Using Co-Design to Test and Refine a Model for Three-Dimensional Science Curriculum that Connects to Students' Interests and Experiences

William R. Penuel[1] Brian J. Reiser[2] Michael Novak[2] Tara McGill[2]

Kim Frumin[3] Katie Van Horne[1] Tamara Sumner[1] Douglas Adam Watkins[1,4]

[1]University of Colorado Boulder

[2]Northwestern University

[3]Harvard University

[4]Denver Public Schools

March 2018

Paper to be presented at the Annual Meeting of the American Educational Research Association, New York, NY.

This material is based upon work supported by the National Science Foundation under Grant Numbers IIS-1147590 and DRL-1748757 and by the Gordon and Betty Moore Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders.

Using Co-Design to Test and Refine a Model for Three-Dimensional Science Curriculum that Connects to Students' Interests and Experiences

For decades, curriculum has been a lynchpin in efforts to improve science teaching and learning. We have looked to curriculum materials as direct supports for improving learning, as evidenced by multiple efforts to test the efficacy of materials in large-scale randomized controlled trials (e.g., Borman, Gamoran, & Bowdon, 2008; Rethinam, Pyke, & Lynch, 2008). And, we have developed innovative ways to integrate learning supports for teachers within curriculum materials (Davis, Palincsar, Smith, Arias, & Kademian, 2017), as well as some powerful programs of professional development (e.g., Seidel, Stürmer, Blomberg, Kobarg, & Schwindt, 2011; Taylor, Roth, Wilson, & Stuhlsatz, 2017). What remains under-explored, however, is how curriculum might figure as a support for building the capacity of a system to produce more equitable opportunities to development in science.

Research-practice partnerships are long-term collaborations that, among other purposes, seek to develop capacities of systems for educational improvement. In many design-based partnerships, design functions as a leading activity—that is, it is a central focal point for organizing joint work (Coburn, Penuel, & Geil, 2013). The purpose of such activity is typically to meet some concrete educational need, including for new materials that might better address standards or realize some new possibility for improving teaching and learning outcomes (Penuel, Roschelle, & Shechtman, 2007). But the hope is also that the extensive time investment required for co-design activity yields more than just a new product. Rather, the promise of co-design is that it develops new capabilities for engaging in ongoing or continuous improvement, as well as understandings of the underlying purposes and structures of educational innovations.

In this paper, we describe the evolution of the curricular co-design process in our research-practice partnership to build the capacity of the partnership to support the implementation of the vision of equitable science teaching and learning outlined in *A Framework for K-12 Science Education* (National Research Council, 2012). The curriculum co-design effort we describe here was focused on developing a set of high school biology curriculum units, with the aim of building the capacity of the partnership itself to develop teachers' skill in engaging all students in meaningful science learning. We describe three iterations made to the co-design process over the past three years, in an effort to build knowledge that other partnerships may draw upon when organizing curriculum co-design efforts in science. These descriptions seek to address two questions we have sought to answer over the past four years:

1. What capacities are needed for and developed through co-design of curriculum aligned to the vision for science teaching and learning articulated in *A Framework for K-12 Science Education*?
2. What tools and routines support designing for helping students meet three-dimensional science learning goals and ensuring that units connect to students' interests and experiences?

The Vision of Science Teaching and Learning in the *Framework*

The Framework for K-12 Science Education outlines a bold new vision for K-12 science education (National Research Council, 2012) that has guided the development of the Next Generation Science Standards (NGSS; NGSS Lead States, 2013) and other new

Framework-derived state standards. As of November 2017, 19 states had formally adopted the NGSS, accounting for 36% of all U.S. children in public schools (National Science Teachers Association, 2018). At the heart of these reforms is the Framework's definition of science education in terms of three dimensions: science and engineering practices, disciplinary core ideas, and crosscutting concepts. The Framework proposes integrating these dimensions to make science and engineering more meaningful to students by engaging them in science and engineering practices to develop and apply the target science ideas (Schwarz, Passmore, & Reiser, 2017).

