

Learning Theory, Classroom Assessment, and Equity

Hosun Kang,  University of California, Irvine and Erin M. Furtak, University of Colorado Boulder

Abstract: *Despite increasing awareness about the role of classroom assessments in perpetuating educational inequities, the research community continues to struggle with how to support teachers to design and use classroom assessments for achieving equity. In response to recent calls to better connect learning theory to the design of classroom assessments, we explore the links among contemporary learning theories, classroom assessments, equity, and teachers' professional learning. Building a conceptual argument that we should shift our attention from assessment tasks to a classroom activity system to better support minoritized students' learning via classroom assessment, we examine how teachers participate in assessment codesign activities in two research-practice partnerships (RPPs), and then identify emerging tensions in relation to promoting equity. Each RPP drew upon contemporary learning theories—sociocognitive and sociocultural learning theories, respectively—to create a coherent system of curriculum, instruction, and assessment. The examples show that the tensions emerging from each project are at least partially related to the learning theory that led the researchers to set up professional learning settings in a particular way. Our findings suggest that managing these tensions is an inherent part of the work as researchers seek to support equitable student learning. We discuss specific implications for the assessment community.*

Keywords: classroom assessment, equity, learning theory, professional development, science education

Introduction

Scholars of educational measurement have recently called for increased attention to classroom assessment, or the assessment that involves teachers and students on a daily basis. This emphasis is reflected in Mark Wilson's presidential address to the National Council on Measurement in Education [NCME] (Wilson, 2018), as well as three recent NCME-sponsored conferences focusing on classroom assessment held in 2017, 2018, and 2019. A core element of the argument is that classroom assessment has the greatest potential impact on student learning (as compared to large-scale assessment).

Shepard, Penuel, and Pellegrino (2018), like Wilson (2018), also argued that classroom assessment design and validation must be grounded in an adequate, research-based theory of learning. In particular, Shepard and colleagues, following Penuel and Shepard (2016), state that sociocognitive and sociocultural theories of learning are more likely to support ambitious teaching and further equitable outcomes in classrooms. These assessments focus on student engagement in disciplinary practices as a core element of curriculum, instruction, and assessment, and encourage assessment designers to build from students' experiences and identities as starting points rather than what Wilson (2018) called "top-down" approaches.

In the field of science education, the sociocognitive approach to classroom assessment is often associated with learning progressions (Corcoran, Mosher, & Rogat, 2007). Learning progressions, as a theory of disciplinary learning, are hypothesized learning pathways for students within disciplinary domains. Research shows that these progressions can be used to align large-scale and classroom assessment and to produce thoughtfully designed assessment tasks that provide real-time diagnostic information about the status of student learning to inform ongoing instruction (Shepard, 2019; Wilson, 2018). Progressions may focus not only on content, but also on disciplinary practices, such as modeling in science education (Penuel & Shepard, 2016).

In contrast, sociocultural approaches to classroom assessment rest on the premise that learning is fundamentally a social and cultural activity. Different from cognitive views of learning as a process of acquiring knowledge or changing mental structures in one's head, sociocultural and situative perspectives view learning as "a process in which individuals participate more proficiently in practices that have structure" (Greeno & Gresalfi, 2008, p. 171). Scholars in this research tradition highlight that teachers and students who bring their own life histories coconstruct the opportunities for learners to participate more proficiently in community-valued practice (Greeno & Gresalfi, 2008; Moss, 2008). Attending to students' identities, everyday experiences, and home lives is essential to create classroom assessment systems for equity (Shepard et al., 2018; Tzou & Bell, 2010).

Hosun Kang and Erin M. Furtak contributed equally to this study.

We agree with the emphasis on well-articulated theories of disciplinary learning as the foundations of curriculum and assessment. However, these theories can guide more than just the way assessment tasks are designed; in addition, we can consider, like Shepard and colleagues, the ways in which situated and sociocultural perspectives on learning can inform a broader view of a classroom assessment *system*. By a classroom assessment system, we mean the system of activities for classroom assessments, including actors, materials, and tools that can support productive, learning-focused “assessment cultural practices” (p. 27). As noted by Greeno and Gresalfi (2008), Moss (2008), and others, this perspective of a classroom assessment system broadens the aperture to include teachers and students as the key actors who set up and provide learning opportunities through an assessment task. The lived histories and experiences of those participants are also fundamental to understand in this context.

Given the enormous impact of assessment on students’ lives both current and future, we argue that the goal of promoting equitable engagement in learning environments, particularly for students who have been historically marginalized in educational settings, should be prioritized in considering classroom activity systems, including the design of classroom assessment tasks. Since teachers are core enactors of this activity system, it is essential to engage teachers’ insights as part of the task design process. It is necessary to support teachers in shifting possibly long-held and implicit perspectives on the role and function of assessment in order to promote opportunities to learn for historically marginalized students.

The purpose of this article is to further advance the conversation about designing well-grounded classroom assessment tasks with a commitment to equity. We examine how different learning theories (sociocognitive and sociocultural) guided the construction of equitable classroom activity systems through codesigning classroom assessment tasks in the contexts of two research-practice partnerships (RPPs). The analysis focuses on examining tensions that arose in each RPP as we supported teachers in creating classroom activity systems for equity. The following questions guide our inquiry:

1. What are the similarities and differences in the design of professional learning experiences between the two RPPs? How did contemporary learning theory (sociocognitive and sociocultural) guide the design and facilitation of professional learning to improve classroom assessments with a commitment to equity?
2. What are the tensions within each RPP in terms of promoting equity through assessment?

Theoretical Framework

In this section, we first discuss changing perspectives on disciplinary learning and its implications for designing classroom assessments in light of contemporary learning theory. Next, we unpack the role of teachers with special attention to creating equitable classroom activity systems. Finally, we present codesigning within partnerships as one approach to facilitate teachers’ professional learning.

