

Running Head: SOCIOCULTURAL SYSTEM OF ASSESSMENT

**Building a System of Assessment in a School System to Promote Equity and  
Epistemic Justice:  
A Use-Case of a Research-Practice Partnership in Science Education**

William R. Penuel

*University of Colorado Boulder*

Douglas A. Watkins

*Denver Public Schools*

Key words: assessment, science, equity, epistemic justice, research-practice  
partnerships

This material is based in part upon work supported by the Spencer Foundation, the Gordon and Betty Moore Foundation, and the National Science Foundation (Grant Number DRL-1748757). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funders. A special thanks to partnership colleagues Jean Moon and Sarah Michaels for proposing the framing of our efforts to create inclusive classrooms as engaging in epistemic justice work. Also thanks to Rich Lehrer, Jim Pellegrino, and David Stroupe, who provided comments on an earlier draft of this manuscript.

Corresponding author:

William R. Penuel, School of Education, University of Colorado Boulder, UCB 249,  
Boulder, CO 80302, email: [william.penuel@colorado.edu](mailto:william.penuel@colorado.edu), tel: 415-269-0902

### **Abstract**

Efforts to imagine new possibilities for principled ways to align assessment with sociocognitive foundations of learning in social interaction and with instruction are limited by the absence of credible *use cases* that show how such systems can be developed and maintained within complex educational systems such as large urban school districts. Such a use case can be found within evolving joint work of a research-practice partnership between a university and a large urban school district. The system we are building together focuses on the goals of informing and supporting the redistribution of educational opportunities in science classrooms (equity) and on supporting and repairing students' perceptions of themselves as having agency as a knower and reasoner in their science classrooms and the community (epistemic justice). In this chapter, we describe different components and practices that work together to accomplish these aims: a curriculum with embedded assessments that allows students to figure out rather than be told core ideas; regularly administered exit tickets that elicit students' experience of the classroom; and an instructional guidance system that is focused on iterative refinement of teacher learning opportunities to support student agency. We present evidence of teacher perceptions and uptake of these components and practices, to develop an account of what this particular case can and cannot tell us about the possibilities for designing assessment systems in a district that are thoroughly grounded in contemporary theories of teaching and learning.

**Building a System of Assessment in a School System to Promote Equity and  
Epistemic Justice:**

**A Use-Case of a Research-Practice Partnership in Science Education**

Although a number of scholars have concluded that formative assessment can greatly improve teaching and learning outcomes (e.g., Black and Wiliam 1998; Kingston and Nash 2011), interventions that seek to improve outcomes through assessment are informed by widely differing theories of action. Some of the approaches are guided by specific theories of how to support student learning, while others are guided by broad theories of how organizations can improve by using data from assessments can guide improvements to instruction (Penuel and Shepard 2016a). Interventions with the strongest evidence base are those that offer a social model of how students learn in the disciplines, that is, they emphasize the way that knowledge and practice develop together through participation in meaningful classroom activities (Penuel and Shepard, 2016b; Mislevy this volume; Shepard this volume).

Some interventions that have shown promise draw on a sociocultural model of learning. Such models emphasize the ways that learning always entails becoming a certain kind of person, as one joins in, contributes to, and changes valued social practices (Lave and Wenger 1991; Lave 2012; Rogoff 2003). Sociocultural models of assessment accordingly emphasize how learning and identity are intertwined (Penuel and Shepard 2016b). They offer a particularly powerful account for helping us to see how eliciting and making use of students' interests, concerns, experiences, and identities in instruction can/might promote equity in classrooms, insofar as they can help us illuminate ways to expand our understandings of disciplinary practices and who can take part in them (Bell,

Tzou, Bricker, and Baines 2012; Tan and Calabrese-Barton 2012). At the system level, they may also help to promote justice, by supporting equitable implementation of inclusive instructional practices that explicitly link learning opportunities to learners' identities.

Sociocultural models of assessment are not likely to be implemented fully at the level of a school, district, or state any time soon, however. They demand that new forms of assessment be implemented that would allow teachers to tailor instruction more readily to students' interests and experiences (e.g., González, Andrade, Civil, and Moll 2001). They demand attention to the way that students' lived experiences in classrooms vary and shape learning outcomes, because of how they are positioned in classrooms with respect to race and gender (e.g., Langer-Osuna and Nasir 2016). They would demand that we pay closer attention to how, when, and which students begin to identify as people who can and do appropriate disciplinary practices to their own ends or to address broader concerns of communities and the planet (Nasir and Hand 2008; Penuel 2016). In today's accountability regimes that emphasize test score growth over meaningful learning outcomes and that are informed by an institutional logic that views educational opportunity as scarce, trying to develop an assessment system that monitors and addresses inequities in student experience and encourages teachers and students to connect disciplinary goals to broader goals for social justice is like trying to push a river upstream.

Yet within educational systems there can be "niches" where new possibilities can emerge, if partly buffered from the currents around them (Cohen and Mehta 2017). In this chapter, we describe such a niche within a long-term research-practice partnership that is

developing and testing a system of assessments grounded in sociocultural theories of learning. It is a small niche within a larger school district that (like all districts) includes competing aims, priorities, and initiatives, and the system is still very much a work in progress. In presenting this system here as a pair of “use cases,” we hope to offer an image of the possible that is informed by practical tools and some real life success stories, serves both the purpose of addressing the field more broadly about the possibilities and challenges of developing an assessment system grounded in sociocultural theory and providing a tool for fostering internal dialogue and reflection with stakeholders in schools and our community about the system in development.

