

Running Head: DISTRICT PARTNERSHIP TO IMPROVE ASSESSMENT

A Research-Practice Partnership to Improve Formative Assessment in Science

William R. Penuel¹

University of Colorado Boulder

Angela Haydel DeBarger

SRI International

¹Correspondence for this chapter should be addressed to: Bill Penuel, University of Colorado Boulder, UCB 249, Boulder, CO 80309. Email: william.penuel@colorado.edu.

Abstract

This chapter describes an approach to supporting and investigating teacher learning of formative assessment in middle school Earth science. The approach illustrates a framework for conducting research on and supporting instructional improvement at scale in partnership with school districts. In the project, researchers worked in partnership with teachers and leaders in a large urban district to design and test classroom assessment resources intended to improve the efficacy of district-adopted curriculum materials. The project tested the value the resources add to the curriculum in a small-scale, quasi-experimental field trial. This chapter highlights both the potential and challenges of research-practice partnerships for promoting district-wide improvements to classroom assessment, highlighting the ways the partnership supported efforts to create vertical, horizontal, and developmental coherence to a district's assessment system.

A Research-Practice Partnership to Improve Classroom Assessment in Science

Many district reform efforts today focus on improving instruction as a strategy for improving learning outcomes for all students (Corcoran, Fuhrman, & Belcher, 2001; Honig, Copland, Rainey, Lorton, & Newton, 2010; Supovitz, 2006). A key aim of many of these efforts is to bring greater *coherence* to the key elements of the local educational system¹, especially standards, curriculum, assessment, and professional development (Fuhrman, 1993; Rorrer, Skyrla, & Scheurich, 2008; Supovitz, 2006). In addition, a key district function in instructional improvement is to *promote equity* with respect to student opportunities to learn, by supporting uniformly high quality teaching across all classrooms and schools (Rorrer et al., 2008; Supovitz, 2006; Talbert & McLaughlin, 1993; Trujillo, 2011). Developing coherence and ensuring equity, research suggests, requires that districts develop a vision for high-quality instruction and build the commitment and capacity of teachers to enact and evolve that vision (Cobb & Jackson, 2012; David & Shields, 2001; Supovitz, 2006).

At present, most districts rely on external partners to provide services and resources to help realize their visions for instructional improvement. In part, this is because district central offices historically have been organized principally to manage schools, and they are more limited in their capacity to provide vision, focus, and support for improving instruction in schools (Crowson & Morris, 1985; Honig, 2013). External partners often can bring broader experience, resources, and technical expertise to improvement efforts than district leaders can (Supovitz, 2006). This is especially true for services that require both specialized knowledge and extensive resources, such as curriculum design and assessment development.

In this chapter, we describe a kind of arrangement between districts and external partners that is organized to provide long-term support for district reform initiatives and to support

instructional coherence. In *design research partnerships* like the one we describe, district leaders, teachers, and researchers collaboratively design of resources for improving instruction that address jointly negotiated problems of practice. Such partnerships' provide subject matter, design, and research expertise to districts' own efforts to bring coherence to curriculum, instruction, assessment, and professional development. As we illustrate with an example of a partnership in which we have been engaged since 2008, these partnerships can produce tools that teachers find useful, adaptable to diverse students' needs and strengths, and that support short-term improvements to instructional practice aligned to a district vision.

Design Research Partnerships in District Reform

Research-practice partnerships are long-term collaborations between practitioners and researchers that are intentionally organized to investigate problems of practice and solutions for improving district outcomes (Coburn, Penuel, & Geil, 2013). Policymakers and funders see promise in the potential of research-practice partnerships to enable greater use of research evidence in decision making (e.g., Tseng, 2012). Advocates from within the research community argue that such partnerships can address persistent problems of practice and improve educational outcomes (Bryk, 2009; Donovan, 2013).

Design research partnerships are a type of research-practice partnership. A key aim of design research is to develop theories of the process of learning and the means of supporting these processes (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; Design-Based Research Collective, 2003; Kelly, Baek, Lesh, & Bannan-Ritland, 2008; Sandoval, 2014; Sandoval & Bell, 2004). Though most design research focuses either on a single classroom or a few classrooms (e.g., Confrey & Lachance, 2000), in design research partnerships, efforts can span an entire school district (e.g., Cobb & Jackson, 2012).

In design research partnerships, collaborative, iterative design is a leading activity. That is, teachers, educational leaders, and researchers work together to design, implement, and test strategies for improving teaching and learning outcomes in a school district. The focus of the designs, moreover, is on problems of practice facing district leaders (Donovan, 2013). As in action research, a key aim of research in design research partnerships is to develop knowledge about strategies that is relevant to leaders in the district. Design research partnerships also aim to generate understandings of principles and mechanisms for change through systematic inquiry that can be used elsewhere and in the future by district leaders themselves to support change (Bryk, 2009; Penuel, Fishman, Cheng, & Sabelli, 2011).

