# Partnership for Building Capacity for Improvement in State Science Education

This proposal is to build foundational knowledge about the role that state agencies can play in supporting improvement to science education when partnered with researchers and each other. Building on 4 years of prior collaboration, the proposed effort involves establishing a formal partnership between the Council of State Science Supervisors and researchers at the University of Colorado Boulder and the University of Washington. State science supervisors from 13 states will form teams and a network that will develop and test state-level strategies for improving formative assessment as a policy instrument for aligning curriculum, instruction, and assessment. We aim to build knowledge and theory about the conditions under which a network of state teams can promote coherent guidance for instruction in local districts and schools. Using the language of the solicitation, we will accomplish this aim by engaging in "design-based iterative research on foundational knowledge for the implementation of new models of teaching and learning," where the new model is the "3-dimensional" model of science and engineering learning that integrates disciplinary core ideas, science and engineering practices, and crosscutting concepts (National Research Council, 2012).

The central problem we will investigate is developing knowledge about the conditions under which *coherent* state systems of science education can emerge in which all students have opportunities to meet challenging new standards. The incoherence of state and local science systems are well documented, as are the ways that incoherence makes system-level improvements difficult to attain (Cohen, Moffitt, & Goldin, 2007). Less well understood are the steps state leaders can take to help build coherent systems of science education around a common vision for science education where curriculum, instruction, and assessment align. Creating aligned policies is part of the solution, but the constraints on state leaders and evidence from efforts to craft coherence at the local level suggest additional strategies are necessary. Implementing 3-dimensional formative assessment through an iterative, collaborative research and development effort in relation to dimensions of state and local context will serve as the specific implementation strategy being tested.

Our research will involve the building of state-level teams and a network of state teams who will collaborate with researchers to diagnose current challenges to promoting coherence and develop knowledge about conditions for promoting coherence by testing and studying strategies for cultivating it. A first step will be to identify key system actors in each state and their responsibilities, for the purpose of assembling a participatory research team. A second step will be to engage this team in conducting a rapid assessment of needs by (a) mapping state-level system components and interrelations, (b) surfacing mixed messages, contradictions, and dilemmas faced by local actors related to implementing the vision of 3-dimensional science learning, and (c) identifying key learning needs of educational leaders and teachers. A third step will be to engage in an iterative process of design-based research that focuses on creating contexts for investigating how improving formative assessment can help align curriculum, instruction, and assessment in both states and districts around a vision of 3-dimensional learning. By attempting to create conditions that directly address implementation dilemmas and learning needs of local actors in the system through design, the project can move beyond accounts that document incoherence and begin to identify supportive processes and practices for promoting it.

The **intellectual merit** of this project is that it will build foundational knowledge related to how a network of states can support implementation of ambitious new standards. The proposed activities will develop, apply, and refine concepts from learning sciences, information sciences, and institutional theory to the study of implementation of new visions for teaching and learning. The partnership of researchers and state leaders that will lead the effort and its external advisors bring a unique blend of expertise in organizing effective collaborative design processes and conducting rigorous implementation research, in formative assessment, and in educational leadership practice. The partnership also brings significant experience of working together in focused efforts to marshal resources to address current problems of practice faced by state leaders.

The **broader impact** of the project is that developing knowledge of conditions that support equitable implementation of 3-dimensional science learning will benefit society by developing in all students' sufficient knowledge of science and engineering to engage in public discussions on related issues. Since inequity of opportunity to learn is a chief obstacle to broadening participation of underrepresented groups in STEM fields (NRC, 2012), this knowledge can contribute to the design of solutions to expand diversity in these fields. We have a national-level, consensus vision for K-12 science education (NRC, 2012), although implementation efforts have historically been staged within specific states, regions, districts, and schools. This project seeks to explore how to support broad-scale implementation of central aspects of this vision by developing tools and routines collaboratively in close partnership with many states from across the country. The direct involvement of state teams comprised of leaders with authority to make system-level changes increases the potential for the project to have direct and immediate impacts on practice. The resources and leadership of CSSS as the field's professional organization for state leaders ensures that the knowledge gained will continue to live and grow beyond the life of the project.

## Rationale for the Project

### Theoretical Background

A major obstacle to improving science education is the *incoherence* of guidance teachers receive regarding curriculum, instruction, assessment, and professional development. The messages teachers receive about what students should know and be able to do and how to teach from educational systems are often in conflict or competition. When leaders and teachers in schools and districts try to make sense of these conflicting messages, they often develop ideas that conflict with the intent of a given message (Coburn, 2001). For example, in a study of districts' responses to state-level reforms in mathematics and science in the 1990s, Spillane (2004) found that as new policies were communicated across departments and to schools, even when actors embraced those policies, the messages were distorted as people sought to relate them to past practice and other messages in their environment.

For nearly two decades, policy and science education researchers have posited that a key condition for improving educational systems is to increase the coherence among key elements in the system and across levels of the system (Fuhrman, 1993; Gitomer & Duschl, 2007; Linn, Kali, Davis, & Horwitz, 2008; National Research Council, 2006). This premise is reflected in the *Framework for K-12 Science Education's* chapter on implementation:

> Successful implementation requires that all of the components across the levels cohere or work together in a harmonious or logical way to support the new vision. This kind of system-wide coherence is difficult to achieve, yet it is essential to the success of standards-based science education. (NRC, 2012, p. 245).

The *Framework* also names different forms of coherence that are important in systems, two of which are relevant to state leaders' roles in their respective systems: *vertical coherence,* that is, the degree to which there is a shared understanding of the goals for science learning and consensus about the purposes and uses of assessment, and *horizontal coherence,* that is, the alignment of curriculum, instruction, and assessment policies and practices with standards.

Coherence has been a central focus of knowledge and theory building in education research. Policy research has problematized the notion that simply aligning policies regarding curriculum, instruction, and assessment with standards is sufficient to bring about coherence (Honig & Hatch, 2004; Knapp, Bamburg, Ferguson, & Hill, 1998). Traditions of local control complicate state level attempts to bring policies into alignment. Top-down efforts are also challenged by local actors' varied responses to new external guidance regarding curriculum, instruction, and assessment. Differences in individuals' past experience, their instructional philosophies, and collegial norms can actually lead to divergent instructional practices, rather than coherence (Coburn, 2004; Penuel, Frank, Sun, Kim, & Singleton, 2013).

Institutional theory (March & Olsen, 1984; Meyer & Rowan, 1977; Powell & DiMaggio, 1991; Scott, 2008) focuses on the coupling of policy and practice, specifically on the mechanisms that both bring them into alignment and that separate them. Two lines of inquiry framed by institutional theory offer a promising alternative to inform efforts to improve the coherence of educational systems. The first focuses on districts' efforts to design infrastructures for instructional improvement, and the second focuses on schools' efforts to create contexts for productive sensemaking to achieve a shared vision for improvement. Our research will apply, test, and refine these concepts to build knowledge of how states and networks of states can craft vertical and horizontal coherence in systems.

**Redesigning infrastructures for tailored instructional improvement.** Though the sources of different messages teachers receive about how to organize teaching and learning are varied, schools and districts filter these messages through what might be called an "instructional guidance infrastructure." This term refers to the structures and resources that school districts use to guide school leaders' and teachers' efforts to organize instruction (Hopkins, Spillane, Jakopovic, & Heaton, 2013). What makes up an instructional guidance infrastructure include governance structures (e.g., who decides what goes on district tests), collaborative work structures (e.g., curriculum writing teams), and leadership practices (e.g., coaching). The resources include curriculum guides and materials, assessments, professional development programs, and frameworks for teacher quality.