This chapter is organized as follows. First, we describe the purposes of use cases in the context of collaborative design and identify who is involved in developing this system. Second, we elaborate on the goals of this system and how they reflect sociocultural theories of learning. Third, we present two different use cases intended to show how some of the different components and practices work together to accomplish our aims.

### **Use Cases and Partnerships**

*Use cases* are a popular method in software development for developing requirements for multi-component technical systems. In that field, use cases help a development team and stakeholders to imagine and communicate clearly how people will use the system to perform particular tasks (Jacobson, Christerson, Jonsson, and Övergaard 1992). Use cases take the form of a schematic narrative that describes how actors interact with each other and components of a system, as a way to help teams and stakeholders arrive at a shared understanding of what a system should accomplish

(Cockburn 2000). In agile and participatory forms of design, use cases can become something that stakeholders shape and argue about, including the purposes that are worth pursuing to functional requirements and constraints the system must observe (Jacucci, Calzà, and Dandrea 2002).

Developing a use case typically begins with an identification of actors who will interact with the system and their goals. It begins here, because a use case is intended to help a development team stay focused on giving users what they want and need to perform tasks (Bittner and Spence 2003). The challenge in developing a use case for any assessment system within an educational agency such as a school district is the multiplicity of actors and purposes that such a system must accommodate. Many purposes will certainly be at odds with a sociocultural perspective on learning, and many uses of assessment data could easily reproduce, rather than transforming inequities. Addressing this challenge begins with naming *who* is at the table when deciding on the system's purposes and designing processes to support the *mutual learning* of actors that yields expanded sense of possibilities for assessment.

Our approach to developing the theory-driven assessment system described in this paper has been to develop and test this system within a single central office department in a large urban school district as part of a research-practice partnership. Research-practice partnerships are long-term collaborations between research and practice focused on investigating and searching for solutions to persistent problems of practice (Coburn and Penuel 2016). The core members of our partnership are district curriculum leaders and teachers in science in the Denver Public Schools and researchers from the University of Colorado Boulder. Representatives from five other organizations are involved in

supporting the work described here—Northwestern University, the Tidemark Institute, Clark University, BSCS, and Project VOYCE—that bring community perspectives, academic expertise, and professional development capacity to bear on the joint work.

As a partnership, when we develop and test innovations, we employ a participatory form of design research called Design-Based Implementation Research (DBIR; Fishman, Penuel, Allen, Cheng and Sabelli 2013), in which changing systems—rather than designing and testing the efficacy of a single innovation—is a key object of design. For us, participation is not just a means to improving the usability of systems—though that is an important goal—it is also a means for participants to have a say in the goals and strategies used that they are expected to implement or use to support their learning (cf., Kensing and Greenbaum 2013). Design teams are made up of actors and stakeholders across different levels of the system and members of the community, and we structure research and development activities to allow for participants to influence both the process and products of design (Severance, Penuel, Sumner and Leary 2016). The system in development described here is no different, though the forms of participation vary from use case to use case in ways that are consequential for the level of observed buy-in to different components, as we elaborate further below.

### **Equity and Epistemic Justice as Shared Purposes in the Partnership**

In our partnership, we define *equity* in a way that is informed by sociocultural perspectives on teaching and learning.<sup>1</sup> With respect to equitable teaching, we focus on supporting and monitoring implementation of teaching practices that build from students' cultural funds of knowledge, practices, and identity resources (Esteban-Guitart and Moll 2014; Ladson-Billings 1995; Moll, Amanti, Neff, and González 1992; Paris 2012). In

science, the kinds of practices we seek to support include selecting phenomena to study that students say are interesting to them and relevant to their communities (Penuel 2017), using students' own questions to drive learning (Schwarz, Passmore and Reiser 2017), eliciting and making use of students' own knowledge and experience to explain phenomena (González-Howard and McNeill 2016; Reiser, Novak and McGill 2017), and encouraging students to use everyday language to help them get a grasp on core ideas and practices of science they are learning (Brown and Ryoo 2008). Further, in our work we define an equitable educational *system* as one in which all students encounter opportunities where they can connect what they are learning to their lives outside of school and that help them imagine and pursue futures where they can apply knowledge and practices at work, in civic and family life, and at play (Penuel, Clark, and Bevan 2016). As a result of experiencing such opportunities, we would hope that students' appropriation of science practices shifts, as does their image of themselves as scientists, and that these shifts would be important to assess.<sup>2</sup>

We also aim for an educational system that promotes a particular form of justice that we call *epistemic justice*. Epistemic oppression or injustice occurs when listeners discount a speaker's knowledge claims in a cumulative fashion (that is, repeatedly, across encounters and contexts) due to some prejudice against them based on their social identity (Fricker 2009). Discounting others' claims in this way can be a systemic form of oppression, when it targets entire groups of people and when it is reinforced by particular institutions' and groups' powers to enforce a particular interpretation of experience on others and when it is linked to the distribution of other social goods (Anderson 2012; Dotson 2014). School science historically is a social setting where students' contributions

to building knowledge are often discounted because their ideas do not reflect canonical views (i.e., they hold misconceptions, from one point of view) and because the authority of texts and the teacher as source of knowledge is privileged (Aguiar, Mortimer, and Scott 2010; Lemke 1990). Certain ideas in science education—such as the idea that a central purpose of teaching is to confront and challenge students’ misconceptions—reinforce or help to perpetuate this epistemic oppression.<sup>3</sup> (Stroupe, personal communication). In addition, studies of interaction in science classrooms underscore the ways that gender and race shape whose contributions are taken up as valuable for knowledge construction, reproducing rather than challenging inequities (e.g., Carlone 2004; Carlone, Haun-Frank, and Webb 2011).