How Design Research Partnerships Can Support Coherence

Design research partnerships have the potential to support district leaders' efforts to build coherence among key elements of the educational system: standards, curriculum, assessment, and professional development. For example, the collaborative design process employed in design partnerships can increase the likelihood that new tools developed to support any one of these elements will fit together into a harmonious and logical whole, a key condition for coherence (cf., National Research Council, 2012). The collaborative design process can also help develop teachers' sense of ownership over new tools that can support common goals for instructional improvement (Penuel, Roschelle, & Shechtman, 2007). This is significant, because successful district reform depends on districts' balancing guidance with efforts to build ownership of reform goals (David & Shields, 2001).

But building a coherent system requires more than bringing key elements into alignment with reform goals that are decided ahead of time. Implementation of district-level reforms creates its own problems that must be addressed in real-time, and so district leaders must engage in

continuous diagnosis of problems and devise new strategies for dealing with them (Spillane & Coldren, 2010). Creating coherence among different elements of the system is a “never-ending grind” that demands opportunities for individual actors across multiple levels of a district gaining a “grasp of the whole” that guides their moment-to-moment decision making (Kirp, 2013, pp. 13, 10). It is about having resources at hand to solve new problems that enable both people and systems to learn as they design and implement innovations (Lissack & Letiche, 2002).

Design research partnerships have the potential to support these aspects of building coherence as well. For example, research activities can provide feedback about how well strategies are working and about how widely shared particular visions of reform are among teachers (Cobb, Jackson, Smith, Sorum, & Henrick, 2013). In addition, organizational routines that engage researchers, district leaders, and teachers in collaborative sensemaking about emerging challenges and opportunities can help individuals get a better grasp of the “whole environment” (Glazer & Peurach, 2013).

How Design Research Partnerships Can Promote Equity

Design research partnerships also have the potential to broaden access to opportunities to learn for all students, a key dimension of equity. Opportunities to learn encompass access to the resources, practices, and skilled guidance needed to develop proficiency in a given domain (Guiton & Oakes, 1995). Opportunities to learn are defined both by participation in learning activities within classrooms and by students’ access to classrooms and courses that are shaped by district policies such as tracking (Hand, Penuel, & Gutiérrez, 2012; Oakes, 1990; Quiroz & Secada, 2003).

Design research partnerships can support a focus on equitable opportunity to learn in multiple ways. Innovations collaboratively designed for classrooms can incorporate specific

equity goals, such as drawing on a wide range of students' "funds of knowledge" to organize subject matter instruction (e.g., Civil, 2007) and promoting more equitable participation in classroom discussions (e.g., Hoadley, 2004). Research can investigate whether innovations produce different kinds of outcomes for different student groups (e.g., Roschelle et al., 2010) and explore how and when participation helps students from diverse backgrounds identify with subject matter (Nolen, Tierney, Goodell, Lee, & Abbott, 2014). There are examples of partnerships helping to broaden students' access to advanced courses, though in these instances, partnerships have been between researchers and community groups rather than districts (e.g., Oakes & Rogers, 2006).

Case Study: A Design Research Partnership with Denver Public Schools

We investigate the claims for the potential of design research partnerships to support district efforts to create coherence among standards, curriculum, assessment, and professional development and to promote equitable participation in classrooms within the context of an ongoing (as of 2014) research-practice partnership that began in 2007. The partnership was formed initially between the authors of this chapter and district curriculum leaders in Denver Public Schools when both authors were research scientists at SRI International in Menlo Park, California. Today, the partnership between the first author and district leader continues, with the support of a research team at the first author's current institution, the University of Colorado Boulder.

Initial Focal Problem of Practice: Improving Classroom Assessment

The initial focus of the partnership's work was on improving classroom assessment, a goal of mutual concern to researchers and district leaders. Researchers are particularly interested in improving classroom assessment, because there is strong evidence that classroom-based

assessment can improve student outcomes (Black & Wiliam, 1998a; Crooks, 1988; Fuchs & Fuchs, 1986; Kingston & Nash, 2011), but there is also evidence that improving assessment in ways that research suggests is necessary is challenging to teachers (e.g., Penuel & Gallagher, 2009; Yin et al., 2008). District leaders, for their part, are interested in improving alignment among standards, classroom assessment, curriculum, and district assessment practices. In Denver, as in many other districts, (e.g., Black & Harrison, 2001; Brandon et al., 2008; Penuel, McWilliams, et al., 2009) the assessment component includes annual standardized achievement tests designed and scored by state officials but administered by district staff, interim or benchmark tests designed, administered, and scored by the district, and quizzes and tests that teachers give in their classrooms (Goertz, 2009).

In most schools and districts, system-level coherence and coherence within assessment system are difficult to achieve. Policies and practices of assessment may not cohere with policies and practices with respect to curriculum, instruction, and professional development (*horizontal* incoherence). There may also be limited agreement among the purposes of assessment among actors (e.g, curriculum and instruction leaders, principals, teachers) at different levels of systems (*vertical* incoherence). Assessment systems may also include parts that are inconsistent with research about how children's understanding develops over time (*developmental* incoherence). Researchers presume that assessment systems necessarily have multiple interacting parts, but they also warn that improvement is not possible unless assessment systems are horizontally, vertically, and developmentally coherent (Herman, 2010; National Research Council, 2006, 2012). Thus, incoherence along any of these dimensions threatens the success of interventions to improve formative assessment in individual classrooms.