Realizing the vision of *the Framework* will require making substantive shifts in curriculum materials to support teachers and students in meeting the performance expectations of the NGSS (National Research Council, 2015). For example, the *Framework* calls out the need for curriculum developers to address all three dimensions in lessons and units. This integration requires more than simply including these dimensions as separate areas of attention — engagement in science and engineering practices requires that students' participation in these practices is directly motivated by their goals of making sense of phenomena or solving problems they have identified (National Research Council, 2012; Reiser, Novak, & McGill, 2017). At the same time, the Framework left open questions about which practices, crosscutting concepts, and core ideas to feature in lessons and units, in order to ensure that all receive sufficient attention. In addition, while the *Framework* and others (Fortus & Krajcik, 2012; Fortus, Sutherland Adams, Krajcik, & Reiser, 2015) have called for curriculum developers to consider materials that help students to develop increasingly sophisticated understandings from kindergarten to twelfth grade, it does not offer a completely specified path for doing so.

Similarly, while the *Framework* calls for curriculum that addresses science as a human endeavor that is shaped by and informs historical, cultural, social, and ethical issues, it asks curriculum developers to take up questions of how.

Our particular design team has taken on two related challenges for developing curriculum, namely how best to support the goals of achieving *coherence* and *relevance* from students' perspectives. Reiser and colleagues (2017) argue that supporting meaningful engagement in three-dimensional learning requires developing curricular materials that are coherent from the student's perspective. They suggest that is insufficient to organize curriculum according to the logic or structure of the discipline, as Bruner (1960) argued, because the logic or need to move from one topic to the next that is apparent to the disciplinary expert may not be apparent, convincing, or compelling to students. To engage students meaningfully in science and engineering practices requires that students address questions or problems they have identified and committed to address, and for students to be partners with the teacher and the curriculum materials in figuring out what to do next, in order to make progress on questions or problems.

Another challenge we tackled is how best to select overarching Driving Questions for our units, to provide a framework for students to pursue questions that would on the one hand provide opportunities for students to develop toward targeted performance expectations, while also connecting to students' interests and experiences. For science to be meaningful to students—it is not sufficient for adults to select phenomena that they presume are meaningful to students, especially because teachers' life experiences and interests may be quite different from those of their students. Nor is it wise to assume, as Bruner once asserted, that academic subjects have intrinsic attraction to students (Takaya, 2008). Our design efforts therefore sought, as the *Framework* authors assert is necessary, to develop realize the possibility that curriculum can

> …build on prior interest and identity, [which] is likely to be as important as instruction that builds on knowledge alone. All students can profit from this approach, but the benefits are particularly salient for those who would feel disenfranchised or disconnected from science should instruction neglect their personal inclinations. (National Research Council, 2012, p. 287)

Building Collective Pedagogical Design Capacity through Co-Design

In this section, we describe how we have organized, re-organized, and analyzed our co-design process to build what we are calling the *collective pedagogical design capacity* of the partnership. By this term—which extends Brown's (2009) notion of pedagogical design capacity—we refer to the effort of our partnership as a whole to develop the knowledge, skills, and capabilities for designing coherent and relevant materials that embody the vision of science teaching and learning in *A Framework for K-12 Science Education.*

Creating Affordances for Teacher Design in Materials: Key Dimensions

A particularly useful framework for our work has been Brown's (2002) Design Capacity for Enactment Framework. This framework takes the view that curriculum resources should be designed as tools for influencing instruction, but that they should support teachers' productive adaptation of them to students' interests and needs (Brown, 2009). The framework encompasses both key features of curriculum materials that support design and adaptation, as well as teacher characteristics and how these shape teachers' perceptions of the affordance of curriculum materials.

In this particular analysis of the co-design process, we focus on the "curriculum resources" part of this framework, which emphasizes the need for designers to attend to three different aspects of those resources. A first element in the Design Capacity for Enactment framework are *domain representations.* These are representations of the content to be addressed in the materials, including the relationships among key ideas, or—in the case of the Next Generation Science Standards—the way that performance expectations and their component parts are represented in the materials and used to guide design. Second, Brown's framework emphasizes the need for designers to attend to *representations of tasks.* This includes the various instructions, procedures, or guidance to teachers in the form of lesson plans, activity sheets for students, and recommendations for how to structure or present a lesson. These representations may also include diagrams and artifacts that represent the overall structure of a unit or sequence of lessons. Though not always considered by designers up front, when bringing materials to scale, these representations must be coordinated with other representations of curriculum to which teachers must attend, such as pacing guides and instructional models used in their school and district (Penuel, Harris, et al., 2015). Last, designers must consider *physical materials* needed—supplies, tools for carrying out different investigations, and the instructions for assembling and using these materials in class. Schools and districts may have limits on how much such materials can cost, which can be used, and also formal mechanisms through which these can be purchased and distributed. Thus, material aspects of curriculum intersect directly with larger systems that make use of them.