Changing Perspectives on Disciplinary Learning and Classroom Assessments

Assessment is defined as an inferential process by which one determines learners’ progress toward meaningful disciplinary

learning goals valued by the community (National Research Council [NRC], 2001). Historically, as learning goals within science education have shifted, both what to assess and how to assess have been reconceptualized (NRC, 2007, 2014). This process is by nature not objective. Learning goals are expressed in ways that make them seem settled despite reflecting Whiteness and privileging the values and stories of members of the dominant communities above others (Bang, Warren, Rosebery, & Medin, 2012); as such, assessment of these goals is a fundamentally value-laden enterprise. Those who hold power define learning goals and, as a result, the perspectives and priorities of those in power are privileged through what is assessed, often at the expense of the values, ideas, and perspectives of members of minoritized communities. As a result, assessment has historically functioned to further separate and marginalize students from participation in school (e.g., Au, 2017; Davis & Martin, 2018; Gutiérrez, 2008).

In the field of science education, establishing goals for science learning has similarly been a value-laden and political process (e.g., Calabrese-Barton, 1998; Collins, 1998). What constitutes science as a discipline (“what is science?”), science learning (“what do we mean by understanding science?”), and who is served (“who participates in science, and who is science for?”) are contested (e.g., Bang et al., 2012; Harding, 1991, 2015; Latour & Woolgar, 1979). With the paradigm shift from cognitive to sociocognitive or sociocultural perspectives on learning between the 1980s and 2000s, the science education community broadened its perspectives on science and science learning (Ford & Forman, 2006). A comprehensive report, *Taking Science to School, [TSTS]* (NRC, 2007) noted this important shift as follows:

Expectations of what it means to be competent in doing science and understanding science have also broadened. Beyond skillful performance and recall of factual knowledge, contemporary views of learning prize understanding and application or knowledge in use. . . . Accomplished learners know when to ask a question, how to challenge claims, where to go to learn more, and they are aware of their own ideas and how these change over time. . . . Young learners, not unlike scientists, use knowledge and language to ask questions and make sense of the world. (p. 19)

Building upon this consensus report, the *Framework for K–12 Science Education* (NRC, 2012) and the Next Generation Science Standards [NGSS] (NGSS Lead States, 2013) presented a new vision of science learning grounded in the view of “science-as-practice.” In this vision, learning goals are presented as “performance expectations” that integrate disciplinary core ideas, science and engineering practices, and crosscutting concepts.

The resultant shift in science and science learning calls for a new approach to the design and conceptualization of classroom assessment, including both what and how to assess. Instead of locating knowledge inside the heads of individual learners and privileging symbolic representation, sociocultural theorists posit that learning is manifested through embodied experiences *in* activity situated in contexts. Accordingly, attending to the relationships and interactions between a person and an environment is essential to recognize one’s learning.

Informed by contemporary theories which posit that learning, activity, and context are inseparable (National Academies of Sciences, Engineering, and Medicine [NASEM], 2018), the process of recognizing learning—that

is, the process of assessment—necessitates shifting the unit of analysis from individuals to *activity systems* that are deeply situated in context and culture. In the process of assessment, a teacher selects an assessment task, sets the norms, rules, and expectations, allocates tools and resources, and determines participation structure. The characteristics of activity systems, largely determined by a teacher's pedagogical actions, provide particular *affordances* for learners' participation in a community of practice. When a system of activities affords some aspect of participation for some individuals, it makes it relatively easy for those individuals to participate in that way (Gibson, 1977; Greeno & Gresalfi, 2008). Importantly, the affordances are *relational* (Gee, 2008; Greeno & Gresalfi, 2008); that is, the perceived feasibility of participating in the task varies depending on learners—who they are and their historical relationship with the discipline, and actors in classrooms.

In short, the sociocultural perspective fundamentally shifts our perspective on what “counts” as science learning. Our focus widens to examine the affordances of activity systems in changing learners' participation in practices to infer that one has learned. This is fundamentally different from cognitive views that attribute observed changes to a reflection of some changes in one's mental characteristics. That is, whether and to what extent a learner shows desirable changes in participation over time is deeply related to the affordances of a classroom activity system that is dynamically and relationally coconstructed with learners in context (Kang, 2018).

The Role of Teachers in Designing Assessments for an Equitable Classroom Activity System

Teachers are a core component of classroom activity systems and play a crucial role in reducing the opportunity gap for students from nondominant communities. This involves increasing the affordances of the features of that activity system for students from nondominant communities to more proficiently participate in community-valued practices.

Sociocultural theories reveal multiple ways in which teachers can increase the affordances of activity systems for students from nondominant communities. First, teachers can legitimize learners' nontraditional ways of thinking, talking, and doing. As discussed in the prior section, assessment is fundamentally a value-laden enterprise. Historically, being “good” at science in schools (i.e., meeting learning goals) tends to privilege White, middle-class, Eurocentric values, practices, and experiences over those from nondominant backgrounds (Bang et al., 2012). In a mainstream classroom, nondominant perspectives or ways of making sense of the world are often treated as different, therefore “wrong,” which puts students from nondominant communities in an epistemically untenable position (Bang & Medin, 2010; Tan, Calabrese-Barton, & Benavides, 2019). Teachers can unsettle this settled hierarchical structure by strategically designing classroom activities and assessment systems that disrupt, rather than reinforce, these historical hierarchies, and which integrate students' community, home, and school lives (e.g., Foster & Peele, 2001). For example, nontraditional assessment tasks that legitimize and celebrate diverse ways of thinking, doing and talking can increase opportunities for students from nondominant communities to successfully participate and demonstrate their proficiencies.

In addition, teachers can increase the affordances for marginalized students' participation by designing or selecting classroom assessment tasks that students can relate to. Students from nondominant communities are more likely to engage in disciplinary learning when the task is relevant and important to them (e.g., Tzou et al., 2019; Penuel & Shepard, 2016). By leveraging students' everyday experiences, interests, and concerns to set up a meaningful assessment task, teachers can increase opportunities for students to participate in the work meaningfully. In addition, sociocultural perspectives on learning complicate our perception of what “counts” as competence, encouraging teachers to embrace heterogeneous learning processes (Nasir, Rosebery, Warren, & Lee, 2014; Suárez, 2020).

Teachers can also reconfigure the components of activity systems, such as tools, resources, and participation structures, to increase access and opportunity for marginalized students. For example, a teacher can invite students to draw an individual model that shows interdependent relationships in ocean ecosystems on blank paper. The characteristics of this activity system are changed if students are asked to complete the same work after talking with a partner who supplies cognitive and linguistic resources. The characteristics are further changed if the task is contextualized in students' everyday experiences around the issue that matters to them or their communities. For example, the assessment tasks described above can be framed using a locally contextualized phenomenon such as the following: “draw a model that shows what causes the dangerous yellow-bellied snakes, that have not appeared over the last 30 years, to appear in Huntington Beach.”