(Coburn et al., 2013) The *Contingent Pedagogies* project was an NSF-funded project that brought University of Colorado at Boulder and SRI researchers together with district curriculum leaders, a curriculum publisher, and subject matter experts to investigate how using classroom response systems (CRSs or “clickers”) could support classroom assessment with two units of *Investigating Earth Systems*, a middle school Earth Systems science curriculum. The intervention we developed focuses on the core disciplinary idea of Earth systems and answers the question, “How and why is Earth constantly changing?” (National Research Council, 2012, p. 179). The aim was to support five steps of formative assessment posited as essential or critical for improving learning outcomes (Black & Wiliam, 2009) among sixth grade science teachers in a single school district:

- Teachers elicit student thinking;
- Students respond to teachers’ elicitation activities, revealing their thinking;
- Teachers interpret students’ responses to make sense of where students are relative to their goals for student learning;
- Teachers’ take action (e.g., trying a new strategy) on the basis of their interpretation to move students in the desired direction; and
- Teachers’ re-assess student understanding to measure success of the action.

As with many other large urban districts, the district office in Denver Public Schools is organized into multiple departments, and our partnership required coordination of departments related to science curriculum and instruction and instructional technology. We had infrequent and peripheral contact with the office of assessment, evaluation, and accountability in the district. As in many other districts as well, there were multiple assessment strategies being implemented. These included state assessments given every few years in science, as well as

district benchmark assessments that were introduced toward the end of our study that teachers were expected to administer.

As researchers, we began with the premise that if Contingent Pedagogies were to be brought successfully to scale and sustained, we would need to anticipate and address the ways that the suite of tools could cohere with and support district goals and assessments. We therefore organized the project as a research-practice partnership with the district, and we set about making sure that Contingent Pedagogies supported horizontal, developmental, and vertical coherence in the district. Ours is a design research partnership, in which teachers and district leaders contributed to the overall design of the tools and supported their implementation and testing. Through the collaborative design process and in implementation, equity emerged as a concern of teachers, and we incorporated strategies for promoting equitable participation in classrooms into designs, particularly for the growing English Language Learner population in the school system.

Below, we describe the ways we organized the Contingent Pedagogies project to (1) build horizontal, developmental, and vertical coherence and (2) promote equity.

Promoting Horizontal, Developmental, and Vertical Coherence

We were particularly purposeful about promoting horizontal and developmental coherence in structuring the partnership but somewhat less successful in efforts to structure participation in such a way as to promote vertical coherence. We organized the assessment purposes to fit within district goals for science, using a particular perspective on the development of student understanding in science (knowledge-in-pieces or facets-based view of cognitive development; diSessa & Minstrell, 1998a; Minstrell, 1992). We structured design to include extensive

involvement of a cadre of teacher leaders and to include selected district and school leaders in helping shape the project.

Strategies for promoting horizontal coherence. Horizontal coherence refers to alignment among four key interrelated elements pertaining to the “technical core” of education: curriculum, assessment, instruction, and professional development (National Research Council, 2006). A key aim of instructional improvement efforts for science education in Denver Public Schools was to build children’s understanding of big ideas through a constructivist, inquiry-oriented approach to teaching. Each grade level in the middle grades focuses on core ideas in a different domain of science: sixth grade on Earth science, seventh grade on life science, and eighth grade on physical science. Students are expected to learn through direct engagement with phenomena, through direct, guided inquiry and teacher demonstrations.

For the district leaders in the partnership, the district-adopted curriculum materials were essential supports for helping teachers meet the student learning goals reflected in standards documents; they expected us to use those materials to anchor our design work together. The particular curriculum materials used in Contingent Pedagogies are taken from a 10-module research-based middle school curriculum called *Investigating Earth Systems* published by It’s About Time, Inc. A diverse team of scientists, curriculum developers, and teachers led by the American Geological Institute developed the materials. The National Science Foundation, the American Geological Institute Foundation, and the Chevron Corporation supported its development. Both initial field tests and subsequent research have found that when combined with high-quality professional development, use of the *Investigating Earth Systems* curriculum materials can have positive impacts on science teaching and learning (Penuel & Gallagher, 2009; Penuel, Gallagher, & Moorthy, 2011; Penuel, McWilliams, et al., 2009). Implementation of the

curriculum was well supported by the district, with all teachers receiving professional development from the curriculum developers and restocking of materials and supplies each year for students to conduct investigations.

Prior to the partnership's formation, district leaders expected teachers to use the adopted curriculum materials' embedded assessments formatively to inform their instruction. From the vantage point of the researchers as well as the teacher leaders, the curriculum-embedded activities and assessments provided few opportunities for students to relate hands-on investigations to disciplinary core ideas or to science practices. District leaders who agreed to be part of the original proposal effort were interested in learning whether tools we would develop in collaboration with their teachers could improve both teachers' classroom assessment practice and student learning from the curriculum materials. Thus, we partnered with the curriculum developers to improve the assessment activities and developed a research design aimed at comparing changes to teacher practice and student learning in classrooms where teachers would receive the additional assessment resources to those who did not. In both groups, however, teachers would implement the adopted *Investigating Earth Systems* curriculum units chosen for the sixth grade by district leaders.