Infrastructure redesign centers on the strategies that districts use to figure out what combination of reforms are needed to promote specific reform goals and what structures must be adjusted to support implementation of those reforms (Spillane & Coldren, 2010; Spillane, Parise, & Sherer, 2011). Focusing on infrastructure redesign represents a big shift for research and development. Most efforts to develop tools and resources focus on developing resources to improve a single component of an educational system (e.g., designing new curriculum materials or a professional development program), not on how to design mechanisms to coordinate and bring into alignment different components to support a vision for teaching and learning within particular sites of implementation. As a consequence, theory that can guide the re-design of infrastructures in education lags behind the kinds of theories that can guide the design of interventions to improve teaching and learning.

As an initial set of principles, we draw on ideas from the information sciences about the design of information infrastructures, where such efforts are sometimes referred to as "infrastructuring" (Le Dantec & DiSalvo, 2013; Pipek & Wulf, 2009; Star & Ruhleder, 1996). A first principle is that it is necessary to engage with the existing base of infrastructure (Star, 2010). In education, that means identifying current governance structures, mapping system components and their interrelations, and surfacing contradictions and dilemmas faced by local actors (Penuel & Spillane, 2014; Sabelli, Penuel, & Cheng, 2015). A second principle is that change must be modular and incremental; it is not possible to change an infrastructure all at once (Star, 2010). Concretely, that means choosing key *leverage points* within a complex system where a focused intervention can have effects that are not limited to a single component of the system (Maroulis et al., 2010). A third principle is that infrastructuring must be participatory, because robust infrastructures require actors from different parts and levels of systems to use them for different purposes (Le Dantec & DiSalvo, 2013; Pipek & Wulf, 2009). Infrastructures change, too, as users engage with them, and considering users as co-designers codifies users' role in thoughtful redesign (Hanseth & Lundberg, 2001).

> *Our project aims to apply and refine these initial principles for supporting local, tailored redesign of instructional guidance infrastructures and extending the current knowledge base on school and district infrastructures to the study of state-level infrastructures. We will do so by engaging in a process of "design-based iterative research on foundational knowledge for the implementation of new models of teaching and learning" (EHR CORE Research Solicitation), which we detail in our research plan.*

**Designing contexts for productive sensemaking.** The concept of sensemaking in institutional theory describes the ways that actors "structure the unknown" (Waterman, 1990, p. 41) within organizational settings such as schools (e.g., Coburn, 2001). Actors engage in sensemaking to resolve ambiguity and manage uncertainty within their environment and make retrospective, as well as prospective sense of change (Weick, 1995). In science education research, teachers' sensemaking activities related to the coherence of messages they receive in professional development with messages from other components of the instructional guidance infrastructure has been shown to predict implementation of programs and curriculum materials (Penuel, Fishman, Gallagher, Korbak, & Lopez-Prado, 2009). In addition, research has documented the roles coaches and colleagues can play in supporting *productive* sensemaking, that is, sensemaking in which teachers work through implement challenges they face in their local environment and go on to make changes to their practice (Allen & Penuel, 2015; Furtak & Heredia, in press; Heredia, 2015).

Of particular relevance to the role of sensemaking in promoting coherence is Honig and Hatch's (2004) synthetic review of strategies that local leaders use to "craft coherence." They begin their review by contrasting their own approach—which begins with the premise that creating a coherent system of instructional guidance for teachers is an active, ongoing process of groups of leaders and teachers—with the notion that coherence is achieved simply by aligning policies with one another. They then go on to identify from the literature (which we have updated to reflect current research) three strategies by which schools can craft and maintain coherent instructional guidance for teachers:

1. *Create collective decision-making structures:* Developing local leadership teams and professional learning communities of teachers who share authority and responsibility for aligning curriculum, instruction, and assessment (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010; Coburn, Russell, Kaufman, & Stein, 2012; Louis, Marks, & Kruse, 1996; Malen, Ogawa, & Krantz, 1990; McLaughlin & Talbert, 2001; Westheimer, 1998).
2. *Maintain collective decision-making structures:* Provide new and veteran teachers with support tailored to their learning needs, through mechanisms such as coaching (Coburn & Woulfin, 2012; Elmore & Burney, 1997; Newmann, King, & Youngs, 2000; Youngs & King, 2002).
3. *Actively manage information:* Formalize goals to help filter what information is important and collect and use data to assess current practice, supported by external organizations as necessary (Datnow & Stringfield, 2000; Honig, 2004; Honig & Venkateswaran, 2012).

Notably, Honig and Hatch's principles are derived from studies of school and district practice, not states. Though we conjecture that these principles are good starting points for designing contexts for productive sensemaking within state-level endeavors, there are likely to be ways that the principles need to be refined and new principles identified. Activities facilitated by states are likely to include a range of education stakeholders, including leaders from different local educational agencies that each have their own instructional guidance infrastructures. Participants will differ widely with respect to familiarity with the *Framework* and the *Next Generation Science Standards*. In states that have not adopted the new standards, there is likely to be less familiarity than in states where adoption has already taken place. In addition to being a theoretical contribution about educational implementation, we assert that elaborating these principles in collaboration with a network of state leaders is a critical task, both because it is an identified need of the organization, and because of states' responsibilities for establishing consistent policies and allocating resources that support a common vision of science teaching and learning.

> *A core focus of our design efforts will be to design, test, and iteratively refine a set of learning opportunities to be implemented by state teams with different groups that use formative assessment as an entry point for learning about and planning to implement the vision of 3-dimensional science learning in the* Framework. *These learning opportunities may take the form of a workshop or webinar series, a collaborative design process, or action research projects, depending upon the state team. Researchers will study these efforts using a set of common outcome measures and survey protocols for assessing the professional learning opportunities.*

**Building a community of practice across state teams.** Though state science leaders collaborate with other leaders in state education agencies and with leaders and teachers in districts within their states, they have limited opportunities to learn from other state leaders and their efforts to promote improvement. In science, CSSS as an organization is the primary resource for state science supervisors wishing to connect with others who have similar roles in their state. Yet in general, CSSS does not yet constitute a *community of practice,* because the opportunities for interaction are so time limited. As we define it here, a community of practice is a group of people who are share a concern with improvement in a particular domain of practice and who interact with one another regularly about improving practice (Wenger, 2000).

There are a number of efforts underway today through the Regional Education Laboratories (RELs) to support communities of practice across states that our effort can build upon. One example is a multi-state partnership facilitated by the Midwest REL that is focused on identifying and replicating strategies that schools with high percentages of struggling students use to “beat the odds,” that is, to outperform schools with similar students on standardized tests. The partnership includes multiple state teams, each of which has different testing systems and circumstances. However, states have joined together both because of a shared commitment to the common strategy of identifying effective strategies in beat-the-odds schools, but also because of shared concerns about the implications of their methodological choices about identifying schools. This relatively new partnership has already yielded both original research findings related to those methodological choices (Abe et al., 2014) and a number of insights about how such partnerships can build a community of practice focused on continuous improvement (Gerdeman & Scholz, 2015). At present, there are not similar kinds of partnerships focused specifically on improvement in science education and with such a broad geographic scope. Our project aims to develop knowledge about whether and how a community of practice can support teams’ collaborative design of activities and learning from one another within a community of practice.