Lastly, teachers can increase the affordances of marginalized students' participation by attending to their historical relationships with the disciplines, people, and spaces (Kang, 2018). Gee (2008) notes that an affordance only exists when the learners perceive its presence. In other words, the affordances of an activity system depend on what the learner can perceive as *feasible*. Teachers' discursive interactions and relationships can reframe the way marginalized students think of themselves in a classroom learning community. Marginalized students may bring the historical experiences of failure or being denied as a legitimate member of the science community; teachers can make science “thinkable” (Archer, Dewitt, & Osborne, 2015), and facilitate students to feel it is “feasible” by restoring the marginalized students' relationships with the discipline and the people in the space (Kang, 2018).

Codesigning Classroom Assessment Systems through Research-Practice Partnerships (RPPs): A Hypothesis to Facilitate Professional Learning

In this article, we posit that RPPs can create a context for professional learning for equity by engaging teachers and researchers in collaborative activities, such as codesigning classroom assessments. There is a growing awareness among researchers and practitioners that traditional mechanisms for achieving classroom reforms are inadequate, as they promote a one-way flow of knowledge from researchers to practitioners (e.g., Easton, 2013; Rowan, 2006). The “top-down” or “knowledge transfer model” of reform reinforces traditional knowledge hierarchies and which have historically failed to achieve large-scale changes in educational systems (e.g., Cohen, Moffit, & Goldin, 2007; Tyack & Cuban, 1995). Long-term partnerships with educators and educational organizations

may be better equipped to challenge historical inequities and to promote more agency for teachers and learners (Penuel, Fishman, Haugan Cheng, & Sabelli, 2011; Penuel, 2019). They involve researchers directly partnering with educational and community organizations toward mutually beneficial goals (Coburn & Penuel, 2016). In community-based research, the lines may be further blurred between researcher and practitioner, with codesign also entailing participation in designing and implementing alternatives (Ghiso, Campano, Schwab, Asaah, & Rusoja, 2019).

In this article, we seek to understand how sociocognitive and sociocultural views of learning inform the design of professional learning approaches for teachers within classroom activity systems in the contexts of RPPs, and the tensions emerging from this work. This is important because it will provide theoretical and practical implications for creating a classroom assessment system that promotes equity.

Method

To ground our conceptual analysis of sociocognitive and sociocultural approaches to the design of professional learning environments for classroom assessment codesign, we draw on contrasting cases (e.g., Yin, 2003) of several research-practice partnerships conducted with high school science teachers, which we describe in the following sections.

RPPs grounded in a sociocognitive framework [sociocognitive RPPs]

The first example spans three separate partnership projects with science teachers and district science leaders focused on the design, enactment, and reflection upon formative assessment tasks centered on learning progressions (see Briggs & Furtak, 2019; Furtak, 2012; Furtak & Heredia, 2014 for additional details). Across the studies, multiple learning progressions were used, including those designed by the researchers and teachers themselves, as well as some developed by other research teams and used to support teachers in new contexts. Each project featured high school teachers gathering once or twice monthly in school-based disciplinary teams to explore student thinking and engagement in practice with the support of some form of a learning progression, then codesign classroom tasks to organize and coordinate the sharing of student ideas at planned-for points within a curricular unit. Teachers then enacted the tasks in each of their classrooms before bringing examples of student work—such as written responses, group posters, and video-recorded lessons—to discuss and identify how to further support students in their learning. All three partnerships spanned multiple school years, such that teachers had opportunities to iteratively refine formative assessment tasks and learning progressions in response to their classroom enactments.

RPPs grounded in a sociocultural theory [sociocultural RPPs]

The second example spans two partnership projects with high school science teachers focused on the codesigning and enactment of curriculum and assessments to expand Latinx and multilingual students' opportunity to learn in secondary science classrooms. Grounded in sociocultural and critical perspectives, the project uses five design principles as a tool for mediating professional interactions across settings and

over time. The five design principles are: *making it matter, supporting sense-making, attending to race, language, and identities, building a welcoming community, and disrupting power hierarchies*. The design principles translate various research-based recommendations, such as culturally relevant pedagogy (Ladson-Billings, 1995), for teachers to guide their principled and adaptive pedagogical actions in local contexts. Similar to the sociocognitive RPPs, the teachers in these RPPs engaged in a sequence of activities which involved: (a) coplanning curriculum and assessments in focal units, (b) implementing the codesigned curriculum and assessments in their own classrooms, and (c) analyzing student work throughout the year.

Sources of data

In each of these partnerships, we (the two authors) collected myriad data sources, including recordings of professional development meetings and classroom observations, fieldnotes, classroom artifacts, student responses to assessment tasks, and interviews with teachers.

Analytic approach

We conducted data analysis across about two months. First, we examined common data sources—transcriptions of audio or videorecorded professional development meetings with teachers—as a primary data source. Each author dove deeply into their own data, keeping detailed research memos with transcript excerpts and analytic summaries that identified tensions emerging in the dataset. Next, we exchanged research memos and discussed salient characteristics and patterns in weekly meetings. Specifically, we identified areas of commonality as well as discrepancies in the professional development settings (research question #1) and tensions that manifested within them as teachers designed, enacted and interpreted responses from classroom assessment tasks (research question #2). We used two criteria to select the focal episodes presented in this study. First, we identified episodes that signified salient and important commonalities or tensions of each project. Second, these episodes illustrated commonalities or tensions which related to underlying learning theory. We then further developed the selected episodes using additional sources of data while highlighting salient commonalities (research question #1) and tensions (research question #2).