Embedding assessment activities within the district-adopted curriculum was a purposeful and important strategy for promoting horizontal coherence. We hypothesized that assessment activities we developed would be more likely to support the goals the district's goals for student learning and thus would be consistent with teachers' goals for instruction. In addition, we hypothesized that the assessment activities would fit easily within the flow of instruction, because they were embedded in particular places within the curriculum materials.

Strategies for promoting developmental coherence. Developmental coherence refers to how systems help build and assess student learning over time (National Research Council, 2006). The curriculum materials themselves provided one source of coherence at the unit level. Each unit is organized into seven or eight investigations that provide opportunities for students to participate in science practices related to one or two disciplinary core ideas in Earth science. The Contingent Pedagogies project supported teachers' implementation of two units from the curriculum, *Our Dynamic Planet* and *Rocks and Landforms*.

A key goal of assessment activities we designed was to promote more opportunities for eliciting and developing students' conceptual models about disciplinary core ideas that related to the investigations. We employed a *facet-based* perspective on student cognitive development for purposes of developing questions to elicit and develop student thinking. A facet is a construction of one or more pieces of knowledge by a learner in order to solve a problem or explain an event (diSessa & Minstrell, 1998b). Facets that are related to one another can be organized into clusters, and the basis for grouping can either be an explanation or interpretation of a physical situation or around a disciplinary core idea (Minstrell & Kraus, 2005). The facets perspective assumes that, in addition to problematic thinking, students also possess insights and understandings about the core disciplinary idea that can be deepened and revised through additional learning opportunities (Minstrell & van Zee, 2003).

While researchers developed many of the facet clusters using traditional data collection and analysis methods, teachers in the study played key roles in designing questions that would elicit student facets as they introduced new topics in the curriculum and for review. The aim of these questions was to provide evidence of student learning for shorter timescales of development (3-4 days). To that end, researchers and teachers collaboratively designed diagnostic *elicitation*

questions intended to identify aspects of disciplinary core ideas students understood at the beginning of the lesson. They also helped to design what we called *reflect-and-revise* questions that checked student understanding of core ideas at the conclusion of an investigation. Teachers were particularly valuable to the partnership in identifying problem or question contexts that would likely captivate students' attention and in developing wording that could be understood by a wide range of learners. Their contributions point to an important dimension of developmental coherence that is difficult for reviewers to know ahead of time, namely the fit of the challenge of particular activities and vocabulary to the current capabilities of particular groups of students.

Strategies for promoting vertical coherence. Vertical coherence in an educational system is evident when there is agreement on the key goals for student learning and on the purposes and uses of assessment across actors in different levels of the system (National Research Council, 2012). Users of assessment data tend to have different needs, depending on their level within the system. District leaders, for example, may use assessments for accountability purposes and for improving programs. Teachers, by contrast, may use assessment to gauge individual progress in learning and to adjust instruction targeted to individuals, groups, and the whole class. In principle, these purposes for assessment can be brought into alignment, but the process takes time and typically requires multiple assessment instruments and coordination of activities across levels of the system (Herman, 2010; National Research Council, 2006).

Initially, the researchers had approached the curriculum developers of *Investigating Earth Systems* about conducting a study focused on embedding technology-supported assessments into selected units of study. The researchers asked the developers to help identify a district that might be interested; they readily identified Denver Public Schools. From the district's perspective, being part of the project fit into their ongoing improvement efforts within the district, especially

as researchers refined the goals to match theirs for the study. Once funded, however, the district's main champion for the curriculum and the study had left; researchers had to develop a relationship with the new science curriculum coordinator. The coordinator was open to being part of the project, though the research team was initially uncertain about whether the project could succeed with its champion gone. To help build teacher support, researchers asked the new coordinator to nominate teachers who could participate in co-designing the assessments with the research team. She selected two sixth grade teachers who were part of another district reform initiative to digitize and organize curriculum resources in science.

During fall 2008, researchers assembled a design team comprised of the study PI and co-PI, four researchers, three technology specialists, four content experts from partner organizations, one teacher leader, and two co-design teachers from the district. The design process began with a researcher visiting each co-design teacher's classroom visited to conduct an interview and classroom observation. The interview covered teaching experience, teaching approach and practices, details of a recently taught Earth science unit, access to and use of technology at the school, and school and district context. The observation protocol was designed to capture details of classroom organization and teacher and student interactions so that design work could target real-life situations. To further ground design in classroom realities, the technology support lead at each school completed a detailed survey about technology at the school, what was available, how it was used and supported.