Co-Design as a Means to Building Collective Pedagogical Design Capacity

Our team was faced with the challenge of designing curriculum materials that embody a new vision for science teaching and learning, when no materials existed that could serve as fully adequate model for design. At the same time, we had access to expertise—in the form of partners at Northwestern University—to people who had been involved in related curriculum development efforts and who had been involved in the development of the *Framework for K-12 Science Education.* Partners at the University of Colorado, additionally, brought expertise in the study of project-based curriculum in science and in assessing three-dimensional science learning. But none of us had ever developed curriculum materials that were coherent from students' point of view or relevant to a wide range of students who encountered them.

Our situation required us to construct some means by which we could engage in cycles of *expansive learning* (Engeström, 1987; Engeström, Pasanen, Toivainen, & Haavisto, 2005; Engeström & Sannino, 2010). Expansive learning cannot be understood fully either by examining how individuals acquire new knowledge and skills or by studying how their participation transforms in practices that are relatively stable. Rather, expansive learning is a way of supporting and studying learning when "learners learn something that is not yet there," that is, when "the learners construct a new object and concept for their collective activity, and implement this new object and concept in practice" (Engeström & Sannino, 2010, p. 2). Expansive learning brings about new forms of collective and distributed agency, as people construct new objects and realize new concepts in material practice (Engeström, 1996). One way that expansive learning may be supported in a partnership is through different kinds of formative intervention research

(Engeström, Sannino, & Virkkunen, 2014; Penuel, 2014), in which participants engage in systematic and iterative processes of problem analysis, design, and testing of designs in practice.

A commitment to collaborative, iterative design is a key principle of Design-Based Implementation Research (DBIR; Fishman, Penuel, Allen, Cheng, & Sabelli, 2013), an approach to organizing research and development in a research-practice partnership. In DBIR, key stakeholders come together to define and study educational innovations to address a concrete educational problem or goal. The target of co-design is initially somewhat flexible, to allow for the agency of participants in design to shape it (Penuel et al., 2007). The design process itself may also be malleable, responsive to what is learned through analysis of the products of design, from implementation, and feedback from stakeholders (Severance, Penuel, Sumner, & Leary, 2016).

The distribution of expertise on a design team often does require some acquisition of new knowledge of different people's perspectives. Educators may develop new knowledge related to learning goals and how to support them through interactions with researchers. And, they may develop new capacities for design. For this reason, co-design is often thought of as a potentially powerful form of professional development for educators (Couso, 2016; Voogt et al., 2015). Researchers also acquire important knowledge of educational contexts from their educator partners, and educators' practice can be an important source of ideas for design. Still, much of what is learned is through joint work to create new forms of activity, rather than through transfer of knowledge from research to practice or from practice to research (Penuel, Allen, Farrell, & Coburn, 2015). For this reason, we focus here on pedagogical design capacity as a *collective* resource for design, rather than an individual one and explore how it developed within our team to support our ongoing work to design and iterate upon curriculum materials in science.

The Current Study

This particular study is a case study (Yin, 2013) of the co-design process itself aimed at building theory and developing knowledge of participatory structures, tools, and routines that can support curriculum development for the NGSS. The development and refinement of the co-design process has taken place within a partnership of researchers and teachers from multiple research institutions and school districts and iteratively refined through successive unit development efforts that we describe here. This paper's analysis focuses on the capacities needed for and developed through the co-design process, as well as the tools and routines that mediate that process.

Participants

This case study shares some elements with a participatory case study approach (Reilly, 2009), in that participants in this study include the authors of this paper, who represent a wide range of roles in the partnership. The members of the research team in the study include participants with expertise in the learning sciences, science education, computer science, teacher professional development, curriculum development, and organizational change. They include people whose primary identities are not as "researchers" but who are part of a university-based research team. One of the authors (Watkins) is a leader in the district, regular participant in research team meetings, and now a graduate student at the University of Colorado Boulder.

There are several high school biology teachers who are participants in the study but who have not contributed to its design and analysis. From them, we have included their perspectives through an analysis of their written reflections at the end of design workshops, artifacts they created in the design processes, interview transcripts, and survey responses.