The pursuit of epistemic justice requires new practices in classrooms and new commitments and practices at the system level. Re-constructing science classrooms as sites of epistemic justice will require forming classroom communities in which members of the community explicitly negotiate standards of accountability to disciplinary norms and practices (Engle 2012; M. J. Ford 2008; Lucas, Broderick, Lehrer, and Bohanan 2005), make room for student goals and contributions that do not follow a given curricular storyline (Miller, Manz, Russ, Stroupe, and Berland 2018), and allow students to apply their knowledge and grasp of science and engineering practices to social and environmental problems (Penuel 2016; Tan and Calabrese Barton 2012). To make such classroom communities available to all learners will require supports for teachers to develop their “interpretive power” (Rosebery, Warren, and Tucker-Raymond 2016) to see how student contributions to discussions make sense and can be leveraged for collective science learning in the classroom. Further, they must re-position students as having

authority in the classroom for knowledge production (Engle 2012) and allow students to grapple with the kinds of uncertainties that scientists face when conducting investigations (Manz and Suarez 2015).

Educational systems, for their part, will need to find ways to grapple with how conceptions of knowledge and who can be a knower are reinforced in everyday routines and practices in schools (Dotson 2014; Miller et al. 2018). Similarly, they will need to provide all teachers with routines, tools, and insights that can help them elicit and make use of student interests, experiences, and identities that are grounded in subject matter goals and the racial, linguistic, and gender diversity of the system. These tools include not only activities traditionally thought of as assessment, but also curriculum materials that are linked to assessment and support goals of epistemic justice. Finally, they will need to consider how they can provide opportunities for students to have more of a say over what goes on in schools, in a way that moves beyond typical models of student governance and toward genuine shared leadership with adults (Kirshner and Jefferson 2015).

How can the efforts of a research-practice partnership to develop an assessment system informed by sociocultural theory contribute to these aims and help monitor progress toward the changes in practices needed to accomplish those aims? First, we can use the efforts to describe a vision for equitable science teaching and epistemic justice that is grounded in a theory of learning and that is linked to tools and routines for helping teachers and students partner to realizing the vision in their classrooms. Second, we can monitor and seek to address inequities in student experience in science classrooms, using assessment tools that focus on students' perceptions of the coherence of, personal and

community relevance of, and their own contributions to knowledge-building activities in the classroom. Third, we can design and monitor systems of professional learning for teachers to build equitable classroom communities and to create culminating assessments in which students contribute directly to addressing community goals and concerns. The use cases presented below say—at a very broad level—how we are undertaking these activities in our partnership to promote equity and epistemic justice.

### **Use Cases and Arguments for an Emerging Assessment System**

In this section, we present two written use cases in the brief typically used in agile software development as boundary objects for discussion in the partnership, along with a narrative argument for their importance to building an assessment system informed by sociocultural theory. That argument draws both on evidence we have gathered in Denver, as well as in other settings where specific assessment tools and routines have been used.

#### **Use Case 1: Building and Using a Driving Question Board**

Our assessment system is intended to support a particular approach to implementing the vision of teaching and learning presented in *A Framework for K-12 Science Education* (National Research Council 2012) and reflected in the *Next Generation Science Standards* (NGSS; NGSS Lead States 2013). Our approach privileges the student point of view on subject matter learning. That is to say, units we have designed with teachers and our partners are organized so that the sequence of activities make sense from the student point of view and build toward disciplinary understandings (Reiser, et al. 2017). Our strategy for creating curriculum materials that are coherent from the student point of view is to organize units around investigations of observable phenomena in the world, and for each unit to culminate in challenges where students apply their

understandings and use engineering practices to solve a design problem. These engineering practices are derived in part from studies of how engineers really work (Bucciarelli 1994; Trevelyan 2010) and are ones called out in the *Framework for K-12 Science Education* (National Research Council 2012) and include defining problems, as well as designing and testing solutions.

The challenge that our curriculum co-design teams have set for ourselves is to select anchoring phenomena that are both compelling to students and that will help students figure out disciplinary core ideas emphasized in the Next Generation Science Standards (NGSS). To write curriculum materials to support students' epistemic agency in the classroom further requires us to anticipate students' questions well enough to specify a possible pathway to explaining the anchoring phenomena and solving the design problem that includes student activities and assessments and that is therefore coherent from the student point of view (Reiser et al. 2017). Of course, since not all student questions can be anticipated, we need practices teachers can use to help them partner with students to come up with questions the class can pursue together and gives them confidence in their ability to support students in pursuing their own questions. In addition, teachers will need to support students who have encountered more teacher-directed science instruction in the past to learn how to contribute to directing the class' trajectory of investigation.<sup>4</sup>

The "Anchoring Phenomenon Routine" is such a practice, and it is a core part of our assessment system. The Anchoring Phenomenon Routine is triggered by the start of a new unit or a new "plot line" within a unit. Our evolution unit begins with a presentation of a short video of a girl, Addie, who has been hospitalized and who has multiple bacterial infections that are resistant to antibiotic treatment. As part of this gripping

opening, students generate and prioritize as a class a list of questions that they need to answer to explain what is going on with Addie. In the initial lesson, students write questions individually and in small groups, and they identify experiences they have had that might help them understand what is going on. As a class, they build a timeline of the events that they see in the video. Then, students in small groups draw an initial model to explain what they think is going on. This leads students to generate questions about parts they can't explain.