We then held a series of design meetings over the course of two years, in which we structured processes for developing assessment activities for the two focal *Investigating Earth Systems* modules. An additional three teachers from the district joined in the second year as co-designers. As in all forms of design research (Cobb et al., 2003), the design process was iterative:

we developed assessment activities together, teachers tested them in classrooms, and on the basis of their classroom testing, we revised the activities. Over the course of the two years, we made significant changes to the form of activities, their placement within modules, and the ways that technology was expected to support activities. The designs we used in the field trial with the co-design teachers focused on assessment opportunities at the beginning and end of each investigation, and teachers used “clickers” or student response systems to pose questions that were intended to elicit common facets of student understanding and, in turn, spark rich discussion of the diversity of student ideas and how they related to student investigations.

Equity as an Emergent Priority in the Research-Practice Partnership

Some of the teachers who participated in the design of Contingent Pedagogies tools had large percentages of emerging bilinguals (English Language Learners) in their classrooms. They expressed concern over the language demands both of the existing curriculum and of the new tools we were developing to promote participation in classroom discussions. Their concerns included worries about whether emerging bilingual students would be able to follow fast-moving discussions about difficult disciplinary ideas. The teachers suggested that whole class discussions might benefit from being broken up into segments that included pair and small group talk. As the teachers and some researchers studying classroom discussions have argued (Michaels & O'Connor, 2011), providing these additional formats could provide time for emerging bilinguals to develop and rehearse their thoughts. The reporting out of small group discussions, moreover, could help students keep up in discussion, because they would hear fewer ideas and some ideas that were consistent and repeated being reported out.

We incorporated recommendations to use these different formats into a revised set of “pedagogical patterns” or teaching routines (DeBarger, Penuel, Harris, & Schank, 2010) that we

provided to teachers in a field trial as part of their professional development. The use of the expanded formats for “thinking time” proved popular with teachers, who incorporated a wider range of formats than did their counterparts in comparison classrooms when orchestrating classroom discussions (Penuel et al., 2013).

Key Partnership Outcomes

One of the key ways we measured the impact of the partnership was through a study of the effects of Contingent Pedagogies in which we tested the tools with a wide range of teachers in different schools. In the last year of the project, district leaders helped us to recruit 7 additional science teachers to the project as part of a field trial. All teachers who were recruited attended a two-day summer professional development workshop where they were introduced to Contingent Pedagogies. During the first day of the workshop, the teachers participated and practiced as students in using the clicker technologies and activities they would be implementing, while discussing ideas about classroom assessment. They learned about facet-based instruction and how they can be used to support classroom assessment. Discussion and hands on activities about how to elicit and develop student thinking through a set of “talk moves” (Michaels & O'Connor, 2011) were also part of the workshop.

The teachers also had professional development and implementation support throughout the year. During the school year following the workshops the teachers joined 14 two-hour teleconferences. Teleconferences were used to discuss current classroom activities, technology issues, challenges and problems, reflections on the activities, tips for implementation, reminders, and research activities. Quick tips in the form of a weekly newsletter were provided for fourteen different weeks during the implementation of the project to support facet-based instruction and provide refreshers on the workshop material. Beyond discussing classroom strategies, these tips

“unpacked” some of the clicker questions to focus on the ways the teacher could engage students to think deeper about the substance of Earth science content.

During the field trial, we collected data on the perceived usability and value of the Contingent Pedagogies suite of tools and on the impact on teaching and learning. To analyze impacts on teaching and learning, we recruited a group of comparison teachers to participate in research activities. More detailed descriptions of the research design and findings appear in peer reviewed publications (Penuel, Beauvineau, DeBarger, Moorthy, & Allison, 2012; Penuel, Moorthy, DeBarger, Beauvineau, & Allison, 2012) and online at <http://contingentpedagogies.org>. In short, we found good evidence that teachers perceived the tools to be usable and valuable for helping them accomplish their instructional goals. There was a close relationship, moreover, between the tools teachers found valuable and those they actually used with students. Our analysis of student learning outcomes found small to medium effects of the tools (+0.25 SD for one unit and +0.45 SD for the second), as measured by tests we developed that consisted of a mix of multiple-choice and constructed response items.

Despite the promising results of the study, there has been little uptake of the tools beyond our initial field trial, and the district ultimately decided to implement an alternative approach to formative assessment in middle school science that focused on common tasks related to the district’s standards. Anecdotally, we know that some of the teachers continue to use the tools, but many also have become part of other reform initiatives, a pattern that many other research and development efforts have documented.

In addition, although the co-design process created a strong, common sense of purpose among the participating teachers, we were ultimately less successful in aligning the project to the district’s priorities. Concurrent to our research and development efforts were initiatives that were

complementary to ours (notably one focused on promoting academic language development in middle school science) in the district that ultimately district leaders decided were more central to district goals. Some of these competing initiatives involved science instruction, and one involved formative assessment in science. In retrospect, we concluded we had not involved district leaders enough in the research and development process to foster vertical coherence. We did not have sufficient contact with the supervisor of the science coordinator in the district, nor did we involve district staff who were charged with developing district-level assessments, which might have facilitated better coherence among assessments teachers were using. We had also failed to engage principals, who held considerable authority for allowing teachers to participate in the research.