> *We will link state teams together and support their regular engagement with one another, with the aim of forming a community of practice across multiple states where states can share ideas and resources, explore strategies for addressing challenges associated with creating coherent systems of science education, and define potential areas of future collaboration. Researchers will study the team’s efforts to gain insight into the specific problems that they address together and the nature of support they provide.*

## Project Approach

In this project, we will investigate three research questions:

1. How do states compare with respect to the coherence of their instructional guidance infrastructures for supporting the vision of science learning for all presented in the *Framework for K-12 Science Education*?
2. What structures, resources, and practices support the work of crafting vertically and horizontally coherent science education systems in states?
3. How can multiple state teams partnering with researchers provide mutual support, expertise, and resources for supporting improvement efforts across those states and others?

Our approach to addressing these questions is to organize our research around three broad strategies:

- Partnering with state science supervisors committed to implementing the vision of the *Framework for K-12 Science Education;*
- Using formative assessment as a focal point for redesigning infrastructure since it can promote coherence across curriculum, instruction, and assessment; and
- Using a Design-Based Implementation Research (DBIR; Penuel, Fishman, Cheng, & Sabelli, 2011) approach to developing foundational knowledge about the conditions that support crafting coherent state-level systems in science education

### Partnering with State Science Supervisors Focused on *Framework* Implementation

The *Framework for K-12 Science Education* (National Research Council, 2012) calls for significant changes to science education to ensure that *all* students become proficient in science. It calls for systems to be organized around building understanding of disciplinary core ideas over time, engagement of students in the practices of science and engineering, and application of crosscutting concepts that unify science. It is based on a large body of research on how students best learn science (National Research Council, 1999, 2005, 2007, 2009) and on careful observations of the real work of scientists and engineers (e.g., Knorr-Cetina, 1999; Pickering, 1995). In the past two years, a number of states have chosen to adopt the *Next Generation Science Standards* (NGSS Lead States, 2013) or standards based on the *Framework.* Additional changes to curriculum, instruction, teacher preparation and professional development, and student assessment will be required.

State education agencies will need to play key roles in supporting implementation of this new vision. State agencies are responsible for developing and administering policies regarding standards, curriculum frameworks, materials selection and adoption, teacher professional development and adoption, and assessments. In carrying out these responsibilities, state science supervisors work closely with other staff in state agencies and contractors who assist them in carrying out these responsibilities. State supervisors routinely seek input from and provide guidance to districts on matters related to standards, curriculum, assessment, instruction, and professional development. Periodically, they develop and issue requests for proposals and guidelines that may apply to education providers in their states.

The lead professional organization for science leaders in the states, the Council of State Science Supervisors (CSSS), has been supporting its members in planning for *Framework* implementation since fall 2011. Through its Building Capacity for State Science Education (BCSSE) initiative, more than 40 states have sent representatives, including their state science supervisors, to gain fluency with the *Framework,* to create communication tools to disseminate key messages from the *Framework* to multiple state-based audiences, and to develop state-based strategic plans for implementation. As part of this effort, CSSS has enlisted a number of science education researchers as advisors to BCSSE, as presenters at meetings, and as members on working committees established to identify resources states could use to develop more concrete plans for implementation. The two research PIs have been part of these efforts.

While BCSSE has been successful in developing state science supervisors' understanding of the *Framework,* much remains to be done to develop local capacity to help teachers and administrators bring about the changes required to science education. Educators will need to develop understanding within the constraints of limited time for professional development and that work with curriculum materials that are not yet aligned with the *Framework's* vision. State leaders will need to develop plans that address the long-term professional learning needs of teachers and administrators (National Research Council, 2015).

> *As a first step in this research project, we will identify key ways that roles and responsibilities of state science supervisors vary across 13 purposefully selected states, as well as the current strategies states are using to promote implementation of the vision of the Framework.*
>
> *On the basis of what we learn, we will then build state-level teams and prepare them to conduct a rapid assessment of the learning needs of teachers and administrators in their states. Simultaneously, researchers will co-design and field a survey within six states focused on the particular components of leaders and teachers' local instructional guidance infrastructures that influence what happens in classrooms. The information we gather from both sources will be used to create a map of the key challenges to implementing the Framework and identify LEAs and existing professional communities where there is strong will and capacity for change.*

## Focusing on Formative Assessment as a Focal Point

Our proposed effort begins with the premise that *formative assessment*[1] is an ideal leverage point for states to bring about vertical and horizontal coherence to their science education systems. Formative assessment is a key strategy for supporting ambitious disciplinary learning (Black, 1993; White & Frederiksen, 1998). As we argue below, it provides a focus for infrastructure design that requires articulation among curriculum, instruction, and assessment and is within the authority of state science

[1] For purposes of this project, we (the research team and states) employ the CCCSO definition of formative assessment as "a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievements of intended instructional outcomes" (Formative Assessment for Students and Teachers (FAST) SCASS, 2012, p. 4).

leaders to influence. Further, conversations about formative assessment practices and student work provide a means for creating a shared vision about what three-dimensional learning looks like in science classrooms, a key condition for vertical coherence.

Schools, districts, and states are pursuing a number of initiatives to improve formative assessment in classrooms. This presents both an opportunity and challenge: the opportunity is that there is energy and will to improve the quality of the process, and the challenge is that there are many different definitions and approaches being tried even within individual schools. Drawing on a recent review of formative assessment interventions (Penuel & Shepard, in press), we propose to focus our efforts on approaches that (1) are grounded in sound theories of how students learn science, and that (2) show how assessment practices must be linked to curriculum and instruction. We view these approaches as good starting points for improving horizontal coherence, because formative assessment interventions of this type have the potential to link different communities of practice across levels of the system and focus their attention on how curriculum, instruction, and assessment must be integrated.

Examples of the kinds of approaches to formative assessment that will inform our efforts are highlighted in Chapter 4 of the recent NRC report, *Developing Assessments for the Next Generation Science Standards* (2014) and include Lehrer and Schauble's (2012) efforts to embed learning progressions-based assessments into coherent instructional sequences focused on modeling of natural systems and Songer and colleagues' (Gotwals, Songer, & Bullard, 2012) curriculum-based assessments of students' explanations of ecosystems. The report also highlights innovative approaches to identifying relevant interests and experiences of youth that can inform instruction, which are important strategies for helping students from underrepresented groups identify with science (e.g., Tzou & Bell, 2010).

Of course, simply focusing on formative assessment and making good examples available to teachers and educational leaders throughout a state is not sufficient to promote coherence. To promote vertical coherence (coherence around a shared vision or model for teaching and learning), it is important to structure opportunities to analyze assessment tasks and associated student work (National Research Council, 2014). Specifically, conversations should be structured to look for evidence of how tasks elicit and student work evidences "three-dimensional" science learning that integrates disciplinary core ideas, science and engineering practices, and crosscutting concepts. Research on professional development shows that examining instructional practices and student work in tandem can support improvements to teaching and learning (Roth et al., 2011).

> *State teams will collaboratively design professional learning opportunities that incorporate opportunities for teachers and leaders to examine assessment tasks and student work together to develop their understanding of a vision of 3-dimensional learning. The workshops will also include opportunities for people to identify opportunities in their organizations (e.g., schools and districts) to adapt the activities to help others learn about the vision. The aim will be to develop, implement, and study productive formats for sensemaking to develop a shared vision of learning across levels of the system. We will implement and evaluate these in different contexts identified during the first phase of research.*

## Using Design-Based Implementation Research to Develop Knowledge and Theory

The CORE Solicitation is focused on developing both foundational knowledge and theory; it explicitly states that teams can engage in "design-based iterative research on foundational knowledge for the implementation of new models of teaching and learning." This kind of research is appropriate whenever it is necessary to *engineer* or *organize* the phenomenon that one wants to study, that is, when it does not exist under routine conditions of practice (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; Sandoval & Bell, 2004). As noted above, incoherence among key elements of the instructional guidance infrastructure is common; coherence is desired. Therefore, we propose to use a design-based approach to organizing state level processes that (1) reach into districts and local teacher communities of practice and (2) are aimed at creating coherence among curriculum, instruction, and assessment. We will use variation in both enactment and outcomes to help us identify the conditions under which these processes lead to a shared vision of 3-dimensional learning across levels of the state system.