Findings

We were interested to understand both the similarities and differences in the design of professional learning between the two RPPs, and how contemporary learning theory guided the design and facilitation of professional learning to improve classroom assessments with a commitment to equity. There are several common features between the partnerships despite different underlying theoretical foundations. Both RPPs attended to the multiple components of activity systems in classrooms, including teachers, tools, and assessment tasks, routines, roles, and responsibilities. This is evident from a comparison of the structures and activities of professional learning facilitated by the RPPs. Specifically, both RPPs engaged teachers in the *activities* of codesigning assessment and analyzing student work. Second, both RPPs used *tools*—knowledge-embedded materialized artifacts—to guide the

Table 1. Tensions Emerging from Each RPP

	Sociocognitive RPP	Sociocultural RPP
Designing assessment	More time spent on learning what is in learning progression levels vs. starting design process from who students are and what knowledge, practices, and experiences they bring to the classroom. Presence of learning progression draws teacher attention to deficits rather than assets	Working against the existing culture of professional development; positioning teachers as codesigners of student experiences increases the workload. Some teachers rather want to get the expert-developed curriculum and assessment (“this is too much work”)
Enacting assessment in classrooms	Listening for “wrong” or lower-level answers and correcting them; learning progression focuses teachers on what’s not there vs. using progression to think about how to support students in attaining higher levels of understanding	Having a well-designed curriculum and assessment is insufficient; expanding minoritized students’ opportunity to learn necessitates a deep <i>relationship</i> between the teacher and students
Look at student responses or analyze student responses	Teachers struggle to match complexity of student responses to the examples provided in learning progression	Devaluing the responses that are deviated from standard or normalized ways of thinking, writing, and doing

design of the classroom assessment tasks. The sociocognitive RPPs made use of a learning progression framework and other tools to support design and interpretation of student responses to formative assessment tasks. The sociocultural RPPs used a set of design principles that highlighted important problems of practice in facilitating marginalized students’ learning at schools. The design principles were complemented with concrete representations of practices (e.g., exemplary assessment tasks, a model unit plan). Third, both RPPs *positioned* teachers as experts with rich contextual knowledge of communities and students. As the learners of student thinking and designers of classroom assessment systems, the teachers were guided to design, study, and revise *their own* assessment tasks, instead of passively delivering an expert-designed curriculum and assessment. Lastly, in both RPPs, teachers and researchers worked together for an extended period of time while building *relationships*, instead of offering 1 or 2 days of a professional development program.

At the same time, there are notable differences in the code-signed assessment tasks and teachers’ discourses between two RPPs. The sociocognitive RPPs using the learning progressions produced curriculum-embedded *formative assessment tasks* that students completed individually or in groups and, to varying degrees, were then used to structure teacher-facilitated whole-class conversations where student ideas were discussed. Any summative assessments—often used to determine how much students learned through participating in formative assessment—consisted of researcher-designed items (including multiple-choice and open-ended questions) aligned with the learning progression (in one study, the proposal reviewers required that the summative assessments be kept secret from teachers to avoid any possibility that teach-

ers were “teaching to the test”). Teachers in the sociocultural RPP produced various kinds of formative and summative assessment tasks including both traditional and nontraditional forms of assessments. For example, in a physics unit about momentum, teachers designed three forms of assessment tasks: (a) initial and final performance assessments that engage students in scientific practices, such as modeling, constructing explanation, development argument with evidence, etc.; (b) communicating students’ ideas with people through a two-minute sales pitch about the safety features of the dream car that they designed for their loved one; and (c) drawing and revising a dream car for the loved one and writing a letter to the person, including the description of how the safety features make the loved one safe in a collision. The assessment provided rich information about students’ unfolding ideas as well as their identities, home languages and personal stories about family. For example, the initial assessment revealed that the majority of students thought that “stiff” materials of the car would be safer than “soft” materials, reflecting their everyday perceptions. In addition, the assessment of designing a car for their loved one revealed the story of students’ family stories. For example, one girl wanted to design a car that offered access to a wheelchair for her mom who drove to take her grandma to chemotherapy.

In the sociocognitive RPP, teachers’ conversations were focused on sorting out students’ responses in light of the learning progression framework when designing assessment tasks. While looking at student responses to the assessment, for example, teacher and researcher *discourses* were mostly focused on understanding and refining the learning progression itself, and then using the learning progression as an interpretive framework for students’ responses. In the sociocultural RPPs, teachers’ conversations were focused on students’

experiences, interests, or concerns to design assessment tasks that matter to students—one of the design principles. Nontraditional assessments revealed students' personal stories, which facilitated further conversations between the teacher and students.

What Are the Tensions Emerging from Each RPP in Terms of Promoting Equity through Assessment? In the following sections, we describe tensions emerging from the RPPs in terms of creating a classroom activity system for equity (see Table 1 below). These tensions manifested differently in ways related to their theoretical foundations in sociocognitive as compared to sociocultural frameworks. The main tension in the sociocognitive RPPs was maintaining an asset perspective on student ideas and maintaining an openness to students' nonlinear growth. The main tension in the sociocultural RPPs was unsettling the normalized views, expectations, and practices of professional development as well as disciplinary learning. Together, the analysis shows that creating classroom activity systems for equity inherently involves managing these tensions Table 1

Tensions Emerging from the Sociocognitive RPPs Using the Learning Progression Framework

Designing assessment tasks: focusing on “half-empty” rather than “half-full”. The sociocognitive partnerships began with routines in which teachers explored student thinking with the learning progression. This involved a combination of activities in which teachers would read the learning progression, ask questions, and then explore student work samples—often collected by the researchers in pilot studies—relative to the learning progression. This method by nature prioritized the researchers' and teachers' views on canonical scientific knowledge and practices represented in the learning progression, rather than centering students' experiences. In one study, this was followed by some teachers' framing of their primarily Latinx students based on what the students lacked when coming to school. Another teacher challenged this framing, stating it was as though students were a glass that was “half-empty” rather than “half-full.” University-based researchers acting as facilitators for these teacher meetings similarly worked together to listen for this kind of deficit framing and to encourage teachers to focus on the assets and prior experiences students brought to school. However, the existence of the learning progression as a centerpiece for these conversations privileged learning goals and expectations as reified in the progression.

These and other experiences led researchers to consider a number of design features for the progressions and other tools used alongside the progression. These included sections of the learning progressions that emphasized “what students understand,” as well as, in another study, “look-fors” that were elements of student ideas and explanations that teachers could use to help locate student work at particular levels in the progression.

Enacting assessment in classrooms: staying with a “confront-and-replace” approach while devaluing everyday ideas.