Lessons Learned and Implications for Research-Practice Partnership Designs

In our view, the findings from our study underscore the importance of each of the different kinds of coherence in systems—horizontal, developmental, and vertical. We had considerable success in promoting horizontal coherence because we worked with existing, district-adopted curriculum to improve formative assessment. Had we introduced assessments incompatible with the curriculum or ignored the core ideas and science practices that were focal in the materials, it is unlikely teachers would have found the tools useful in supporting their learning goals. Co-design proved an effective strategy for enhancing developmental coherence, by helping to calibrate researchers' expectations of students with teachers' perceptions of student capabilities.

At the same time, we can conclude that a well-designed suite of formative assessment activities closely aligned to district standards and adopted curriculum and informed by learning theory cannot be sustained unless there is a shared understanding of its goals and contributions at all levels of the system. We needed more involvement not only of a more diverse group of

district leaders in the process, but also principals in the district, who have significant power to influence teachers' participation in professional development activities.

With respect to equity, our studies point to the need for additional research and development related to a specific aspect of classroom assessment, namely participation in classroom discussions. As researchers, we did not anticipate the need to focus specifically on tools and supports for equitable participation of emerging bilinguals in discussions of student ideas, and so this aspect of our work got less attention than it should have. Since concluding our study, we have developed a partnership with a second school district that has adopted the *Investigating Earth Systems* materials and are partnering with teachers on a set of resources for supporting emerging bilinguals in practices of argumentation in science classroom. The tools are intended to augment those in the Contingent Pedagogies toolkit.

One of us remains involved in a research-practice partnership with the district as part of a subsequent research and development project, the Inquiry Hub, also funded by the National Science Foundation. The purpose of the Inquiry Hub is to develop and test models for helping teachers to make principled adaptations to curriculum materials using a digital platform for sharing and adapting resources. Formative assessment activities inspired by the Contingent Pedagogies approach will be part of the Inquiry Hub, though not the specific materials we developed as part of the project, because the grade level focus is different for the new project. Importantly, the design process is structured differently than in Contingent Pedagogies: there are regular meetings between district leaders and researchers, as well as regular meetings with teachers who are helping to design and test models for adapting curriculum materials. Our early qualitative analyses of discussions in meetings provide evidence of greater shared understanding of purpose across the two different role groups of district leaders and teachers.

Our involvement in this research-practice partnership provides us with cautious optimism about their promise. On the one hand, the fate of our particular project tools is similar to that of many research and development projects that are developed with less involvement of district leaders and teachers. In most projects, interventions are not sustained, particularly as district priorities and personnel who champion particular interventions change positions or leave. Yet, we have been able to develop partnerships with a new district where these tools will be accessible to a large number of teachers and students in the coming years. Moreover, within Denver, the relationships built remain, as does a joint commitment on the part of researchers and district leaders to work together to address current problems of practice facing the district. Realizing this commitment requires ongoing negotiation about the focus of joint work. But the foundation built through past work provides a solid basis for engaging in these negotiations, as does a shared understanding across the research-practice divide of the importance of both achieving system coherence and supporting teachers across the district in their efforts to improve their own practice.

Endnote

¹We use the term “system” to refer to the set of interrelated components that make up a school district. There are many kinds of systems and ways of conceptualizing school districts, including as “complex adaptive systems” (McDaniel, 2007). In our view, it may or may not be useful to characterize school districts as complex adaptive systems, given that bureaucratic governance structures limit the potential for self-organization. What matters for the present argument is that designing for one component requires partnerships to consider the fit of that component with other elements in the system, because parts are connected both in policies and in how teachers make sense of them.

References

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, & Accountability*, 21(1), 5-31.
- Bryk, A. S. (2009). Support a science of performance improvement. *Phi Delta Kappan*, 90(8), 597-600.
- Civil, M. (2007). Building on community knowledge: An avenue to equity in mathematics education. In N. i. S. Nasir & P. Cobb (Eds.), *Improving access to mathematics: Diversity and equity in the classroom* (pp. 105-117). New York: Teachers College Press.
- Cobb, P. A., Confrey, J., diSessa, A. A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9-13.
- Cobb, P. A., & Jackson, K. (2012). Analyzing educational policies: A learning design perspective. *Journal of the Learning Sciences*, 21, 487-521.
- Cobb, P. A., Jackson, K., Smith, T., Sorum, M., & Henrick, E. C. (2013). Design research with educational systems: Investigating and supporting improvements in the quality of mathematics teaching at scale. In B. J. Fishman, W. R. Penuel, A.-R. Allen & B. H. Cheng (Eds.), *Design-based implementation research: Theories, methods, and exemplars. National Society for the Study of Education Yearbook*. (pp. 320-349). New York, NY: Teachers College Record.
- Coburn, C. E., Penuel, W. R., & Geil, K. (2013). Research-practice partnerships at the district level: A new strategy for leveraging research for educational improvement. Berkeley, CA and Boulder, CO: University of California and University of Colorado.