Sources of Data

Research team meeting notes. One source of data for this analysis is two sets of running meeting notes kept by the research team. One set is for a weekly meeting that includes members of the research team at the University of Colorado and the science coordinator at Denver Public Schools. There are notes for 87 meetings between June 2015 and October 1, 2017 that comprise this part of the data set. A second set is for a weekly meeting that includes some members of the research team at the University of Colorado and the team at Northwestern University. There are notes for 36 different meetings between January 2016, when these meetings began and October 2, 2017, after the conclusion of the third summer design cycle. We used these notes to identify tools and routines that were developed to support the design process; the paper presents tools that were recurring topics in planning meetings.

Design workshop artifacts. Another source of data for this analysis are agendas, slides, and participant artifacts created for and during annual summer design workshops, when most of our intensive co-design work took place. The workshops were a time for teams to construct and/or iteratively refine a *storyline* for a unit organized around helping students build explanatory models of phenomena and solving design challenges that arise from phenomena (Reiser, 2014). In this approach to developing NGSS-aligned units, phenomena serve as curricular anchors (Bransford, Sherwood, Hasselbring, Kinzer, & Williams, 1990) around which a class defines a shared and evolving mission focused on accounting for the “how” and “why” of the anchoring phenomenon. Guided by the teacher, student-generated questions are what motivate exploration of disciplinary core ideas and crosscutting practices. Culminating design challenges give students a chance to contribute solutions to community problems that relate to the anchoring phenomenon, supporting students in seeing how science can be relevant to their everyday lives.

Between spring 2015 and 2017 we planned for and held a total of seven design workshops whose artifacts were reviewed for this paper. In summer 2015, there was a workshop focused on an initial re-design of a unit on ecosystems and a workshop for a new evolution unit. In summer 2016, there was a workshop focused on re-design of the evolution unit and an initial design for a unit on genetics. In summer 2017, there were three different workshops, all focused on redesign of the three units the team had been developing for ecosystems, evolution, and genetics.

Participant reflections. At the conclusion of each co-design workshop, we asked all participants to write reflections to a set of prompts. Participants wrote for between 20-30 minutes, and we also invited participants to share after writing individual reflections with the whole group insights from the co-design. The prompts varied from workshop to workshop, and we analyzed them here to gain insight into what participants felt were important joint capacities, new insights, and new perspectives on possibilities for science teaching that emerged through the intensive co-design process.

2015-2016 Design Cycle

Spring 2015 we began planning for what would be our second year of unit design. We start our account in this second year, because the tools chosen to support co-design reflected some initial lessons learned; a fuller account of the tools that mediated our co-design process is available (Severance et al., 2016). That year, the CU team worked most closely with colleagues at Denver Public Schools, the University Corporation for Atmospheric Research, and with a fourth partner, BSCS, a nonprofit organization focused on curriculum development and research located in Colorado Springs, CO. The primary focus of our work was on re-designing our ecosystems unit and developing and piloting first version of a unit on evolution. Our colleagues at Northwestern University were not able to join us for the planning workshops, though the team sought out their advice regarding storyline development.

Domain Representations Supporting Co-Design

The key domain representations we relied for this cycle were relevant sections of *A Framework for K-12 Science Education* that focused on the specific disciplinary core ideas (DCIs) we had decided would be the focus for each unit. The choice to focus on bundles based on DCIs was rooted in the existing practices of the district, which organized pacing guides around topics in the district-adopted curriculum at the time, *BSCS Biology: A Human Approach.* We engaged people directly in reading and taking notes on hard copies of relevant sections of Chapter 6 the *Framework,* which focused on core ideas in the life sciences, with the intent of familiarizing ourselves as a team as to just what student understandings we needed to develop for the Framework. At this point, our discussions did not involve a deep analysis of the *Framework* text or even make reference to the performance expectations of the NGSS. We did, however, consider the specific grade band guidance for high school, and we used this to help us select an anchoring phenomenon for the unit.