The particular approach we will employ to design-based iterative research is one in which both research PIs are involved in helping develop: Design-Based Implementation Research (DBIR; Penuel et al., 2011). We emphasize that a core principle of DBIR is to develop **knowledge** and **theory** related to both classroom learning and implementation through systematic inquiry. Though some interpret DBIR as a methodology for scaling proven innovations, this is incorrect. DBIR can and will be used in this project to develop foundational knowledge about implementing the vision for learning presented in the *Framework.* It will do so by engaging in implementation research, which systematically documents evidence of implementation and explains patterns of implementation in a way that is both informed by theory and develops it (Spillane, Reiser, & Reimer, 2002). As noted above, we draw on theories from information sciences (infrastructuring) and sociology (institutional theory) to inform our design and adaptation of measures and analysis of implementation evidence, which will focus on documenting how and when team activities support vertical and horizontal coherence.

As other DBIR projects do, we will engage in collaborative, iterative design of strategies we plan to test. The efforts to engineer coherence we will test will be collaboratively designed by state teams and researchers, to ensure that they address the needs identified in our research of local science leaders and teachers. Where possible, we will design workshops and resources that can be similarly or variably used across different states, not just a single state, so that we can iteratively refine our designs across successive implementation trials. With each trial, we will document enactments and collect data from participants about their understanding of the vision for 3-dimensional science learning in the *Framework.*

> *We will embody **the principles of DBIR** by jointly negotiating the specific challenges related to coherence that states wish to address using evidence from a needs assessment (**Principle 1**), collaboratively designing experiences for local actors to engage with the vision of the* Framework *through encounters with formative assessment tasks (**Principle 2**), study implementation and whether it develops a common vision for science learning anchored in the* Framework *(**Principle 3**), and build state teams' capacity to engage in systematic, rapid needs assessment that can be used in future endeavors (**Principle 4**).*

## Research Plan

Our research plan is divided into a sequence of three major tasks, as depicted in the table below.

Table 1. Research Plan: Sequence of Major Tasks

| *Task* | *Subtasks* | *Rationale for Task* | *Timeframe* |
|---|---|---|---|
| **1. Build and Convene State Networks of Teams** | 1.1. Determine team membership<br>1.2. Convene network of teams | Purposeful composition of teams is needed to ensure actors with authority, resources are included | *Year 1:*<br>Summer 2016 (1.1, 1.2) |
| **2. Develop Descriptions of Infrastructures and Learning Needs of Teachers and Leaders** | 2.1. Design protocols for data collection.<br>2.2. Carry out data collection (survey and focus groups)<br>2.3. Analyze rapid assessment data | Identify key dilemmas and contradictions related to incoherence<br>Build local capacity for systematic, rapid needs assessment | *Years 1-2*<br>Fall 2016 (2.1)<br>Fall-Winter 2017 (2.2)<br>Spring 2017 (2.2) |
| **3. DBIR Focused on Formative Assessment** | 3.1. Codesign of learning activities.<br>3.2. Measuring changes to vertical coherence within states<br>3.3 Supporting and assessing horizontal coherence | Develop theory and knowledge related to conditions that foster coherence | *Years 2-3*<br>Summer 2017 (3.1, Initial design)<br>Summer 2018 (3.1, Redesign)<br>Fall 2017-Spring 2019 (3.2, 3.3) |

## Task 1: Build and Convene Network of State Teams

A total of 13 states have agreed to build design, implementation, and study teams for this project. We purposefully selected these 13 states to meet different criteria. The criteria we used were size and urbanicity and adoption of the Next Generation Science Standards. We sought to include small states with no large urban center that had and had not adopted NGSS, states with one large urban center (metro area > 800,000) that had and had not adopted NGSS, and states with more than one large urban center that had and had not adopted NGSS (Table 2). We reasoned that these represented sufficiently varied policy contexts so as to warrant the claim that implementation knowledge developed from study of these 13 states would be generalizable to a broad range of other states.

Table 2. States Agreeing to Build Design, Implementation, and Study Teams

| *NGSS Adopted States* | *Non-NGSS States* |
|---|---|
| Delaware | Arkansas |
| Kansas | Michigan |
| Kentucky | Oklahoma |
| New Jersey | Pennsylvania |
| Oregon | South Dakota |
| Washington | Utah |
| West Virginia | |

**Task 1.1. Determine team membership.** Prior to composing these teams, the leaders of the state teams will conduct policy analyses and develop preliminary profiles of the instructional guidance infrastructure in each state and network of ties between researchers and science leaders at the state level, districts, and intermediary organizations. The profiles will identify key people and workgroups responsible for creating and maintaining policies and resources related to curriculum, instruction, and assessment in science. Within each state, a team of 5-6 the most influential people identified in the network will be selected to be part of the team.

**Task 1.2. Convene network of teams.** At our first convening of the state teams, which will be held in Boulder, Colorado, at the University of Colorado, we will provide teams with an overview of the research team, engage them in collaborative design of survey instruments designed to identify challenges to implementing the vision of the *Framework* in their states and leaders and teachers' perceived learning needs, and prepare them to engage in a rapid assessment of needs (see Task 2 below) in their own states using focus groups and interviews.

**Task 2: Develop Descriptions of State-Level Infrastructures and Learning Needs of Science Teachers and Leaders (RQ1)**

To develop an understanding of the current level of *incoherence* in the science education and learning needs of teachers and leaders in states, we will adapt a methodology from public health for participatory research called a rapid assessment (Trotter, Needle, Goosby, Bates, & Singer, 2001). Rapid assessment includes a suite of methods designed for quick data collection and analysis, the involvement of relevant experts and community members, and triangulation across data sources. The design elements of rapid assessment ensure that researchers help local teams comprised of practitioners develop expertise to complete observations and focus groups and that the ethnographic work is accompanied by analysis of existing data sets. We will employ interviews, local focus groups, and a survey conducted in each state as part of this process.

**Task 2.1. Design protocols for data collection.** There will be three protocols that the network will co-design: key informant interviews, local focus groups, and a survey. The design process will be facilitated by researchers from the University of Colorado and the University of Washington, using a construct-centered approach to instrument design. As such, the design of each instrument will begin with an elaboration of the foci identified below for each instrument, before beginning the process of identifying questions (in the case of interviews and focus group protocols) and items (in the case of the survey). Researchers will collaborate with 5-6 team leaders to pilot test each protocol and make refinements to the protocols on the basis of their pilot tests before they are broadly used.

*Key informant interviews.* Interviews will target district and school level leaders in science in six of the 13 states. Key informant interviews will be focused on the topic of identifying challenges, dilemmas, and "double binds" faced by local leaders and teachers that are the result of incoherent instructional guidance (e.g., how best to coordinate the teaching of skills outlined in Common Core State Standards and the practices in the *Framework*, how to coordinate teacher evaluation with the uptake of strategies to promote student argumentation). They will also focus on leaders' perceptions of their own learning needs and those of teachers with respect to implementing the vision of teaching and learning in the *Framework.* A third focus will be on the areas of the *Framework* where leaders are giving emphasis, in their own instructional leadership in science. This information will be useful especially in identifying where there is political will at the local effort to direct improvement efforts.

