workshops and resources: 2016 catalogue. Retrieved September 7, 2018 from <https://asiasociety.org/files/issn-catalog-20160909.pdf>
- Baartman, L., Bastiaens, T., Kirschner, P. A., & Van der Vleuten, C. (2006). The wheel of competency assessment: Presenting quality criteria for competency assessment programs. *Studies in Educational Evaluation*, 32(2), 153–170. <https://doi.org/10.1016/j.stueduc.2006.04.006>
- Brown, C., & Mevs, P. (2012). *Quality performance assessment: Harnessing the power of teacher and student learning*. Boston, MA: Center for Collaborative Education.
- Christophersen, K.-A., Elstad, E., & Turmo, A. (2014). Organisational factors influencing teachers' inner sense of affective commitment to their schools in high- and low-accountability environments. *International Journal of Management in Education*, 9(1), 111–128.
- Coburn, C. E., Penuel, W. R., & Geil, K. E. (2013). *Research-practice partnerships: A strategy for leveraging research for educational improvement in school districts*. New York, NY: William T. Grant Foundation.
- Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64(1), 5–31.
- Darling-Hammond, L. (2017). Developing and measuring higher order skills: Models for state performance assessment systems. *Council of Chief State School Officers (CCSSO)*, 49.
- Darling-Hammond, L., & Adamson, F. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning*. Stanford, CA: Stanford Center for Opportunity Policy in Education.
- Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work*. New York, NY: Teachers College Press.
- Darling-Hammond, L., & Falk, B. (2013). *Teacher learning through assessment: How student-performance assessments can support teacher learning*: Center for American Progress.
- Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, 34(3), 3–14. <https://doi.org/10.3102/0013189X034003003>
- DeWalt, K. M., & DeWalt, B. R. (2010). *Participant observation: A guide for fieldworkers*. New York, NY: Rowman Altamira.
- Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. New York, NY: Macmillan.
- Dimopoulos, I., Petropoulou, O., & Retalis, S. (2013). Assessing students' performance using the learning analytics enriched rubrics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 195–199). ACM.
- Earl, L. M. (2012). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.
- Finch, P. A. (2016). *Teacher-created student growth assessments used for teacher evaluation and the effects on teacher efficacy*. (EdD), National Louis University.
- Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18(9), 27–32.
- French, D. (2017). The future is performance assessment. *Voices in Urban Education*, 46, 6–13.
- Frey, B. B., Schmitt, V. L., & Allen, J. P. (2012). Defining authentic classroom assessment. *Practical assessment, research & evaluation*, 17(2), 1–18.
- Geier, R., Blumenfeld, P. C., Marx, R. W., Krajcik, J. S., Fishman, B. J., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching*, 45(8), 922–939.
- Gerber, N. (2018). Pop quiz: Why should teacher candidates be trained in assessment. Retrieved September 7, 2018 from <https://www.nctq.org/blog/POP-QUIZ:-Why-should-teacher-candidates-be-trained-in-assessment>
- Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96(3), 606–633. <https://doi.org/10.1525/aa.1994.96.3.02a00100>
- Greenberg, J. (2012). What teachers aren't learning about assessment...And we mean any kind of assessment. *National Council on Teacher Quality Blog*. Retrieved September 7, 2018 from <https://www.nctq.org/blog/What-teachers-arent-learning-about-assessment...and-we-mean-ANY-kind-of-assessment>
- Grob, R., Holmeier, M., & Labudde, P. (2017). Formative assessment to support students' competences in inquiry-based science education. *Interdisciplinary Journal of Problem-based Learning*, 11(2), 6.
- Gulikers, J. T. M., Kester, L., Kirschner, P. A., & Bastiaens, T. J. (2008). The effect of practical experience on perceptions of assessment authenticity, study approach, and learning outcomes. *Learning and Instruction*, 18(2), 172–186. <https://doi.org/10.1016/j.learninstruc.2007.02.012>
- Harlen, W. (2005). Trusting teachers' judgement: Research evidence of the reliability and validity of teachers'