We recognized from the first year of our work the consequentiality of choosing a phenomenon that could anchor a unit of instruction. The phenomenon needed to be more than just a hook for students, but something that would provide opportunities for them to develop understandings of core ideas and crosscutting concepts while engaging in science and engineering practices, and we knew already that sustaining the interest of students could be a problem, especially because students were not used to problem-based units that extended over multiple weeks. We hoped when we came to choose an anchor for our new evolution unit that a single phenomenon could be chosen that would address all the performance expectations for a given domain, though we also had struggled with meaningfully addressing all of the ecosystems performance expectations relying on a single anchor, that is, the problem of deciding what species of tree to plant in a schoolyard that could maximize biodiversity and benefit humans and other organisms.

This year, we introduced a new routine, an interest survey, to try and ensure that our choice of phenomenon would be engaging to students. In the first year, teachers had brainstormed possible anchoring phenomena, and we came to agreement on what would be interesting based on our collective judgment. On the one hand, teachers reported that most of their students had been excited by the opportunity to choosing a tree to plant in their school that our community partner, Denver Parks and Recreation, would provide to their school. On the other, the length of the unit gave rise to many complaints from students that they were spending too much time on trees. Some couldn’t wait to move on to new topics. So, we began to extend and formalize a process for selecting anchoring phenomena that could address the targeted bundle of standards that we sought to address and that could captivate and sustain students' attention for the multiple weeks that students would be studying a particular disciplinary core idea. We created a simple survey that we used to solicit high school biology students' opinions about the interestingness of ten different candidate phenomena. These were the current, rather than prospective, students of biology teachers on our design team, but we reasoned that they might provide a good approximation—better than our own judgment at least—about what phenomena might interest them.

We created a fairly simple representation of the candidate phenomena for the design team to use in our summer workshop for evolution, a representation that might be considered intermediate between a *domain representation* and a *task representation* in Brown's (2002) framework. It included the question linked to a particular phenomenon, the ranking on the student interest survey, a list of the performance expectations we though the phenomenon might address, and a column for recording team votes for particular phenomena. We did not just vote on the basis of this information, however. We created a shared Google doc for small groups to form and then develop arguments in favor of a particular phenomenon. In addition to presenting a driving question and brief description of the phenomenon, this form required advocates for a phenomenon to sketch arguments for why the phenomenon would have been a "good context for developing student understanding of Natural Selection and Evolution and Inheritance and Variation of Traits?" and to identify some investigations that students might conduct, while investigating this phenomenon, and say how students might be involved in refining the question and planning the investigation. In the form, small group members drew heavily from curriculum resources with which they already had familiarity, including *BSCS Biology: A Human Approach.*

The process resulted in the choice of three different initial anchors, one of which changed as the group began to work on it: antibiotic resistance or the evolution of "superbugs"; species extinction and adaptation when an environment changes; and observable differences between hominid and modern human fossils. A small group that focused its design work on the second choice—because it did not specify a particular species or extinction/adaptation event—had to find a case that could illustrate one or both of these ideas. After some research, the group settled on the case of the microevolution of a dark-eyed junco population on the University of California San Diego campus, which involved the birds becoming bolder over successive generations as the population adapted to city life.

Task Representations Supporting Co-Design

To support the design teams' work to designing the sequence of tasks in the evolution unit, we used a version of the initial tool we had used in the first year of design (Figure 1), along with two other designs of the CU Boulder team's creation (Figure 2). Figure 1 shows a rendering of a blank "storyline" used by Northwestern team members to support the design of a coherent sequence of lessons to support students developing understandings needed to explain the anchoring phenomenon. The storyline has some key affordances for supporting the process of building curriculum that is coherent from the student point of view, namely the idea that each lesson is designed around a student question where students engage with some investigative (lesson-level) phenomenon that students engage in science practices to explain, which in turn builds toward a larger explanation of the anchoring phenomenon.