- Confrey, J., & Lachance, A. (2000). Transformative teaching experiments through conjecture-driven research design. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 231-266). Mahwah, NJ: Erlbaum.
- Corcoran, T. B., Fuhrman, S., & Belcher, C. L. (2001). The district role in instructional improvement. *Phi Delta Kappan*, 83(1), 78-84.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481.
- Crowson, R., & Morris, V. C. (1985). Administrative control in large-city school systems: An investigation of Chicago. *Educational Administration Quarterly*, 21, 51-70.
- David, J. L., & Shields, P. M. (2001). *When theory hits reality: Standards-based reform in urban districts*. Menlo Park, CA: SRI International.
- DeBarger, A., Penuel, W. R., Harris, C. J., & Schank, P. (2010). Teaching routines to enhance collaboration using classroom network technology. In F. Pozzi & D. Persico (Eds.), *Techniques for fostering collaboration in online learning communities: Theoretical and practical perspectives* (pp. 222-244). Hershey, PA: IGI Global.
- Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5-8.
- diSessa, A. A., & Minstrell, J. (1998). Cultivating conceptual change with benchmark lessons. In J. G. Greeno & S. V. Goldman (Eds.), *Thinking practices in learning and teaching science and mathematics* (pp. 155-187). Mahwah, NJ: Erlbaum.
- Donovan, M. S. (2013). Generating improvement through research and development in educational systems. *Science*, 340, 317-319.

- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287-312.
- Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53(3), 199-208.
- Fuhrman, S. H. (1993). The politics of coherence. In S. H. Fuhrman (Ed.), *Designing coherent educational policy: Improving the system* (pp. 1-34). San Francisco, CA: Jossey-Bass.
- Glazer, J. L., & Peurach, D. J. (2013). School improvement networks as a strategy for large-scale education reform: The role of educational environments. *Educational Policy*, 27(4), 676-710.
- Goertz, M. E. (2009). *Overview of current assessment practices*. Paper presented at the Workshop of the Committee on Best Practices in State Assessment Systems: Improving Assessment while Revisiting Standards, National Research Council, Washington, DC.
- Guiton, G., & Oakes, J. (1995). Opportunity to learn and conceptions of educational equality. *Educational Evaluation and Policy Analysis*, 17(3), 323-336.
- Hand, V., Penuel, W. R., & Gutiérrez, K. D. (2012). (Re)framing educational possibility: Attending to power and equity in shaping access to and within learning opportunities. *Human Development*, 55, 250-268.
- Herman, J. L. (2010). *Coherence: Key to Next Generation Assessment Success (AACC Report)*. Los Angeles, CA: University of California.
- Hoadley, C. M. (2004). Methodological alignment in design-based research. *Educational Psychologist*, 39(4), 203-212.
- Honig, M. I. (2013). Beyond the policy memo: Designing to strengthen the practice of district central office leadership for instructional improvement at scale. In B. J. Fishman, W. R.

- Penuel, A.-R. Allen, & B. H. Cheng (Eds). *Design-based implementation research. National Society for the Study of Education Yearbook, 112(2), 256-273.*
- Honig, M. I., Copland, M., Rainey, L. R., Lorton, J. A., & Newton, M. (2010). *Central office transformation for district-wide teaching and learning improvement.* Seattle, WA: The Center for the Study of Teaching and Policy.
- Jackson, K., & Cobb, P. (2013). Coordinating professional development across contexts and role group. In M. Evans (Ed.), *Teacher education and pedagogy: Theory, policy and practice* (pp. 80-99). New York, NY: Cambridge University Press.
- Kelly, A. E., Baek, J. Y., Lesh, R. A., & Bannan-Ritland, B. (2008). Enabling innovations in education and systematizing their impact. In A. E. Kelly, R. A. Lesh & J. Y. Baek (Eds.), *Handbook of design research methods in education* (pp. 3-18). New York: Routledge.
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30(4), 28-37.*
- Kirp, D. L. (2013). *Improbable scholars: The rebirth of a great American school system and a strategy for America's schools.* New York, NY: Oxford University Press.
- Lissack, M. R., & Letiche, H. (2002). Complexity, emergence, resilience, and coherence: Gaining perspective on organizations and their study. *Emergence, A Journal of Complexity Issues in Organizations and Management, 4(3), 72-94.*
- McDaniel, R. R. (2007). Management strategies for complex adaptive systems sensemaking, learning, and improvisation. *Performance Improvement Quarterly, 20(2), 21-41.*
- Michaels, S., & O'Connor, C. (2011). *Talk science primer.* Cambridge, MA: TERC.

- Minstrell, J. (1992). Facets of students' knowledge and relevant instruction. In F. Duit, F. Goldberg & H. Niedderer (Eds.), *Research in physics learning: Theoretical issues and empirical studies* (pp. 110-128). Kiel, Germany: IPN.
- Minstrell, J., & Kraus, P. (2005). Guided inquiry in the science classroom. In National Research Council (Ed.), *How students learn: History, mathematics, and science in the classroom* (pp. 475-514). Washington, DC: National Academies Press.
- Minstrell, J., & van Zee, E. (2003). Using questioning to assess and foster student thinking. In J. M. Atkin & J. Coffee (Eds.), *Everyday assessment in the science classroom* (pp. 61-74). Arlington, VA: NSTA Press.
- National Research Council. (2006). *Systems for state science assessment*. Washington, DC: National Academies Press.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Research Council.
- National Research Council. (2013). *Developing assessments for the Next Generation Science Standards*. Washington, DC: National Academies Press.
- Nolen, S. B., Tierney, G., Goodell, A., Lee, N., & Abbott, R. D. (2014). Designing for engagement in environmental science: Becoming “environmental citizens”. In J. L. Polman, E. A. Kyza, D. K. O'Neill, I. Tabak, W. R. Penuel, A. S. Jurow, K. O'Connor, T. F. Lee & L. D'Amico (Eds.), *Learning and becoming in practice: The International Conference of the Learning Sciences (ICLS) 2014* (Vol. 2, pp. 962-966). Boulder, CO: International Society of the Learning Sciences.
- Oakes, J. (1990). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science*. Santa Monica, CA: RAND.