The underlying hypothesis of using learning progressions to support teachers' classroom enactment was that they would support teachers in identifying and responding to student thinking in the course of facilitating classroom discus-

sions (Furtak, 2012). However, in some teachers' classrooms, the learning progressions also seemed to reinforce teachers' confront-and-replace approaches to instruction. While learning progressions can have a number of design features and affordances (Wilson, 2009), two common approaches are to either begin from students' everyday ideas as they develop toward standards-based, scientifically accepted ideas, or to articulate student learning goals as they build across grade bands (e.g., Duschl, Maeng & Sezen, 2011; Shavelson & Kurpius, 2012).

One study, which deliberately began with students' common everyday ideas at the lower-anchors of the progressions, yielded unexpected results in which teachers developed increased competence in listening for particular student ideas—labeled “misconceptions”—in classroom discussions (Furtak, 2012). Some teachers even created posters that were lists of these ideas, hung them on the walls, labeled them as “no-no's” and referred to them in class discussions as ideas they didn't want to hear and were wrong.

A later study took a different approach, employing the learning progressions included in Appendix F of the NRC's *Framework for K–12 Science Education* (2012) as artifacts to guide middle school teachers' discussions of learning goals and expectations as their school district shifted to the NGSS. When reviewing practice progressions that articulated growth in grade bands, district science coordinators intended that teachers would see how they were “under-teaching” learning goals and needed to raise their standards for what their students might learn under the new standards. Many teacher conversations around these progressions, however, focused upon what their students lacked, and how their understandings were only at elementary levels of the progressions, rather than supporting them in participating in learning experiences at the appropriate grade band.

Looking at student responses: struggling to work with “messy” student responses that didn't neatly map on the learning progression framework.

Teachers used the learning progression to interpret student responses to the code-signed tasks and to identify next steps for instruction. Despite the multiple levels represented in the learning progressions, teachers would often begin by placing the student work in just two stacks—one where students were correct, and one where students were wrong, reflecting common “get it or don't” perspectives on student thinking (e.g., Otero & Nathan, 2008). University-based researchers acting as facilitators adopted the guidance, after multiple instances of this two-pile approach, to encourage teachers to look for more nuances or “flavors” of student understandings, encouraging teachers to make at least three piles.

However, while some of these piles would map onto levels of the learning progressions, in other instances teachers would notice that what was present in student work didn't map neatly onto the learning progression levels. The operating idea behind sorting student work in this way was to use the learning progression as a guide for instructional adjustments that might be provided to those students in order to support them in advancing in their learning. What, then, was to be done for the students whose work landed in piles that weren't reflected in the learning progression? Ultimately, teachers sorted the student work into their own sets of piles based on patterns that emerged from the students' responses.

Tensions Emerging from the Sociocultural RPPs Using the Design Principles Framework

Designing assessment tasks: Working against the culture of “doing school” for both the teacher and students. The first tension emerging from the sociocultural RPPs has to do with working against the culture of “doing school” (Bloome, Puro, & Theodorou, 1989; Jiménez-Aleixandre, Bugallo Rodriguez, & Duschl, 2000). Doing school describes a procedural display or going through the motions without deep intellectual engagement. Some teachers expressed their difficulties in transforming curriculum and assessment that goes against what some of their students expect or are accustomed to. For example, one teacher commented:

The kids are kids and they just want to come in and do school. And that’s a reality. And so, when you put that ownership on them, sometimes they’re like, would you just please lecture? Because it’s just easy for them, it’s easy for me. It’s easy for them to just sit there and we tell them exactly what they want and they’re great honor students. So, they’ll just write down and they’ll learn everything that I tell them to learn. Some kids are engaged when they’re challenged to create their own understanding. Other kids are inherently like, I’m too busy or I got this to worry about and you know, I would like you to just tell me what you want me to know.

Ironically, it appears that some teachers’ responses to the professional development facilitated by the sociocultural RPPs were parallel to their students’ responses. The teachers, who were already overwhelmed by so many tasks and initiatives from their schools or districts, expressed frustration about the increased workload as they were positioned as the designers. Different from prevalent professional development programs where teachers sit and get expert-developed materials, tips, and strategies, the sociocultural RPPs guided the teachers to design their own curriculum and assessments tailored for various needs of *their* students in *their* classrooms. Some teachers were excited to be the designers of curriculum and assessments by leveraging their knowledge about students, community, and content. In contrast, some teachers were frustrated about the increased demand, workload, and uncertainty that came with the call for continuously attending and responding to student thinking. For example, one teacher said, “We worked together for three days and I didn’t have a single document to walk away from after this.”

In addition, some teachers, especially those who felt obligated to prioritize meeting standards and covering content more than anything else, didn’t feel it was effective to yield too much control to students. For example, the same teacher also commented:

In terms of disrupting power hierarchies...at the end of the day, I’m still responsible. It’s still my name on their file and I still have to make sure that they are meeting certain standards and so it would be nice if we could have branched off from [traditional teaching], and give [students] a little bit more control over what they wanted to research or pursue or talk about. But, again, how effective is that and how much time do I have?

As much as the teachers expressed tension with students who seemed to “do school,” the researchers felt tensions as they sought ways to encourage teachers to adopt critical perspectives while the teachers struggled to manage increased workload and navigate multiple competing expectations with little support.

Enacting assessment in classrooms: Having a well-designed curriculum and assessment is essential, but not enough.

The second tension has to do with the critical role of the teacher and the relational work that they do in the classroom with students from nondominant communities. In these RPPs, all participating teachers enacted the codesigned curriculum and assessments tailored to the students in their classrooms. Not surprisingly, there were substantial variations across the classrooms in terms of the type and nature of experiences provided for Latinx or multilingual students. It appeared that differences in student experiences, reflected in the quality of student work produced from the assessment tasks, were at least partially related to classroom culture. The classroom culture was coconstructed through cumulative interactions between the teacher and students regarding *both* academic and nonacademic matter. For example, all participating teachers struggled to deal with students’ use of cell-phones during instruction. There were notable differences in the ways in which the teachers responded to the situation, which created a different texture of the space. For example, one teacher responded, “Put the phone away. Guys! Put the phone away” while reinforcing the rules of the school. The students in this class followed directions and completed the assessment task, but the work they produced often didn’t provide strong evidence of their meaningful engagement with sciences despite the teacher’s efforts to enact the codesigned assessment as intended. When the researchers asked about one student while looking at student work, for example, the teacher said, “I don’t know a lot about this student outside the classroom.”