*Local focus groups.* The purpose of local focus groups is two-fold; first, to develop relationships between members of the multi-state network through in-person training on rapid ethnographic methods: conducting focus groups, and second, to have local experts gathering data to support the needs assessment work. Focus groups present an opportunity to engage local science teachers and leaders in wrestling with competing priorities for focusing efforts to implement the *Framework.* They also present an opportunity to identify differences in teachers' and local leaders' responses to incoherence, some of which may lead to productive engagement with the vision of the *Framework.* By focusing on these responses, we can develop initial conjectures about strategies that can be integrated into learning experiences developed as part of task 3.

*Teacher Survey.* The survey will be targeted to teachers in states, and it will focus on three main topics: (1) the sources that teachers rely upon for guidance regarding how they teach science, (2) their own perceived needs regarding the *Framework*, and (3) professional development they have received related to the *Framework* and/or NGSS. This instrument will undergo a more in-depth piloting process by CU Boulder and UW researchers prior to implementation that will involve cognitive interviews and an analysis of scales using classical test theory.

**Task 2.2. Carry out data collection.** Data collection will take place in year 1. Researchers will carry out key informant interviews with 10 people from each state. State teams will identify prospective informants, but researchers will approach informants about the study to obtain their consent to participate. All interviews will be transcribed immediately with personal identifiers removed.

Each state team will be responsible for conducting and developing audio recordings from four different focus groups. As part of their training and preparation, all team members conducting focus groups will receive appropriate training in protecting the rights of human participants in research. Researchers will develop transcripts from focus groups, stripping all personal identifiers from transcripts.

Researchers will use an online system to administer the survey to a sample of 150 teachers from randomly selected districts in each of the 13 states. Researchers will purchase an email list from identified districts for purpose of selecting teachers independent of state teams (i.e., more representative of the teaching force). We will send an invitation to participate in the survey to this email, but no identifying information will be collected from them. Researchers will conduct email follow up until a 50 percent response rate or higher is achieved in each state.

**Task 2.3. Analyze rapid assessment data.** Analysis will proceed in two different ways, on a rapid timescale, to inform the process of design-based research of state teams and on a slower timescale, to triangulate among and assess the validity of different data sources and develop analyses that are publishable in social science journals (see Cobb, Jackson, Smith, Sorum, & Henrick, 2013, for an example of this strategy). For the rapid analysis, researchers will develop summaries of key themes from interviews, focus groups, and the surveys and bring them to a meeting of the network. Summaries will be developed for each state and for the network as a whole. We will structure a discussion of the data in which states work in teams and in cross-team groups to develop hypotheses that could account for the pattern of results and conjectures about structures, resources, and practices support the work of crafting vertically and horizontally coherent science education systems (relates to Research Question 2).

Over the course of year 2, we will develop richer, theoretically grounded descriptions of states' instructional guidance infrastructures. We will adopt a multiple-case study approach to developing descriptions of states' current instructional guidance infrastructure as part of the rapid assessment of needs. In a multiple-case study approach, researchers often draw on multiple sources of data to construct a coherent, interpretive account of cases and how they differ from one another along a common set of dimensions (Yin, 2013). In our study, primary sources of data will be focus group and interview data from state teams, survey data we field, and observations of meetings of the state network we will conduct. So that the state infrastructures can be compared, we will work backward from a case description template to be developed for each state, which is grounded in earlier empirical accounts of infrastructures cited in the first part of the proposal. We can map back data sources from the template to be integrated, which informs decisions about what types of questions are be needed for each data source. We will then develop sample interview and survey questions as models for state teams to use or adapt for their own rapid needs assessment. We do anticipate needing to revise the template, on the basis of what we learn, to reflect elements of state-level infrastructures not adequately captured in contemporary accounts of districts and schools.

### Task 3: Design-Based Implementation Research Focused on Formative Assessment

A primary conjecture is that formative assessment can serve as a leverage point for increasing vertical and horizontal coherence. To test this conjecture, the network of state teams will design and test the effects of strategies foe promoting a shared understanding of the Framework's vision that are anchored in strategies for formative assessment.

**Task 3.1. Co-design of learning activities by state network.** State teams and the network will need to agree on common elements of the vision from the *Framework* to emphasize and to devise learning experiences to implement locally that substantively engage participants with the vision. By "substantive engagement" we mean activities that are sustained over a period of time, focused on a specific science domain, and that guide their interpretation of the vision of the *Framework* in the context of collaborative activity While each state will develop its own sequence of learning opportunities collaboratively with the state team, they will do so in consultation with researchers on the project and in accordance with a set of research-based design principles elaborated by the network of states. The researchers on the project bring extensive experience in developing and testing formative assessment strategies in science, and we view our involvement and support as partners as critical to the success of the teams.

Two core principles we have established ahead of time is that learning experiences should directly expose participants to 3-dimensional multicomponent assessment tasks and to opportunities to analyze student work associated with relevant tasks. Both are principles included in tested models of professional development related to formative assessment (Heller, Daehler, Wong, Shinohara, & Miratrix, 2012; Roth et al., 2011). Because formative assessment demands strategies for interpretation and use of assessment data, we anticipate learning activities will also address how learning progressions can serve as tools for interpreting student thinking (Furtak, Thompson, Braaten, & Windschitl, 2012) and for highlighting alternative pathways to teaching the same content (Penuel, Beauvineau, DeBarger, Moorthy, & Allison, 2012).

In both Washington and Colorado, where members of the research teams can easily visit face-to-face learning activities, we will also conduct observations of a small set (6-10) of multi-event learning sequences. We will select these in collaboration with state teams, selecting those activities that seem particularly promising because of the organization of the activities and commitment of participants. We will develop field notes from these observations, with the aim of documenting different ways that participants engage with the activities. We will also focus on moments when participants give voice to concerns, dilemmas, or challenges related to system incoherence and facilitators' response to participants. These data will provide an additional source of data in this respect to address Research Question 1.

**Task 3.2. Measuring changes in vertical coherence within states.** One of the ways we plan to investigate the success of efforts to use formative assessment as a focal point for state-led learning activities will be to assess participants' vision for science learning via a survey-based measure that we develop. We draw on Hammerness' (2001) definition of vision as a "set of images of ideal classroom practice" (p. 143), and we will focus attention on what their vision of what students should do and learn and on their role in classroom assessment. We focus on vision, because the *Framework* defines vertical coherence as "a shared understanding at all levels of the system (classroom, school, school district, state, and national) of the goals for science education (and for the curriculum) that underlie the standards and (b) that there is a consensus about the purposes and uses of assessment" (National Research Council, 2012, pp. 245-246).

We plan to use a construct-modeling approach (Wilson, 2005) to developing this measure. A construct modeling approach starts with the development of a construct map that defines different dimensions of the construct and levels of depth/sophistication of the construct, then items are created to correspond with the construct map, and finally, pilot survey data is used to model response data using item response theory to understand the characteristics and performance of the items. After analyzing pilot data, the construct map may be revised and new items added or items revised so that the items to elicit specific information about this revised construct map. Table 3 shows the primary constructs for which we will develop survey items to evaluate the vertical coherence of participants' visions for science teaching and learning.