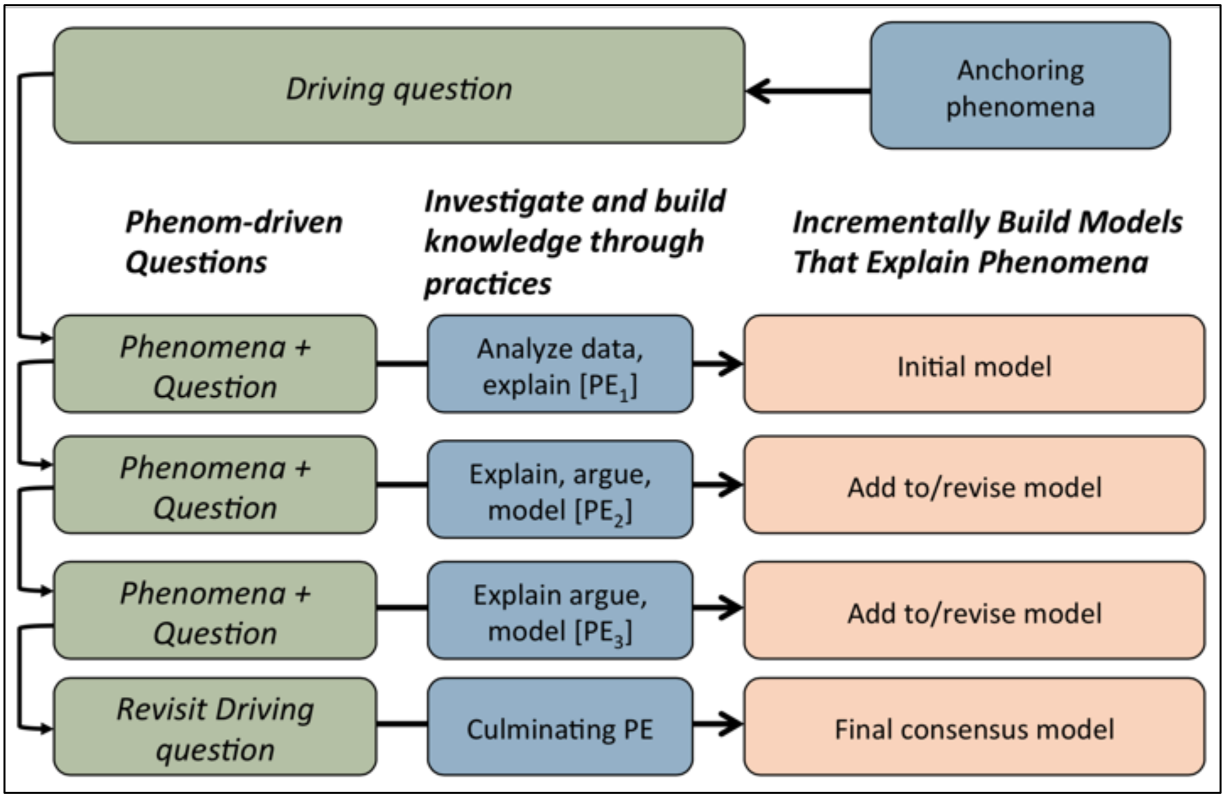


Figure 1. Storyline Representation to Aid Co-Design

In the previous year, the design team had found it difficult as a team to use this representation to write lessons that fit together into a whole. The CU Boulder team had expertise in learning technologies design, as well as regular interactions with another team from the University of Washington, which led us to develop representations that could show how science practices were not isolated but interconnected, as in a "cascade." These representations did not capture the full range of science and engineering practices, and missing in the "cascade of practices" representations were phenomena and student questions, elements that were central the Northwestern team's storyline representation.