- assessment used for summative purposes. *Research Papers in Education*, 20(3), 245–270.
- Helker, K., & Wosnitza, M. (2014). Responsibility in the school context—development and validation of a heuristic framework. *Frontline Learning Research*, 2(3), 115–139.
- Holdorf, W. E., & Greenwald, J. M. (2018). Toward a taxonomy and unified construct of responsibility. *Personality and Individual Differences*, 132, 115–125.
- Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences*, 4(1), 39–103.
- Learning Forward. (2011). *Standards for professional learning*. Oxford, OH.
- Lenz, B., Wells, J., & Kingston, S. (2015). *Transforming schools using project-based deeper learning, performance assessment, and common core standards*. San Francisco, CA: John Wiley & Sons.
- Matteucci, M. C., Guglielmi, D., & Lauermann, F. (2017). Teachers' sense of responsibility for educational outcomes and its associations with teachers' instructional approaches and professional wellbeing. *Social Psychology of Education*, 20(2), 275–298.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement*. New York, NY: Macmillan.
- Mislevy, R., Steinberg, L., & Almond, R. (2003). Focus article: On the structure of educational assessments. *Measurement: Interdisciplinary Research & Perspective*, 1(1), 3–62.
- Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education*, 30(1), 109–162.
- Newmann, F. M., King, M. B., & Carmichael, D. L. (2007). *Authentic instruction and assessment*. Des Moines, IA: Iowa Department of Education.
- Palermo, C., & Thomson, M. M. (2018). Large-scale assessment as professional development: Teachers' motivations, ability beliefs, and values. *Teacher Development*, 23(2), 192–212. <https://doi.org/10.1080/13664530.2018.1536612>
- Penuel, W. R., Coburn, C. E., & Gallagher, D. J. (2013). Negotiating problems of practice in research-practice design partnerships. *National Society for the Study of Education Yearbook*, 112(2), 237–255.
- Perry, N. E., & Meisels, S. J. (1996). *How accurate are teacher judgments of students' academic performance? Working paper series*. Washington, DC.: National Center for Education Statistics.
- Pierce, D. (2012). Performance assessment making a comeback in schools. *eSchool News Special Report*. <http://www.eschoolnews.com/files/2012/01/eSNFeb12SR.pdf>
- Redecker, C., & Johannessen, Ø. (2013). Changing assessment—towards a new assessment paradigm using ICT. *European Journal of Education*, 48(1), 79–96.
- Research for Action. (2014). *Using common assignments to strengthen teaching and learning: Research on the 1st year of implementation*. Retrieved from [https://ldc.org/sites/default/files/research/CAS Year 1 Executive Summary_Sept 2014_RFA.pdf](https://ldc.org/sites/default/files/research/CAS%20Year%201%20Executive%20Summary_Sept%202014_RFA.pdf)
- Saldana, J. (2015). *The coding manual for qualitative researchers*. Los Angeles, CA: Sage.
- Salinitri, F. D., Wilhelm, S. M., & Crabtree, B. L. (2015). Facilitating facilitators: Enhancing PBL through a structured facilitator development program. *Interdisciplinary Journal of Problem-based Learning*, 9(1), 11. <https://doi.org/10.7771/1541-5015.1509>
- SCALE. (2018). Performance assessment resource bank. Retrieved September 7, 2018 from <https://www.performanceassessmentresourcebank.org/bin/performance-tasks>
- Shepard, L. A. (1997). *Insights gained from a classroom-based assessment project*. University of California, Los Angeles: Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies.
- Solano Flores, G., Shavelson, R. J., & Schneider, S. A. (2001). Expanding the notion of assessment shell: From task development tool to instrument for guiding the process of science assessment development. *Revista Electrónica de Investigación Educativa*, 3(1). <https://www.redalyc.org/pdf/155/15503103.pdf>
- Star, S. L. (2010). This is not a boundary object: Reflections on the origin of a concept. *Science, Technology, & Human Values*, 35(5), 601–617. <https://doi.org/10.1177/0162243910377624>
- Stecher, B. (2010). *Performance assessment in an era of standards-based educational accountability*. Stanford, CA: Stanford Center for Opportunity Policy in Education.
- Stosich, E. L., Snyder, J., & Wilczak, K. (2018). How do states integrate performance assessment in their systems of assessment? *Education Policy Analysis Archives*, 26(13), n13. <https://doi.org/10.14507/epaa.26.2906>
- Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(3), 743. <https://doi.org/10.1037/a0027627>
- Thomas, A. (2016). *Evaluating the validity of technology-enhanced educational assessment items and tasks: An empirical approach to studying item features and scoring rubrics*. (PhD), City University of New York, New York, NY.
- Tung, R. (2010). *Including performance assessments in accountability systems: A review of scale-up efforts*. Boston, MA: Center for Collaborative Education.
- Wehlage, G. G., Newmann, F. M., & Secada, W. G. (1996). Standards for authentic achievement and pedagogy. In F. M. Newmann (Ed.), *Authentic achievement:*

Restructuring schools for intellectual quality. San Francisco, CA: Jossey-Bass.

Wei, R. C., & Cor, K. (2015). *Assessing what matters: Literacy design collaborative writing tasks as measures of student learning*. Paper presented at the American Education Research Association, Chicago, IL.

Zhao, Y. (2018). *Reach for greatness: Personalizable education for all children*. Thousand Oaks, CA: Corwin Press.

Dr. Vanessa Svihla is an associate professor at the University of New Mexico (UNM) with appointments in the Organization, Information & Learning Sciences program and the Department of Chemical & Biological Engineering. She is particularly interested in how people find and frame problems, and how these activities relate to identity, agency, and creativity. In 2018, Dr. Svihla received the NSF CAREER Award to investigate how to support students to make decisions that matter and help them learn as they work on design projects.

Dr. Tim Kubik is a founding partner of Project ARC, LLC. His own arc of professional learning spans 22 years of teaching in primary through postgraduate learning environments. Holding degrees from Yale University and Johns Hopkins University, he has served as team leader of three different interdisciplinary grade level teams, department chair, associate head of Upper School, and as an independent educational consultant. Over the past 10 years, he has facilitated workshops for over 5,000 educators from around the world.

Ms. Tori Stephens-Shauger is the co-founder, executive director, and principal at ACE Leadership High School in Albuquerque, New Mexico. Ms. Stephens-Shauger has been an educator for 20 years with a passion to design and implement programs that keep opportunities available for students. She has spent the last 10 years building a school that is entirely project-based where performance assessment and teacher development are foundational practices.