As a final part of the routine, the class together assembles these questions and organizes them into major categories, recording them on an artifact called the Driving Question Board (Weizman, Shwartz, and Fortus 2008). The Driving Question Board is intended to be a living artifact in the classroom. Other routines support the teacher in returning at the beginning and end of lessons to the Driving Question Board, to help the class monitor progress toward answering the class questions, as well as questions that individuals or small groups might pursue. It serves both as an initial assessment for the unit, because student questions provide a window into what students already know and can do (Dori and Herscovitz 1999; Minstrell and van Zee 2003), and the assessment attends to the intertwining of learning and identity, because it provides the class with a public record of relevant interests and experiences that the class can use in their pursuit of questions. Because all students have the experience of being sick and living in families that deal with the feelings and consequences of someone being sick—either over the short-term or over the long haul—the phenomenon connects in some way to all students' experiences in ways that are surfaced from each student and made relevant to the ongoing

investigations of the class, even if the particular experiences and their relevance to explaining the phenomenon at hand must be elicited in classroom interactions.

---

Insert Figure 1 about here

---

Of course, there are several ways this practice can fail to support goals of teachers and students in the classroom. From the teacher and curriculum writer's perspective, if most of the student questions are unrelated to the storyline already, the materials developed for students will be useless to them, if they decide to pursue those questions. Some questions—if pursued—could take the class on a long detour from the planned sequence, limiting opportunities for figuring out other core ideas later in the school year. From the student point of view—and from ours as a partnership—the most consequential failure condition would occur when students construct a Driving Question Board, but it is never used or if teachers simply tell students the answers to their questions. In such an instance, epistemic injustices are reproduced rather than faced, because the irrelevance of students' knowledge and experience is underscored by what comes after the Anchoring Phenomenon Routine.

We have devised a number of strategies for preventing and handling these kinds of failures. For one, we select anchoring phenomena using an extended process that includes not only an analysis of the standards (Kracjik, Codere, Dahsah, Bayer, and Mun 2014) but also administration and analysis of results from a survey that allows us to get a handle on whether and for whom candidate anchoring phenomena are interesting and personally relevant. In addition, before developing storylines for units, we enact the anchoring phenomenon routine multiple times with design teams acting as students would, and we

iterate on our storylines on the basis of teacher feedback from implementation. We present the idea of a “detour” as an opportunity to differentiate instruction and encourage teachers to build in time for individual or small group inquiry related to questions. Similarly, we encourage teachers to develop routines where responsibility for returning to the Driving Questions Board periodically is shared among students and the teacher.

Evidence from Denver and elsewhere indicates that the use of the Driving Question Board can have a powerful impact on students’ experience of the classroom. When, for example, students in pilot classrooms in Denver felt the day’s lesson connected to the anchor for the unit, they were more likely to report feeling the emotion of excitement (Penuel, Van Horne, Severance, Quigley, and Sumner 2016). A separate study of five teachers outside the partnership showed that students’ perception of whether the day’s lesson mattered to them was closely related to their perception that they had made progress on a question from the Driving Question Board (Zivic et al. 2018). This same study found that the majority (62%) of student questions on the board had been anticipated in the unit design. Still, this leaves many student questions (36%) that are not answered by the storyline and that the teacher must figure out how to address. In addition, both studies showed evidence of wide variation—that is inequity—in student experiences across teachers.

As a strategy for addressing these failure conditions, we have developed a strong focus on implementation of the Driving Question Board in teacher professional development. In particular, we model the use of the Anchoring Phenomenon Routine for teachers when they participate as students (or with “student hat” on) and as teachers (with “teacher hat”) to give them an experience for how to include all students’ ideas and guide

their organization in a way so that students can see they are partners in deciding on the direction of the unit. We also are providing opportunities for teachers to invent and share strategies for differentiation that focus on student questions that are likely to fall outside the storyline. These take place as part of regularly scheduled district professional development days.

### **Use Case 2: Monitoring and Supporting Equity and Epistemic Justice**

The evidence used to develop analyses of students' experiences comes from a brief survey measure that our partnership has developed to help us monitor goals of equity and epistemic justice at the system level. It is an example of what improvement scientists call a "practical measure," meaning that it is intended to provide evidence *about* practice that is easily gathered and interpreted *in* practice (Bryk, Gomez, Grunow, and LeMahieu 2015). The constructs measured are students' perceptions of the coherence of the day's lesson with respect to the questions on the Driving Question Board, students' perceptions of the relevance of the day's lesson to them, their class, and the community, and students' sense that they made contributions that helped the class make progress on the Driving Question Board that were recognized and appreciated by others. Consistent with guidelines for developing and validating practical measures, each construct has only a few items so that the survey—or Student Electronic Exit Ticket (SEET) as we call it—can be completed in just a few minutes of class time. Also, a short survey makes analyses easier to perform to inform the ongoing work of the partnership to support teacher development. As a partnership, evidence from this measure—coupled with observations of classrooms—has informed a recent shift from a focus on curriculum implementation in

professional district and toward promoting inclusive classroom cultures where students and teachers partner to figure out the core science ideas.