Table 3. Dimensions of Vertical Coherence To Be Measured

| **Dimension of Vision (Hammerness, 2001)** | **Dimension of Vertical Coherence (NRC, 2012)** |
|---|---|
| ***Focus:*** *primary ideas, clarity of vision about goals* | Goals for student learning: Integration of core ideas, practices, crosscutting concepts (three-dimensional learning) |
| | Goals for student learning: Developing student understanding over time (developmental coherence) |
| | Goals for student learning: All standards, all students (equity) |
| ***Range:*** *breadth of aspects of vision* | Goals for curriculum |
| | Purposes of assessment |
| | Uses of assessment |
| ***Distance:*** *discrepancy between one's vision and current practice* | Required shifts in curriculum to meet new goals for student learning |
| | Required shifts in instruction to meet new goals for student learning |
| | Required shifts in assessment to meet new goals for student learning |

Our aim is to design a survey instrument that takes no more than 15 minutes to complete, so that it can be easily integrated into learning events developed by state teams at the start of each sequence of activities they plan. We will also ask participants to complete the survey at the conclusion of a sequence of learning activities. At the conclusion of the sequence of learning activities, we will add items where participants will report back on the quality of their experience. To design this set of items, we will draw on both Desimone's (2009) framework for professional development to develop items that resemble those used in an earlier study by Penuel and colleagues (Penuel, McWilliams, et al., 2009). We will also pilot test a set of research-based standards for professional learning developed jointly by the research team and the CSSS professional learning committee.

The survey we use will be developed in an online survey development tool, Qualtrics, and all data will reside on a University of Colorado Boulder server. CU researchers will take responsibility for developing reports that aggregate data from different learning events and ensuring that it is linked to documentation of what took place at the event.

We propose a relatively straightforward approach to analyzing vertical coherence changes over time. Although cultural consensus methods are appropriate for measuring the degree of sharedness of attitudes across respondents, those methods presumes that a desired cultural stance is not known ahead of time (Weller, 2007). In our case, there is an ideal (the *Framework's* vision) against which to compare respondents' answers. Movement toward greater vertical coherence, then, should be indicated by an increase in mean scores within a state over time on a construct-centered measure of vision, as well as by a reduced standard deviation around the mean within the state (indicating sharedness). Depending on the number of different events within and across states, we can model change over time as a function of professional development quality using standard multi-level modeling techniques (e.g., HLM; Raudenbush & Bryk, 2002) in order to identify promising practices for promoting vertical coherence that could be tested in future research.

**Task 3.4: Supporting and assessing horizontal coherence at the local level.** Into all learning experiences we design together, we will integrate an activity in which participants work with others in their organizations to articulate shifts needed to make their local system more horizontally coherent with respect to guidance regarding formative assessment. The activity will focus participants' attention on elements of local instructional guidance infrastructure related to curriculum, instruction, and assessment. Participants will identify what shifts are needed in each component, identify persons or established work groups with authority to implement those shifts, and name next steps that one of the participants will take subsequent to the workshop.

Participants will develop their plans in a digital format, using a template we provide and that they agree to allow the research team for the project to analyze. We will develop a rubric for analyzing participants' plans, characterizing them with respect to *specificity* (the degree to which they make specific plans to make changes), *comprehensiveness* (degree to which they address multiple elements of the guidance infrastructure with special attention given to the promotion of educational equity), and *authority* (degree to which plans name specific people with authority to make changes).

Although an analysis of plans is not a direct measure of whether the designs promote horizontal coherence, we conjecture that specific, comprehensive plans with clear identification of persons responsible for implementing components is a necessary condition for horizontal coherence. We also expect to see wide variation in plans, and we can analyze whether this variation is linked to the nature and quality of activities in which they participated, state-level specifics, or personal histories. This analysis will help us surface "positive deviance" (Bryk, Gomez, Grunow, & LeMahieu, 2015) that can be investigated in future studies.

## Results of Prior Work

The proposed effort builds directly from the mutual engagement of PIs Penuel and Bell with CSSS as part of the Research+Practice Collaboratory (DRL- 1238253; http://researchandpractice.org). The Collaboratory is a multi-institution center whose mission is to explore new ways to relate research and practice in STEM education for the purpose of creating more equitable innovations for STEM teaching and learning. Led by the Exploratorium of San Francisco, both the University of Colorado Boulder and University of Washington (along with EDC, Inverness Associates, and SRI International) are key partners in the Collaboratory. Through their involvement in the Collaboratory, Co-PIs Penuel and Bell have partnered with CSSS to develop robust collaborations we call "critical engagements," which are focused on bringing research-based tools to professional associations like CSSS on a just-in-time basis. Through our critical engagement with one another, CSSS and the Collaboratory have embarked on work to support the committees established in spring 2014 as external members of those committees. As part of that work, researchers have identified research resources related to the issues discussed by committees, created research and practice briefs linked to those issues (see, e.g., http://stemteachingtools.org/brief/34), and co-designed documents like the Professional Learning Standards. The researchers have also helped CSSS leaders devise a strategy for integrating equity as a theme across committees. These efforts are described in an article presently in review at the *Journal of Educational Change* (Penuel, Bevan, Bell, Falk, & Buffington, under review). In addition, evaluation data collected by CSSS leaders on researchers' involvement, CSSS committee members highlighted the value of the research partners, "great discussion" among participants, and a sense that co-chairs utilized the expertise of participants. Participants also showed a commitment to wanting to make this collaboration stronger, with one person suggesting that future meetings include examples of research partnerships that work well. The proposed effort aims to address precisely these expressed needs.

## Project Leadership and Management

A team of researchers and practitioners at the three collaborating institutions will lead the project. This team will meet every two weeks by teleconference to plan and manage activities at the respective sites. Overall project management will be the responsibility of the University of Colorado Boulder team. Research activities will be jointly led by CU Boulder and UW researchers, while all state team activities will be led by the CSSS co-Principal Investigators, with researchers serving as co-designers of network-level activities.

The Principal Investigator for the project will be ***Dr. William Penuel,*** Professor of Learning Sciences and Human Development at the University of Colorado Boulder. Dr. Penuel brings expertise in formative assessment, implementation research, and design-based research in science education. He has extensive experience in cultivating, sustaining, and studying research-practice partnerships at the district level, and he has served as an advisor to the Building Capacity for State Science Education (BCSSE) initiative of CSSS. He was a member of the committee that developed the National Research Council report, *Assessment and the Next Generation Science Standards.* He will oversee all aspects of the research and development effort and coordinate communication with key stakeholders.

Co-Principal Investigator ***Dr. Philip Bell*** is Shauna C. Larson Chair in Learning Sciences and Executive Director, Institute for Science & Math Education at the University of Washington Seattle. Dr. Bell brings expertise in how people learn science across multiple structures of cultural practice using design-based and ethnographic methods. He has studied everyday expertise and cognition in science and health, the design and use of novel learning technologies in science classrooms, youth argumentation, culturally expansive science instruction, and scaled implementation of educational improvement. Bell served as a member of the Board on Science Education with the National Academy of Sciences for eight years and served on the committee of the NRC *Framework for K-12 Science Education* that was used to guide development of Next Generation Science Standards. Dr. Bell has served on the advisory board of the BCSSE initiative of CSSS for four years. Dr. Bell will assist Dr. Penuel with all aspects of the research activities for the project.

Co-Principal Investigator ***Matt Krehbiel*** is Science Program Consultant at the Kansas Department of Education and is currently the elected President of the Board of the Council of State Science Supervisors. Prior to joining the Department of Education, Mr. Krehbiel taught Physical Science, Environmental Ecology, Prairie Ecology, and Biology to 9th-12th graders at Junction City High School (KS), where he served as a science teacher and Science, Technology, and Engineering Academy Leader. He is a current member of the Board on Science Education and was a committee member for the NRC Report, *Guide to Implementing the Next Generation Science Standards.* Mr. Krehbiel, along with co-Principal Investigator Sam Shaw, will convene the network of state teams and lead activities to analyze needs data and design learning activities with them. He will also lead a state team in Kansas.