In contrast, the other teacher in the same school responded to the situation by suggesting to the student, “I can charge your phone for you.” The majority of the students came from low-income families, and used old phones that had a short battery life. Students always looked for ways to recharge their phones at the school. This teacher purchased a charger that had multiple adapters, so any model of the phones could be recharged during instruction. In this warm and welcoming environment that was often filled with laughter among the teacher and students, the students produced learning artifacts that showed deep and meaningful engagement with sciences while expressing their stories and creativity.

The partnership activity primarily focused on designing and enacting curriculum and assessment, which helped the teachers to begin their instruction with high quality resources (Kang, Windschitl, & Thompson, 2014; Kang, Windschitl, Stroupe, & Thompson, 2016). As noted by sociocultural theorists (Greeno & Gresalfi, 2008; Gee, 2008), the teacher is one key actor of the classroom who shapes the characteristics of the activity system. Without supporting teachers, who often come from dominant communities, to engage in deep and meaningful relational work with students who were from nondominant communities, a well-designed curriculum and assessment did not seem to do much.

Looking at student responses: “It doesn’t follow the rules”.

The last tension has to do with expanding the settled expectations about what it means to be good at science. This tension arose when teachers looked at student work and interpreted student responses in light of normalized grading practices. When students expressed their ideas with the deviation from normalized practices or expectations, such as following the claim-reasoning-evidence format, those responses were

sometimes treated as incompetent or a failure to show mastery. During a meeting followed by the teachers' enactment of codesigned curriculum and assessment in the evolution unit, for example, four high school teachers were invited to look at samples of student work produced from the assessment. The researchers brought three samples of student work that showed a range of responses. During the professional development, the teachers told the researchers that one response the researcher identified as the most sophisticated and scientifically accurate was scored as "zero" because this student didn't follow the directions. The assessment asked students to predict what might happen to the polar bear in 1,000 years, using the idea of evolution by natural selection. The student predicted that the polar bear would disappear because it does not make a lot of babies and won't fit with the changing environment (i.e., increasing the global temperature, reducing the habitat). For the purpose of grading, teachers used the rule of following the "claim-evidence-reasoning" format in the structure of the assessment, which represents the valued way of thinking and talking in school that privileges Western, Eurocentric sciences. In addition, students were instructed not to predict the extinction.

Critical scholars call for attending to and expanding the settled hierarchy of what it means to be good at science in order to support minoritized students' learning at schools (Bang & Medin, 2010; Bang et al., 2012; Tan et al., 2019). Despite this growing awareness, tensions arise constantly because teachers work under the existing culture of learning and grading at schools.

Discussion and Implications

In this article, we have examined classroom activity systems to explore the links among learning theory, assessment, equity, and professional development. By foregrounding learning theory in our analysis, we surfaced not only similarities and differences in the RPPs, but also tensions that emerged in the process of promoting equity through the transformation of classroom assessments. The examples illustrate how the tensions emerging from each project are at least partially related to the learning theory that led the researchers to set up professional learning settings in a particular way.

We situate our findings in the context of recent assertions that assessment designs be informed by well-articulated learning theories (e.g., Shepard et al., 2018; Wilson, 2018). While we echo the recommendations of Shepard et al. (2018) in emphasizing the importance of explicating the learning theory or assumptions about disciplinary learning that guide the design and use of assessment, we broaden these assertions to focus on how these perspectives can go beyond the ways we might think about the coherence of systems of assessment (e.g., vertical and horizontal alignment). We have taken a different perspective to explore how these theories inform the design of a classroom *activity system*, which enables us to look more deeply at the elements of one level of what Shepard et al. (2018) referred to as systems of assessment. In doing this, we must include the teacher as an integral element of the activity system. It follows that we must consider how to support teachers and their learning as part of these efforts, and guided by well-articulated theories of learning. The learning theories can thus go farther than just the design of tasks—it can (and should) inform our approaches to supporting teachers' professional learning.

We also situate our argument in the urgent call for improving classroom assessments with a commitment to equity. Our analysis supports the argument that we need to pay close attention to the kind of learning theory that guides the design of curriculum, assessment, and instruction, as the learning theories influence classroom assessment systems in ways that have consequences for equity. While certain elements of an assessment system guided by a particular learning theory—such as a learning progression—might afford particular perspectives and activities (e.g., task design across grade bands), they also constrain others (centering design conversations on canonical ideas instead of students' experiences and community assets). Assessment systems guided by situative theory might afford opportunities to disrupt and expand normalized expectations and practices of discipline, teaching, learning, and assessment. However, the assessment system appeared to be limited in drawing attention to sociopolitical aspects of learning and minoritized students' struggles associated with a deeply racialized schooling system itself. In addition, the sociocultural design principles led some teachers to grapple with power hierarchies at their school, and their role in reinforcing and replicating these hierarchies through their design and enactment of classroom assessment. Recall the teacher who reinforces the rule of schooling with limited understanding and relationship with the student's whole self.

This study provides several implications for classroom assessment designers as well as the designers of large-scale assessments. For classroom assessment designers seeking to support equitable student learning through theory-driven RPPs, we offer three practical implications. First, we suggest going beyond the codesign of classroom assessment tasks linked to learning theories. These processes of codesign must be supported by teacher-friendly interpretive tools and frameworks that help teachers take asset-based perspectives and value students' experiences when interpreting students' responses to the codesigned assessment tasks. Teachers' dichotomous or at times deficit discourses toward students' responses that seemingly did not align with canonical scientific ideas was one common tension across the RPPs. Some of these tensions may be mitigated by upfront conversations about how and why it is so important to notice and attend to students' nontraditional ideas while withholding any judgment in order to better support minoritized students' learning. It might be also useful to discuss how we can use the provided interpretive tools or frameworks to have professional conversations that are equitably consequential.