In this use case, the triggering event is the conclusion of a lesson in a unit that calls for using a SEET. Such lessons are built into units and occur with regular frequency: about four times over the course of an 8-week unit. Students complete the survey on an Internet-connected device—a mobile phone, a tablet, laptop, or desktop computer. The first time they complete the survey, students respond to questions about their gender, race, and home language, in addition to answering questions about their experience of the day's lesson. Researchers aggregate data from multiple classrooms after a specified period (usually once per semester) and perform analyses of the data. We examine variation in equity of experience and epistemic justice both within classrooms and across classrooms, looking for patterns that show evidence of epistemic injustice (e.g., fewer African American students are contributing to large group discussions or feeling that their voices are consequential in such discussions) as well as to inequity of opportunity (e.g., some teachers are not using the Driving Question Board at all, while others are using it to partner with students in setting the direction for the units). Then, in a meeting that includes district leaders and partners who help us design and provide professional learning opportunities for teachers, we discuss results and their implications for supporting teachers in ways that can better meet our partnership's goals for equity and epistemic justice. These could include more intensive coaching in particular schools, supporting teachers' evaluators with knowing how to evaluate this type of instruction, or focused workshops on creating inclusive classroom cultures or using a Driving Question Board effectively.

This particular practice has some failure conditions we have documented in our own research and that we are still searching out strategies for effective repair. For some teachers, the challenge of setting up devices to collect data is not worth the extra effort required to provide data that is of use mainly to the partnership leaders (Penuel, Van Horne, Jacobs, and Turner 2018). Though teachers had been involved in helping create questions and prioritize the questions that appear on it, over time, its value to them appears to have decreased, based on the numbers of teachers who provide us with data from SEETs. We recently integrated some multiple-choice questions related to the performance expectations of the day's lesson into the SEETs, at the request of teachers who said these would make them more valuable as assessments for them in their classrooms. In addition, we are developing a new interface and strategies for data collection that will allow teachers to customize the SEET and to explore how students' SEET ratings change over time, as a source of potential evidence of shifting epistemic agency. A bigger challenge awaits us, if we are able to address these particular concerns, and that is how we ensure that the SEET continues to be used primarily for improvement purposes. It is common for assessments of teaching in school districts and states to serve both formative and summative purposes, and this can compromise their formative purposes (T. G. Ford, Van Sickle, Clark, Fazio-Brunson, and Schween 2017; Hinchey 2010). In addition, there is already an end-of-year survey students complete that is included in teachers' evaluation, so the SEET might be perceived as duplicative to the district, even though it provides more detailed, lesson-level data. As a partnership, we will need to include more teacher voices in a conversation about how to create a balanced assessment system (Shepard this volume; Shepard, Penuel, and Pellegrino 2018) that

meets the district's goals for the partnership and that is sustainable beyond our current lines of research in the district.

### **Discussion and Conclusions**

In this chapter, we have presented two use cases from a research-practice partnership that is seeking to develop a balanced system of assessment within science. The system is part of a protected “niche” within the school district and is intended to support a broader effort to promote equity and epistemic justice as teachers implement new standards. It includes tools and routines to inform classroom teaching, monitoring assessments focused on student experience of the classroom, and a teacher support team comprised of district leaders, teacher leaders, and external partners who can design and implement professional development to address problems identified from systems monitoring tools. Unlike systems that seek to focus primarily on cognitive dimensions of learning, what makes this system unique in our view is its recognition of the importance of assessing how well teaching connects to students' interests, experiences, and identities. In that respect, it more thoroughly reflects a sociocultural view of learning than other systems of which we are aware being implemented in schools.

In the chapter, we have described how the system supports two different goals of equity in the classroom and epistemic justice at the system level as if they were separate. But in fact, the two are related, not only because the two are informed by a common, sociocultural model of learning but also because they are linked through the broader goal of supporting equitable implementation of new science standards in a large school district. The two goals are accomplished by different strategies, though, that attend to change at different levels of the system. What is needed to support assessment use by

individual teachers is different from what is needed to support such use at the level of a system. Our various DBIR projects within the partnership are seeking to elaborate precisely how, drawing on theories of teacher learning as well as organizational change.

Put another way, our system of assessment attends to and links different “planes” of learning and development as Rogoff (2003) might describe them. At the individual level, students’ individual interests, experiences, and identities are an important resource for science learning, and assessment seeks to capture these formatively, to guide teacher decision making. Students also negotiate with peers and the teacher to set the direction for learning, by posing and prioritizing questions and jointly monitoring progress toward answering those questions over time. Teachers are expected to facilitate individual accountability to emergent disciplinary norms of reasoning and warranting of claims, as jointly established within classrooms and with reference to broader norms of scientific communities.

At the community plane, attention is paid to selecting phenomena and design challenges that might help students see how science and engineering can address concerns and realize goals for justice in their neighborhood or city. The classroom is a community where students pursue questions together, evolve standards for what counts as a question worth pursuing, and for how answers to questions are to be warranted. Still another community—of leaders in a partnership—attends to changing patterns of student participation in and perceptions of classroom communities

By no means is our system complete or fully operational, and we recognize the ongoing need to increase teacher, student, parent, and community voice in our assessment system. Further, we need more than just practical measures and interest surveys to assess

the long-term changes to practice-linked identities among students. However, these use cases are intended to help us internally to communicate with stakeholders and invite them to critique and revise elements of the system that reflect their goals and concerns. For the external audiences reading this chapter, the use cases are intended to provide a more concrete (if high-level) image of what is possible, when it comes to creating a sociocultural system of assessments in a large school district.