Co-Principal Investigator ***Sam Shaw*** is Team Leader in the Division of Learning and Instruction at the South Dakota Department of Education and is a member of the Board of the Council of State Science Supervisors. Prior to joining the Department of Education, Mr. Shaw was a middle school teacher at Georgia Morse Middle School (SD). Shaw, along with co-Principal Investigator Matt Krehbiel, will convene the network of state teams and lead activities to analyze needs data and design learning activities with them. He will also lead a state team in South Dakota.

**Advisory Board.** We have assembled and will convene an advisory board comprised of experts in both research and practice to provide external, objective feedback to the team in an ongoing fashion. The advisory board will meet once in person in the first year of the project and virtually in years 2 and 3. At each board meeting, we will provide them with a detailed project update and a set of focused questions to help us address that are related to key dilemmas our project is facing. Meeting 1 will take place after state teams have convened, and once we have preliminary needs assessment data. Meeting 2 will take place after teams have developed preliminary plans for learning activities related to formative assessment, and it will be an opportunity for advisors to provide feedback on state teams' plans. Meeting 3 will be held when the research team has developed preliminary analysis of data on vertical and horizontal coherence.

It will be organized to help refine the communication plan based on emerging study findings. The members of the advisory board bring extensive expertise in the areas of professional development, formative assessment in science, organizational perspectives on improvement, and state, district, and school level leadership in science. They are: *Dr. Erin Furtak*, Associate Professor of Curriculum and Instruction at the University of Colorado Boulder; *Dr. Michael Lach*, Director of STEM Education and Strategic Initiatives at CEMSE and the Urban Education Institute of the University of Chicago; *Dr. Stephen Pruitt* is Senior Vice President for Science at Achieve, Inc.; *Dr. Jennifer Russell,* Associate Professor of Learning Sciences and Policy in the School of Education and a research scientist at the Learning Research and Development Center (LRDC) at the University of Pittsburgh; *Tricia Shelton* is a veteran teacher at Boone County High School in Florence, Kentucky and co-developer of the teacher twitter community #NGSSchat.

## Communication Strategy

The research-practice partnership among the University of Colorado Boulder, University of Washington, and Council of State Science Supervisors brings significant commitment, resources, and networks of researchers and practitioners to the task of communicating findings and sharing tools from our project. Our overall strategy will be to use existing networks and professional associations to share about our activities and tools throughout the project, and as findings emerge, to develop manuscripts for peer reviewed publications targeting researchers and then practitioner journals.

The *CSSS Annual Conference* provides opportunities to engage a broader range of states in selected activities (e.g., informing design of protocols) and present findings to the entire network of state science supervisors. These include potential plenary presentations and workshops on focused topics, including findings from the research.

CU Boulder and the University of Washington are formal collaborators in the *Research+Practice Collaboratory,* an NSF-funded effort that is developing and testing new approaches to relating research and practice to promote equity in STEM education. The Collaboratory maintains an active website (http://researchandpractice.org) where activities of this partnership will be described and tools in development shared. The LearnDBIR web site (http://learndbir.org) targets primarily researchers, and it is a go-to resource center for scholars interested in tools and findings related to the Design-Based Implementation Research approach. We plan to post descriptions of the rapid assessment methodology and associated tools there.

The University of Washington's STEM Teaching Tools (http://stemteachingtools.org/) web site contains resources targeting practitioners and is focused on supporting implementation of new standards. The resources include both *research briefs* and *practice briefs* that distill research findings and practitioner insights on specific topics. The research brief on the NRC report, *Assessment and the Next Generation Science Standards,* has been accessed over 3000 times. We will develop a suite of STEM Teaching Tools during the design phase, specifically to support learning activities states develop.

We also have a strong presence on and participate in regular activities on social media. The regularly hosted #NGSSChat event on Twitter regularly draws dozens of practitioners during the dinner hour. The two research PIs are regular participants, where they will share resources when topics for the chat are most relevant.

We have also targeted several academic peer-review and practice venues for publishing research from the project. Academic journals we will target include *Journal for Research in Science Teaching, Science Education,* and the *Journal of Science Teacher Education.* Practice journals we will target include *Phi Delta Kappan,* which targets administrators; *Science Educator,* which is the journal of the association of district science leaders; and the National Science Teacher Association's three journals targeting elementary, middle, and high school teachers.

## References

Abe, Y., Weinstock, P., Chan, V., Meyers, C., Gerdeman, R. D., & Brandt, W. C. (2014). *How methodology decisions affect the variability of schools identified as beating the odds.* Washington, DC: U.S. Department of Education.

Allen, C. D., & Penuel, W. R. (2015). Studying teachers' sensemaking to analyze teachers' responses to professional development focused on new standards *Journal of Teacher Education, 66*(2), 136-149.

Black, P. (1993). Formative and summative assessment by teachers. *Studies in Science Education, 21*, 49-97.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard University Press.

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago: University of Chicago 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., 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. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis, 23*(2), 145-170.

Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education, 77*(3), 211-244.

Coburn, C. E., Russell, J. L., Kaufman, J. H., & Stein, M. K. (2012). Supporting sustainability: Teachers' advice networks and ambitious instructional reform. *American Journal of Education, 119*(1), 137-182.

Coburn, C. E., & Woulfin, S. L. (2012). Reading coaches and the relationship between policy and practice. *Reading Research Quarterly, 47*(1), 5-30.

Cohen, D. K., Moffitt, S. L., & Goldin, S. (2007). Policy and practice: The dilemma. *American Journal of Education, 113*(4), 515-548.

Datnow, A., & Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed at Risk (JESPAR), 5*(1, 2), 183-204.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181-199.

Elmore, R. F., & Burney, D. (1997). Investing in teacher learning: Staff development and instructional improvement in Community School District #2. New York: National Commission on Teaching and America's Future.

Formative Assessment for Students and Teachers (FAST) SCASS. (2012). Distinguishing formative assessment from other educational assessment labels. Washington, DC: Council of Chief State School Officers.

Fuhrman, S. H. (1993). *Designing coherent educational policy: Improving the system*. San Francisco, CA: Jossey-Bass.