Second, our theory-driven RPPs shed light on the importance of the relational work that teachers do. It follows that we recommend researchers support teachers as they engage in deep and meaningful relational work with students—in particular students from nondominant communities—in order to promote equity through the transformation of classroom assessments. As illustrated by the RPP grounded in sociocultural learning theory, minoritized students' successful engagement in assessment tasks is at least partially mediated by the relationships among the key actors in the classroom activity system (see more details in Kang, 2018). Despite the fact that relationships are a key aspect of the classroom activity system, this has been largely disconnected from or unaddressed in reforming efforts of classroom assessments.

Lastly, we recommend explicitly communicating the nature of professional learning experiences linked to a commitment to equity from the onset of a project, including

recruitment. One common feature of both RPPs was that teachers were positioned as codesigners and agents of change who worked together along with researchers and fellow teachers with the commitment to promoting equity, which is different from prevalent modes of professional development. Explicit communication about commitment, goals, and activities of professional learning might help researchers avoid some of the tensions associated with new roles, positioning, and responsibilities expected by the participating teachers, and therefore help the team to collectively work toward goals of better supporting minoritized students' learning by transforming the classroom assessment system.

In this study, we bounded our analysis to focus on the tensions emerging in the work of designing and using assessments at the classroom level. However, we acknowledge that classroom activity systems are inherently situated in a broader system of activities at the district and state levels; as such, our analysis also suggests possible implications and questions for the designers of large-scale assessments who are interested in further advancing the conversations in the assessment community with a commitment to equity. First, we acknowledge that the existence of large-scale assessments that are not aligned with classroom assessment systems can and do create additional sources of tension, pulling the activities of teachers and students away from community-valued disciplinary learning. Instead, designers of large-scale assessments might consider how contemporary learning theory (e.g., sociocognitive or situative perspectives) can guide the design of large-scale assessment tasks. This may shift a focus from the individual student to more components of the activity system. In addition, we encourage designers of large-scale assessment to design assessment opportunities and tasks that better capture the range of student ideas and experiences (in ways that are consistent with practice-based science learning). Furthermore, how might advanced technologies, including automated scoring and learning analytics, help to capture the range of ideas and experiences beyond the evaluation of correctness? Lastly, we recommend that designers of large-scale assessments articulate tensions in the process of aligning assessment with contemporary learning theory. What frameworks, resources, practices or tools might help the designers to mitigate these tensions? We call for collective and critical conversation among those involved in designing and using assessments *across* levels of educational systems (e.g., classroom, district, state levels) to reimagine systems of assessment that support the powerful disciplinary learning of students who have been and continuing to be marginalized in schools.

Acknowledgments

The authors thank Lorrie Shepard for providing thoughtful and substantive feedback on earlier drafts of this manuscript. This material is based upon work supported by the National Science Foundation under grant numbers DRL-0953375, 1561751, and 1846227, and Research Fellowships from the Knowles Science Teaching Foundation and the Hellman Foundation.

References

Archer, L., Dewitt, J., & Osborne, J. (2015). Is science for us? Black students' and parents' views of science and science careers. *Science Education, 99*(2), 199–237.

- Au, W. (2017). Can we test for liberation? Moving from retributive to restorative and transformative assessment in schools. *Critical Education, 8*(13), 1–13.
- Bang, M., & Medin, D. (2010). Cultural processes in science education: Supporting the navigation of multiple epistemologies. *Science Education, 94*(6), 1008–1026.
- Bang, M., Warren, B., Rosebery, A. S., & Medin, D. (2012). Desettling expectations in science education. *Human Development, 55*(5–6), 302–318.
- Bloome, D., Puro, P., & Theodorou, E. (1989). Procedural display and classroom lessons. *Curriculum Inquiry, 19*(3), 265–291.
- Briggs, D. C., & Furtak, E. M. (2019). Learning progressions and embedded assessment. In S.M. Brookhart & J.H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 146–169). New York: Routledge.
- Calabrese-Barton, A. (1998). Reframing “Science for all” through the politics of poverty. *Educational Policy, 12*(5), 525–541.
- Coburn, C. E., & Penuel, W. R. (2016). Research–practice partnerships in education: Outcomes, dynamics, and open questions. *Educational Researcher, 45*(1), 48–54.
- Cohen, D. K., Moffitt, S. L., & Goldin, S. (2007). Policy and practice: The dilemma. *American Journal of Education, 113*(4), 515–548.
- Collins, A. (1998). National Science Education Standards: A political document. *Journal of Research in Science Teaching, 35*(7), 711–727.
- Corcoran, T. B., Mosher, F. A., & Rogat, A. (2009). Learning progressions in science: An evidence-based approach to reform. CPRE Research Reports. Retrieved from https://repository.upenn.edu/cpre_researchreports/53
- Davis, J., & Martin, D. B. (2018). Racism, assessment, and instructional practices: Implications for mathematics teachers of African American students. *Journal of Urban Mathematics Education, 11*, 45–68.
- Duschl, R., Maeng, S. & Sezen, A. (2011) Learning progressions and teaching sequences: A review and analysis. *Studies in Science Education, 47*(2), 123–182.
- Easton, J. Q. (2013, June). Using measurement as leverage between developmental research and education practice [Presentation]. *The Center for the Advanced Study of Teaching and Learning*, Curry School of Education, University of Virginia, Charlottesville, VA.
- Engeström, Y. (1999). Expansive visibilization of work: An activity-theoretical perspective. *Computer Supported Cooperative Work (CSCW), 8*(1), 63–93.
- Ford, M. J., & Forman, E. A. (2006). Chapter 1: Redefining disciplinary learning in classroom contexts. *Review of Research in Education, 30*(1), 1–32.
- Foster, M. L., & Peele, T. B. (2001). Ring my bell: Contextualizing home and school in an African American community. In E. McIntyre, A. Rosebery, & N. Gonzáles (Eds.), *Classroom diversity: Connecting curriculum to students' lives* (pp. 27–36). Portsmouth, NH: Heinemann.
- Furtak, E. M. (2012). Linking a learning progression for natural selection to teachers' enactment of formative assessment. *Journal of Research in Science Teaching, 49*(9), 1181–1210.
- Furtak, E. M., & Heredia, S. C. (2014). Exploring the influence of learning progressions in two teacher communities. *Journal of Research in Science Teaching, 51*(8), 982–1020.
- Greeno, J. G., & Gresalfi, M. S. (2008). Opportunities to learn in practice and identity. In P. A. Moss, D. Pullin, J. P. Gee, G. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 170–199). New York: Cambridge University Press.
- Gee, J. (2008). Sociocultural perspective on opportunity to learn. In P. A. Moss, D. Pullin, J. P. Gee, G. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 76–108). New York: Cambridge University Press.
- Ghiso, M. P., Campano, G., Schwab, E. R., Asaah, D., & Rusoja, A. (2019). Mentoring in research-practice partnerships: Toward democratizing expertise. *AERA Open, 5*(4), 2332858419879448.
- Gibson, J.J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67–82). Hillsdale, N.J: Lawrence Erlbaum.