A sociocultural approach, we hope we have shown, foregrounds the broader system in which classroom teaching takes place. It recognizes, too, that institutions like school systems have “epistemic powers” (Anderson 2006, 8) that is, the ability through various policies and assessments to signal what is worth knowing and who counts as a knower. Ultimately, the success of schools as institutions depends on them becoming, as Anderson (2006) writes, a “model of democracy in which its epistemic success is a product of its ability to take advantage of the epistemic diversity of individuals” (11). A sociocultural system of assessments, collaboratively designed with multiple stakeholders in a system, is well-positioned to work toward this kind of success.

### **End Notes**

<sup>1</sup>From the standpoint of assessment, research on equity has focused primarily on developing tests that are fair and unbiased, and while our partnership is certainly concerned with testing fairness, our vision provides first and foremost an image of an equitable classroom community that recognizes and makes use of the contributions of all its members.

<sup>2</sup>Thanks to David Stroupe for suggesting this particular entailment of our claims about what an equitable educational system should look like.

<sup>3</sup>David Stroupe, personal communication (9/22/18).

<sup>4</sup>David Stroupe, personal communication (9/22/18).

### References

- Aguiar, Orlando G., Eduardo Mortimer, and Phil Scott. 2010. Learning from and responding to students' questions: The authoritative and dialogic tension. *Journal of Research in Science Teaching*, 47(2): 174-193.
- Anderson, Elizabeth. 2006. The epistemology of democracy. *Episteme*, 3(1-2), 8-22.
- Anderson, Elisabeth. 2012. Epistemic justice as a virtue of social institutions. *Social Epistemology*, 26(2): 163-173.
- Bell, Phil, Carrie Tzou, Leah A. Bricker and Annemarie D. Baines. 2012. Learning in diversities of structures of social practice: Accounting for how, why, and where people learn science. *Human Development*, 55: 269-284.
- Bittner, Kurt, and Ian Spence. 2003. *Use case modeling*. Boston, MA: Addison-Wesley.
- Black, Paul, & Dylan Wiliam. 1998. Assessment and classroom learning. *Assessment in Education*, 5(1): 7-74.
- Brown, Bryan A., and Kelly Ryoo. 2008. Teaching science as a language: A “content-first” approach to science teaching. *Journal of Research in Science Teaching*, 45(5): 529-553.
- Bryk, Anthony S., Louis M. Gomez, Alicia Grunow, and Paul G. LeMahieu. 2015. *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard University Press.
- Bucciarelli, Louis L. 1994. *Designing engineers*. Cambridge, MA: MIT Press.
- Carlone, Heidi B. 2004. The cultural production of science in reform-based physics: Girls' access, participation, and resistance. *Journal of Research in Science Teaching*, 41(4): 392-414.
- Carlone, Heidi B., Julie Haun-Frank, and Angela Webb. 2011. Assessing equity beyond knowledge- and skills-based outcomes: A comparative ethnography of two fourth-grade reform-based science classrooms. *Journal of Research in Science Teaching*, 48(5): 459-485.
- Coburn, Coburn E., and William R. Penuel, William R. 2016. Research-practice partnerships in education: Outcomes, dynamics, and open questions *Educational Researcher*, 45(1): 48-54.
- Cockburn, Alistair. 2000. *Writing effective use cases*. Boston, MA: Addison-Wesley Professional.
- Cohen, David K., and Jal Mehta. 2017. Why reform sometimes succeeds: Understanding the conditions that produce reforms that last. *American Educational Research Journal*.

- Dori, Yehudit J. and Orit Herscovitz. 1999. Question-posing capability as an alternative evaluation method: Analysis of an environmental case study. *Journal of Research in Science Teaching*, 36: 411-430.
- Dotson, Kristie. 2014. Conceptualizing epistemic oppression. *Social Epistemology*, 28(2): 115-138.
- Engle, Randi A. 2012. The productive disciplinary engagement framework: Origins, key concepts, and developments. In *Design research on learning and thinking in educational settings: Enhancing intellectual growth and functioning*, David Y. Dai (Ed.), 161-200. New York, NY: Routledge.
- Fishman, Barry J., William R. Penuel, Anna-Ruth Allen, Britte H. Cheng, and Nora Sabelli. 2013. Design-Based Implementation Research: An emerging model for transforming the relationship of research and practice. In *Design-based implementation research. National Society for the Study of Education Yearbook*, Barry J. Fishman, William R. Penuel, Anna-Ruth Allen, & Britte H. Cheng (Eds), 136-156.
- Ford, Michael J. 2008. Disciplinary authority and accountability in scientific practice and learning. *Science Education*, 92(3): 404-423.
- Ford, Timothy G., Mary Elizabeth Van Sickle, Lynn V., Clark, Michelle Fazio-Brunson, and Dorothy C. Schween. 2017. Teacher self-efficacy, professional commitment, and high-stakes teacher evaluation policy in Louisiana. *Educational Policy*, 31(2): 202-248.
- Fricker, Miranda. 2009. *Epistemic injustice: Power and the ethics of knowing*. New York, NY: Oxford University Press.
- González, Norma, Roy Andrade, Marta Civil, and Luis Moll. 2001. Bridging funds of distributed knowledge: Creating zones of practices in mathematics. *Journal of Education for Students Placed at Risk*, 6(1-2): 115-132.
- González-Howard, Maria and Katherine L. McNeill. 2016. Learning in a community of practice: Factors impacting English-Learning students' engagement in scientific argumentation. *Journal of Research in Science Teaching*, 53(4): 527-553.
- Hinchey, Patricia H. 2010. *Getting teacher assessment right: What policymakers can learn from research*. Boulder, CO: National Education Policy Center.
- Jacobson, Ivar, Magnus Christerson, Patrik Jonsson, and Gunnar Övergaard. 1992. *Object-Oriented Software Engineering: A use case driven approach*. Boston, MA: Addison-Wesley.
- Jacucci, Gianni, Diego Calzà, and Vincenzo Dandrea. 2002. *Use of use cases in design for end user design in use*. Trento, Italy: Department of Sociology and Social Research, University of Trent.
- Kensing, Finn, and Joan Greenbaum. 2013. Heritage: Having a say. In *Routledge International Handbook of Participatory Design*, Jesper Simonsen & Toni Robertson (Eds.), 21-36. London, UK: Routledge.
- Kingston, Neal, and Barbara Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4): 28-37.
- Kirshner, Ben and Antwan Jefferson. 2015. Participatory democracy and struggling schools: Making space for youth in school turnarounds. *Teachers College Record*, 117 (6).