Furtak, E. M., & Heredia, S. (in press). Exploring the influence of learning progressions in two teacher communities. *Journal of Research in Science Teaching, 51*(8), 982-1020.
Furtak, E. M., Thompson, J., Braaten, M., & Windschitl, M. (2012). Learning progressions to support ambitious teaching practices. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning progressions in science: Current challenges and future directions* (pp. 405-434). Rotterdam, the Netherlands: Sense Publishers.
Gerdeman, R. D., & Scholz, C. (2015). *Applying an adaptive positive deviance process in support of a teacher-driven continuous improvement community*. Paper presented at the Carnegie Foundation Summit on Improvement in Education, San Francisco, CA.
Gitomer, D. H., & Duschl, R. A. (2007). Establishing multilevel coherence in assessment. *Yearbook of the National Society for the Study of Education, 106*(1), 288-320.
Gotwals, A. W., Songer, N. B., & Bullard, L. (2012). Assessing students' progressing abilities to construct scientific explanations. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning progressions in science: Current challenges and future directions* (pp. 183-210). Rotterdam, the Netherlands: Sense Publishers.
Hammerness, K. (2001). Teacher's visions: The role of personal ideals in school reform. *Journal of Educational Change, 2*(2), 143-163.
Hanseth, O., & Lundberg, N. (2001). Designing work oriented infrastructures. *Computer Supported Cooperative Work, 10*(3-4), 347-372.
Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. *Journal of Research in Science Teaching, 49*(3), 333-362.
Heredia, S. (2015). *Dilemmas of reform: An exploration of science teachers' collective sensemaking of formative assessment practices.* (Ph.D. doctoral dissertation), University of Colorado Boulder, Boulder, CO.
Honig, M. I. (2004). The new middle management: Intermediary organizations in education policy implementation. *Educational Evaluation and Policy Analysis, 26*(1), 65-87.
Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher, 33*(8), 16-30.
Honig, M. I., & Venkateswaran, N. (2012). School-central office relationships in evidence use: Understanding evidence use as a systems problem. *American Journal of Education, 118*(2), 199-222.
Hopkins, M., Spillane, J. P., Jakopovic, P., & Heaton, R. M. (2013). Infrastructure redesign and instructional reform in mathematics. *Elementary School Journal, 114*(2), 200-224.
Knapp, M. S., Bamburg, J. D., Ferguson, M. C., & Hill, P. T. (1998). Converging reforms and the working lives of frontline professionals in schools. *Educational Policy, 12*(4), 397-418.
Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.
Le Dantec, C. A., & DiSalvo, C. (2013). Infrastructuring and the formation of publics in participatory design. *Social Studies of Science, 43*(2), 241-264.
Lehrer, R., & Schauble, L. (2012). Seeding evolutionary thinking by engaging children in modeling its foundations. *Science Education, 96*(4), 701-724.

Linn, M. C., Kali, Y., Davis, E. A., & Horwitz, P. (2008). Policies to promote coherence. In Y. Kali, M. C. Linn, & J. E. Roseman (Eds.), *Designing coherent science education* (Vol. 201-210). New York: Teachers College Press.
Louis, K. S., Marks, H. M., & Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal, 33*(4), 757-798.
Malen, B., Ogawa, R. T., & Krantz, K. (1990). What do we know about school based management? A case study of the literature: A call for research. In W. H. Clune & J. Witte (Eds.), *Choice and control in American education, volume 2: Decentralization and school restructuring* (pp. 289-342). Philadelphia, PA: Falmer Press.
March, J. G., & Olsen, J. P. (1984). The new institutionalism: Organizational factors in political life. *American Political Science Review, 78*(3), 734-749.
Maroulis, S. J., Guimerà, R., Petry, H., Stringer, M. J., Gomez, L. M., Amaral, L. A. N., & Wilensky, U. (2010). Complex systems view of educational policy research. *Science, 330*(6000), 38-39.
McLaughlin, M. W., & Talbert, J. E. (2001). *Professional communities and the work of high school teaching*. Chicago, IL: University of Chicago Press.
Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, *83*(2), 340-363.
National Research Council. (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
National Research Council. (2005). How students learn: History, mathematics, and science in the classroom. Washington, DC: National Academies Press.
National Research Council. (2006). *Systems for state science assessment*. Washington, DC: National Academies Press.
National Research Council. (2007). Taking science to school: Learning and teaching science in grades K-8. Washington, DC: National Academies Press.
National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits*. 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. (2014). Developing assessments for the Next Generation Science Standards. Washington, DC: National Academies Press.
National Research Council. (2015). Guide to implementing the Next Generation Science Standards. Washington, DC: National Academies Press.
Newmann, F. M., King, M. B., & Youngs, P. (2000). Professional development that addresses school capacity: Lessons from urban elementary schools. *American Journal of Education, 108*(4), 259-299.
NGSS Lead States. (2013). Next Generation Science Standards: For states, by states. Washington, DC: National Academies 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., Bevan, B., Bell, P., Falk, J., & Buffington, P. (under review). Using social networks to enhance research use in planning for and supporting educational change. *Journal of Educational Change*.
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., Fishman, B. J., Gallagher, L. P., Korbak, C., & Lopez-Prado, B. (2009). Is alignment enough? Investigating the effects of state policies and professional development on science curriculum implementation. *Science Education, 93*(4), 656-677.
Penuel, W. R., Frank, K. A., Sun, M., Kim, C., & Singleton, C. (2013). The organization as a filter of institutional diffusion. *Teachers College Record, 115*(1), 1-33.
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., & Shepard, L. A. (in press). Assessment and teaching. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of Research on Teaching*. Washington, DC: AERA.
Penuel, W. R., & Spillane, J. P. (2014). Learning sciences and policy design and implementation: Key concepts and tools for collaborative engagement. In R. K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 649-667). Cambridge, UK: Cambridge University Press.
Pickering, A. (1995). *The mangle of practice: Time, agency, and science*. Chicago, IL: University of Chicago Press.
Pipek, V., & Wulf, V. (2009). Infrastructuring: Toward an integrated perspective on the design and use of information technology. *Journal of the Association for Information Systems, 10*(5), 1.
Powell, W. W., & DiMaggio, P. J. (Eds.). (1991). *The new institutionalism in organizational analysis*. Chicago, IL: University of Chicago Press.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. Z. (2011). Videobased lesson analysis: Effective PD for teacher and student learning. *Journal of Research in Science Teaching, 48*(2), 117-148.
Sabelli, N., Penuel, W. R., & Cheng, B. (2015). The role of informatics in education research and policy. In E. W. Johnston (Ed.), *Governance in the information era: Theory and practice of policy informatics* (pp. 213-228). New York, NY: Routledge.
Sandoval, W. A., & Bell, P. (2004). Design-based research methods for studying learning in context: Introduction. *Educational Psychologist, 39*(4), 199-201.
Scott, W. R. (2008). *Institutions and organizations: Ideas and interests* (4th ed.). Thousand Oaks, CA: Sage.
Spillane, J. P. (2004). *Standards deviation: How schools misunderstand education policy*. Cambridge, MA: Harvard University Press.
Spillane, J. P., & Coldren, A. F. (2010). *Diagnosis and design for school improvement: Using a distributed perspective to lead and manage change*. New York, NY: Teachers College Press.

Spillane, J. P., Parise, L. M., & Sherer, J. Z. (2011). Organizational routines as coupling mechanisms: Policy, school administration, and the technical core. *American Educational Research Journal*, *48*(3), 586-619.

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research, 72*, 387-431.

Star, S. L. (2010). This is not a boundary object: Reflections on the origin of a concept. *Science, Technology, and Human Values, 35*, 601-617.

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research, 7*(1), 111-134.

Trotter, R. T., Needle, R. H., Goosby, E., Bates, C., & Singer, M. (2001). A methodological model for rapid assessment, response, and evaluation: the RARE program in public health. *Field Methods*, *13*(2), 137-159.

Tzou, C. T., & Bell, P. (2010). *Micros and Me:* Leveraging home and community practices in formal science instruction. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Proceedings of the 9th International Conference of the Learning Sciences* (pp. 1135-1143). Chicago, IL: International Society of the Learning Sciences.

Waterman, R. H. (1990). *Adhocracy: The power to change*. Memphis, TN: Whittle Direct Books, Inc.

Weick, K. E. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.

Weller, S. C. (2007). Cultural consensus theory: Applications and frequently asked questions. *Field Methods*, *19*(4), 339-368.

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

Westheimer, J. (1998). *Among schoolteachers: Community, individuality, and ideology in teachers' work*. New York: Teachers College Press.

White, B., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3-118.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.

Yin, R. K. (2013). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: Sage.

Youngs, P., & King, M. B. (2002). Principal leadership for professional development to build school capacity. *Educational Administration Quarterly, 38*(5), 643-670.