- Oakes, J., & Rogers, J. (2006). *Learning power: Organizing for education and justice*. New York, NY: Teachers College Press.
- Penuel, W. R., Beauvineau, Y., DeBarger, A. H., Moorthy, S., & Allison, K. (2012). Fostering teachers' use of talk moves to promote productive participation in scientific practices. In J. van Aalst, K. Thompson, M. J. Jacobson & P. Reimann (Eds.), *The future of learning: Proceedings of the 10th international conference of the learning sciences (ICLS 2012) – Volume 2, short papers, symposia, and abstracts* (Vol. 2, pp. 505-506). Sydney, Australia: ISLS.
- Penuel, W. R., DeBarger, A., Kim, C. B., Moorthy, S., Beauvineau, Y., Kennedy, C. A., . . . Allison, K. (2013, April). *Improving learning by improving classroom assessment in Earth science: Findings from the Contingent Pedagogies project*. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, San Juan, PR.
- Penuel, W. R., Fishman, B. J., 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., & Gallagher, L. P. (2009). Preparing teachers to design instruction for deep understanding in middle school Earth science. *The Journal of the Learning Sciences*, 18(4), 461-508.
- Penuel, W. R., Gallagher, L. P., & Moorthy, S. (2011). Preparing teachers to design sequences of instruction in Earth science: A comparison of three professional development programs. *American Educational Research Journal*, 48(4), 996-1025.

- Penuel, W. R., McWilliams, H., McAuliffe, C., Benbow, A., Mably, C., & Hayden, M. M. (2009). Teaching for understanding in Earth science: Comparing impacts on planning and instruction in three professional development designs for middle school science teachers. *Journal of Science Teacher Education, 20*(5), 415-436.
- Penuel, W. R., Moorthy, S., DeBarger, A., Beauvineau, Y., & Allison, K. (2012, July). *Tools for orchestrating productive talk in science classrooms*. Paper presented at the Workshop on Classroom Orchestration: Moving Beyond Current Understanding of the Field, at the International Conference of the Learning Sciences, Sydney, Australia.
- Penuel, W. R., Roschelle, J., & Shechtman, N. (2007). The WHIRL co-design process: Participant experiences. *Research and Practice in Technology Enhanced Learning, 2*(1), 51-74.
- Quiroz, P. A., & Secada, W. G. (2003). Responding to diversity. In A. Gamoran, C. W. Anderson, P. A. Quiroz, W. G. Secada, T. Williams & S. Ashman (Eds.), *Transforming teaching in math and science: How schools and districts can support change* (pp. 87-104). New York, NY: Teachers College Press.
- Rorrer, A. K., Skyrla, L., & Scheurich, J. J. (2008). Districts as institutional actors in educational reform. *Educational Administration Quarterly, 44*(3), 307-357.
- Roschelle, J., Pierson, J., Empson, S., Shechtman, N., Dunn, M., & Tatar, D. (2010). Equity in scaling up SimCalc: Investigating differences in student learning and classroom implementation. In K. Gomez, L. Lyons & J. Radinsky (Eds.), *Learning in the disciplines: Proceedings of the 9th International Conference of the Learning Sciences* (Vol. 1, pp. 333-340). Chicago, IL: International Society of the Learning Sciences.

- Sandoval, W. A. (2014). Conjecture mapping: An approach to systematic educational design research. *The Journal of the Learning Sciences*, 23(1), 18-36.
- Sandoval, W. A., & Bell, P. (2004). Design-based research methods for studying learning in context: Introduction. *Educational Psychologist*, 39(4), 199-201.
- Supovitz, J. A. (2006). *The case for district-based reform: Leading, building, and sustaining school improvement*. Cambridge, MA: Harvard University Press.
- Talbert, J. E., & McLaughlin, M. W. (1993). *Reforming districts: How districts support school reform*. Stanford, CA: Center for the Study of Teaching and Policy.
- Trujillo, T. (2011). *The reincarnation of effective schools research: Rethinking the literature on district effectiveness*. Paper presented at the Thinking systemically: Improving districts under pressure, Rochester, NY.
- Tseng, V. (2012). *Partnerships: Shifting the dynamics between research and practice*. New York, NY: William T. Grant Foundation.
- Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of K-12 mathematics and science education in the United States*. Chapel Hill, NC: Horizon Research.
- Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M., Tomita, M. K., Young, D. B. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education*, 21(4), 335-359.