- Krajcik, Joseph S., Susan Codere, Chanyah Dahsah, Renee Bayer, and Kongju Mun. 2014. Planning instruction to meet the intent of the Next Generation Science Standards. *Journal of Science Teacher Education*, 25(2): 157-175.
- Ladson-Billings, Gloria. 1995. Toward a theory of culturally-relevant pedagogy. *American Educational Research Journal*, 32(3): 465-491.
- Langer-Osuna, Jennifer M., and Na'ilah. 2016. Rehumanizing the "other": Race, culture, and identity in education research. *Review of Research in Education*, 40: 723-743.
- Lemke, Jay L. 1990. *Talking science: Language, learning, and values*. Norwood, New Jersey: Ablex.
- Lucas, Deborah, Nichole Broderick, Richard Lehrer, and Robert Bohanan. 2005. Making the grounds of scientific inquiry visible in the classroom. *Science Scope*, 29(3), 39-42.
- Miller, Emily C., Eve Manz, Rosemary S. Russ, David Stroupe, and Leema Berland. 2018. Addressing the epistemic elephant in the room: Epistemic agency and the Next Generation Science Standards. *Journal of Research in Science Teaching*.
- Minstrell, James, and Emily van Zee. 2003. Using questioning to assess and foster student thinking. In *Everyday assessment in the science classroom*, J. Myron Atkin and Janet Coffee (Eds.), 61-74. Arlington, VA: NSTA Press.
- Moll, Luis C., Cathy Amanti, Deborah Neff, and Norma González. 1992. Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory into Practice*, 31: 132-141.
- Nasir, Na'ilah S., and Victoria Hand. 2008. From the court to the classroom: Opportunities for engagement, learning, and identity in basketball and classroom mathematics. *Journal of the Learning Sciences*, 17(2): 143-179.
- National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: The National Academies Press.
- NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: National Academies Press.
- Paris, Django. 2012. Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher*, 41(3): 93-97.
- Penuel, William R. 2016. Studying science and engineering learning in practice. *Cultural Studies of Science Education*, 11(1): 89-104.
- Penuel, William R. 2017. *Selecting anchoring phenomena for equitable 3D teaching*. Paper presented at the Colorado Science Education Network, Boulder, CO.
- Penuel, William R., Tiffany R. Clark, & Bronwyn Bevan. 2016. Designing for equitable learning across settings. *Afterschool Matters*, 24: 13-20.
- Penuel, William R., & Lorrie A. Shepard. 2016a. Assessment and teaching. In *Handbook of Research on Teaching*, Drew H. Gitomer and Courtney A. Bell (Eds.), 787-851. Washington, DC: AERA.
- Penuel, William R., & Shepard, L. A. (2016b). Sociocognitive and sociocultural models of assessment. In *Cognitive Models of Assessment*, André A. Rupp & Jacqueline P. Leighton (Eds.), 146-173. New York, NY: Wiley.
- Penuel, William R, Katie Van Horne, Jennifer J. Jacobs, and Michael Turner. 2018. *Developing a validity argument for practical measures of student experience in*