- Gutiérrez, R. (2008). A “Gap-Gazing” fetish in mathematics education? Problematizing research on the achievement gap. *Journal for Research in Mathematics Education*, 39(4), 357–364.
- Harding, R. (1991, March). Control of the retained austenite content of ADI. In *1991 World Conference on Austempered Ductile Iron*. Vol. 1 (pp. 22–31). Bloomington IL.
- Harding, S. (2015). *Objectivity and diversity: Another logic of scientific research*. Chicago, IL: University of Chicago Press.
- Jiménez-Aleixandre, M. P., Bugallo Rodríguez, A., & Duschl, R. A. (2000). “Doing the lesson” or “doing science”: Argument in high school genetics. *Science Education*, 84(6), 757–792.
- Kang, H. (2018, April). Re-defining responsive teaching that promotes equity in secondary science classrooms. NARST 2018 Conference, Atlanta, GA, United States.
- Kang, H., Thompson, J., & Windschitl, M. (2014). Creating opportunities for students to show what they know: The roles of scaffolding in assessment tasks. *Science Education*, 98(4), 674–704.
- Kang, H., Windschitl, M., Stroupe, D., & Thompson, J. (2016). Designing learning opportunities for students that advance scientific thinking. *Journal of Research in Science Teaching*, 59(9), 1316–1340.
- Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3), 465–491.
- Latour, B., & Woolgar, S. (1979). *Laboratory life: The construction of scientific facts*. Princeton, N.J: Princeton University Press.
- Love, B. (2019). *We want to do more than survive: Abolitionist teaching and the pursuit of educational freedom*. Boston, MA: Beacon Press.
- Mensah, F. M., & Jackson, I. (2018). Whiteness as property in science teacher education. *Teachers College Record*, 120(1), 1–38.
- Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory into Practice*, 31(2), 132–141.
- Moss, P. A. (2008). Sociocultural implications for assessment I: classroom assessment. In P. A. Moss, D. Pullin, J. P. Gee, G. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 222–258). New York: Cambridge University Press.
- Nasir, N. S., Rosebery, A. S., Warren, B. & Lee, C. D. (2006) Learning as a cultural process: Achieving equity through diversity. *The Cambridge Handbook of the Learning Sciences*. 2nd, United States: Cambridge University Press, 686–706.
- National Academies of Sciences, Engineering, and Medicine [NASEM]. (2018). *How people learn II: Learners, contexts, and cultures*. Washington, D.C.: The National Academies Press.
- National Research Council [NRC]. (2001). *Classroom assessment and the national science education standards*. Washington, D.C.: National Academies Press.
- National Research Council [NRC]. (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, D.C.: National Academies Press.
- National Research Council [NRC]. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, D.C: National Academies Press.
- NGSS Lead States (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.
- Otero, V. K. & Nathan, M. J. (2008) Preservice elementary teachers’ views of their students’ prior knowledge of science. *Journal of Research in Science Teaching*, 45(4), 497–523. <https://onlinelibrary.wiley.com/doi/abs/10.1002/tea.20229>.
- Paris, D., & Alim, H. S. (2017). *Culturally sustaining pedagogies: Teaching and learning for justice in a changing world*. New York: Teachers College Press.
- Penuel, W. R. (2019). Infrastructuring as a practice of design-based research for supporting and studying equitable implementation and sustainability of innovations. *Journal of the Learning Sciences*, 28(659–677), 1–19.
- Penuel, W. R., Fishman, B. J., Haugan Cheng, B., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher*, 40(7), 331–337.
- Penuel, W. R., & Shepard, L. A. (2016). Assessment and teaching. *Handbook of Research on Teaching*, 5, 787–850.
- Rowan, B. (2006). The new institutionalism and the study of educational organizations: Changing ideas for changing times. In H.-D. Meyer & B. Rowan (Eds.), *The new institutionalism in education* (pp. 15–32). Albany, NY: State University of New York Press.
- Shavelson, R. J. & Kurpius, A. (2012) Reflections on learning progressions. Alonzo, A. C. & Gotwals, A. W., *Learning Progressions in Science*. Rotterdam: SensePublishers, 13–26. https://doi.org/10.1007/978-94-6091-824-7_2.
- Shepard, L. A. (2019). Classroom assessment to support teaching and learning. *Annals of the American Academy of Political and Social Science*, 683(1), 183–200.
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, 37(1), 21–34.
- Suárez, E. (2020). “Estoy explorando science”: Emerging bilingual students problematizing electrical phenomena through translanguaging. *Science Education*, 337, 1–42.
- Tan, E., Calabrese Barton, A., & Benavides, A. (2019). Engineering for sustainable communities: Epistemic tools in support of equitable and consequential middle school engineering. *Science Education*, 103(4), 1011–1046.
- Tzou, C., & Bell, P. (2010). Micros and me: Leveraging home and community practices in formal science instruction. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Proceedings of the 9th International Conference of the Learning Sciences (ICLS) 2010*, Vol. 1, (pp. 1127–1134). Chicago, IL: International Society of the Learning Sciences.
- Tzou, C., Meixi, Suárez, E., Bell, P., LaBonte, D., Starks, E., & Bang, M. (2019). Storywork in STEM-Art: making, materiality and robotics within everyday acts of indigenous presence and resurgence. *Cognition and Instruction*, 37(3), 306–326.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward Utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.
- Wilson, M. (2009) Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching*, 46(6), 716–730.
- Wilson, M. (2018). Making measurement important for education: The crucial role of classroom assessment. *Educational Measurement: Issues and Practice*, 37(1), 5–20.
- Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.