- project-based science classrooms*. Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY.
- Penuel, William R., Katie Van Horne, Sam Severance, David Quigley, and Tamara Sumner. 2016. Students' responses to curricular activities as indicator of coherence in project-based science. In *Proceedings of the 12th International Conference of the Learning Sciences*, Chee-Kit Looi, Joseph L. Polman, Ulrike Cress, & Peter Reimann (Eds.), Vol. 2, 855-858. Singapore: International Society of the Learning Sciences.
- Reiser, Brian J., Michael Novak, and Tara A. W. McGill. 2017. *Coherence from the students' perspective: Why the vision of the Framework for K-12 Science Education requires more than simply combining three dimensions of science learning*. Paper presented at the Board on Science Education Workshop on Instructional Materials for the Next Generation Science Standards, Washington, DC.
- Rosebery, Ann, Beth Warren, B., and Eli Tucker-Raymond. 2016. Developing interpretive power in science teaching. *Journal of Research in Science Teaching*, 53(10): 1571-1600.
- Schwarz, Christina V., Cynthia Passmore, and Brian J. Reiser. 2017. Moving beyond 'knowing about' science to making sense of the world. In *Helping students make sense of the world using next generation science and engineering practices*, Christina V. Schwarz, Cynthia Passmore, and Brian J. Reiser (Eds.), 3-21. Washington, DC: NSTA.
- Severance, Sam, William R. Penuel, Tamara Sumner, and Heather Leary. (2016). Organizing for teacher agency in curriculum design. *Journal of the Learning Sciences*, 25(4): 531-564.
- Shepard, Lorrie A., William R. Penuel, and James W. Pellegrino. 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.
- Tan, Edna, and Angela Calabrese Barton. 2012. *Empowering science and mathematics education in urban schools*. Chicago, IL: University of Chicago Press.
- Trevelyan, James. 2010. Reconstructing engineering from practice. *Engineering Studies*, 2(3): 175-195.
- Weizman, Ayelet, Yael Shwartz, and David Fortus. 2008. The driving question board: a visual organizer for project-based science. *The Science Teacher*, 75(8): 33-37.
- Zivic, Alicia, James F. Smith, Brian J. Reiser, Kelsey D. Edwards, Michael Novak, and Tara A. W. McGill. 2018. Negotiating epistemic agency and target learning goals: Supporting coherence from the students' perspective. In *Proceedings of the 13th International Conference of the Learning Sciences*, Judith Kay and Rose Luckin (Eds.), Vol. 1, 25-32. London, UK: International Society of the Learning Sciences.

### **Author Biographies**

**William R. Penuel** is Professor of Learning Sciences and Human Development at the School of Education at the University of Colorado Boulder. He is author (with Dan Gallagher) of *Research-Practice Partnerships in Education* (Harvard, 2017) and a handbook chapter, “Assessment and Teaching” (with Lorrie Shepard) that reviews evidence related to different approaches to formative assessment for improving learning outcomes.

**Douglas A. Watkins** is the High-School Science Curriculum Specialist for Denver Public Schools. He is affiliated with the Colorado Science Education Network, the National Science Teachers Association, and the National Science Education Leadership Association.

<p><b>Name:</b> Developing and Using a Driving Questions Board</p> <p><b>Actors and Goals:</b> The practice will support <i>teachers</i> and <i>students</i> in jointly constructing a learning pathway for the class that addresses the anchoring phenomenon for the unit and related phenomena from student experiences. The practice will also support <i>teachers</i> and <i>students</i> in holding each other accountable for answering students' questions.</p> <p><b>Stakeholders:</b> Students, teachers, district leaders, curriculum writers</p> <p><b>Trigger:</b> Teacher is introducing a new unit related to a disciplinary core idea in science.</p> <p><b>Main Success Scenario:</b></p> <ol style="list-style-type: none"><li>1. Teacher introduces the phenomenon to students via a video, demonstration, or an investigation that students conduct.</li><li>2. Teacher elicits students' initial 'noticings and wonderings' and records them.</li><li>3. Students make and share initial models of the phenomenon and identify additional questions.</li><li>4. Students discuss, share, and record individual knowledge and experiences of related phenomenon that can help them explain the anchoring phenomenon.</li><li>5. Students and the teacher build a public record of student questions, a Driving Questions Board, and generate initial ideas for how to investigate them.</li><li>6. Teacher and student return on a regular basis (at least every 3-4 classes) to the Driving Question Board at the beginning or end to see which ones have been answered and what new questions should be pursued.</li></ol> <p><b>Failure Conditions:</b> Most of the student questions are unrelated to what is in the storyline already. Some of the student questions could take students on a long detour from the planned route through the curriculum. Students construct a Driving Question Board, but it isn't used.</p> <p><b>Failure Handling:</b> Select anchors that connect to student interests and are perceived as relevant to them. Before developing storylines, implement the anchoring phenomenon routine with multiple groups of teachers. Use detours as opportunities for differentiation and make time for individual or small group inquiry. Develop a routine where responsibility for returning to the Driving Questions Board periodically is shared among students and the teacher.</p>
--

Figure 1.

Use Case 1: Building and Making Use of a Driving Questions Board

**Name:**

Monitoring and Supporting Equity and Epistemic Justice

**Actors and Goals:**

The practice will support *district leaders* and *external professional development partners* in monitoring variations in student experience within and across classrooms.

The practice will support *researchers* in developing and communicating understandings about conditions under which systems can become more equitable and just.

**Stakeholders:**

Students, teachers, district leaders, external partners (professional development providers and researchers), community groups advocating for education justice

**Trigger:**

Teacher comes to the end of a lesson that calls for using an electronic exit ticket.

**Main Success Scenario:**

1. Teacher asks students to take out their phones or computers and complete a brief survey of their experience in class.
2. After students from multiple classrooms have completed between 3 and 4 exit tickets, researchers perform analyses focusing on variation across and within classrooms with respect to coherence, relevance, and student contribution.
3. Researchers, professional development providers, and district leaders discuss patterns in data and design professional learning experiences that focus on areas needing attention.
4. Professional development providers guide professional learning experiences.

**Failure Conditions:**

Students lack adequate access to technology for completing surveys.

Teachers do not value and therefore do not assist with data collection.

Data become part of a punitive system of teacher evaluation, rather than used for improvement.

**Failure Handling:**

Provide alternate times and means for gathering data.

Include questions on survey that provide evidence related to student learning that is easy to grade.

Create data agreements with district regarding questions and analyses that can be performed; focus on aggregate patterns in data rather than reporting on individuals.

Figure 2.

Use Case 2: Monitoring and Supporting Equity and Epistemic Justice