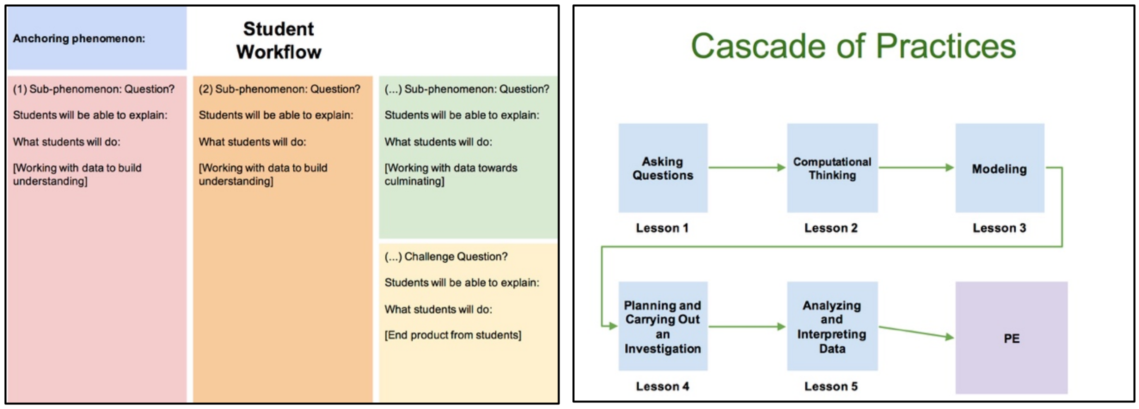


Figure 2. Supplementary Representations of Task Sequences

CU researchers facilitating the team meeting acknowledged that they did not know which tools would prove most useful to the team. We explained, “tools change over time and their purpose may change depending on their use context, and advised teams organized to develop lessons around specific phenomena to “determine what you need to do, choose tool you feel is most appropriate. Ultimately, the design teams used the “Student Workflow” diagram during the workshop to develop ideas for sequences of units, and teams filled out out the storyline template on the basis of that diagram. Each of the team used the three-part structure shown in Figure 2, breaking down their part of the unit related to their anchor into three broad sections, and over the course of the design workshop, we created a rhythm of groups working separately, coming together to share and give feedback to one another, and then returning to their small groups.

As part of the initial design workshop, once teams had outlined workflows and storylines, we assembled as a group to create a rough pacing guide for the unit. The ecosystems unit had proven to take much longer to implement than was allocated in the district pacing guide, and so with district leaders at the table this time, we wanted to conduct an initial check of the “viability” of the pacing for the unit. At this point, no lessons had been developed, and only sketches of individual investigations existed. But the team judged it important to estimate how long the unit would take, to ensure the curriculum could be implemented within the district constraints early in the design process.

One of the challenges in the first year had been a discontinuity that researchers on the team noticed between initial storylines and draft lessons. In addition, teachers provided feedback to the researchers that many of the lesson plans did not include required components, such as a Content Learning Objective or CLO. They complained that different design teams used different vocabulary and created slides that did not follow a consistent format. So, we created a structured lesson plan template to facilitate lesson development, along with templates for slides. Among other elements, the new template included a Content Learning Objective and began with a "Do Now," a requisite component for lessons in many Denver schools that adhered to an instructional approach developed by Lemov (2010). The "Do Now" entailed providing students with something to do independently, immediately upon arrival in the classroom. It also included many elements that teachers would expect to find on a lesson plan they use to plan: information about the expected duration, standards addressed (NGSS, Colorado Academic Standards in Science, and Common Core State Standards in both Math and ELA), assessments, materials needed, and of course, descriptions of the tasks to enact with students. A slide deck template with six basic sections was developed: A Do Now/Warm Up; Agenda; Objectives; Activities; Homework; and Closing Activity. For lessons involving assessment tasks, we encouraged teams to develop a culminating, three-dimensional assessment task (National Research Council, 2014) using a set of task templates for integrating the science and engineering practices into assessments. These templates evolved from an earlier research project that focused on developing such assessments as part of an evaluation of a project-based middle school science curriculum (DeBarger, Penuel, Harris, & Kennedy, 2015).

Material Coordination: Design Challenges

With respect to *physical materials,* there was attention to building an inventory of supplies, but one of the particularly challenging aspects for the team was the material coordination needed to implement unit design challenges. Implementing the design challenge among even a relatively small group of teachers using the ecosystems unit had proven difficult for the team, because of the coordination among researchers, teachers, and volunteers for the Denver Department of Parks and Recreation. Teachers needed information about tree inventories and constraints on planting, and teachers needed to secure permission from building grounds staff at schools for planting. Researchers needed to relay information to Parks and Recreation staff about the timing of tree planting, which varied widely because each teacher took a different amount of time to enact the unit. Parks and Recreation staff had to coordinate with their own volunteers, who then communicated with teachers and researchers to make a concrete plan for planting. In selecting our evolution design challenge, we sought to reduce the complexity of coordination required, while still maintaining the principle of connecting students to the community in the challenge.

To select the challenge, we facilitated a series of discussions over the course of our five-day design workshop in the summer, centered on brainstorming and then researching alternative design challenges that might serve as culminating experiences. We also set the intention to integrate lessons related to the challenge between the different segments of the unit marked by different anchoring phenomena. To scaffold the brainstorming process, teams self-organized around 2-3 ideas to define the goals for the challenge, identify potential community partners, the products students would produce, types of activities needed to support students in developing a product, and the disciplinary core ideas addressed. Though we left without a clear decision, ultimately the team by late summer had agreed to focus the challenge on creating a digital infographic for a school or community health clinic focused on antibiotic resistance. Although this effort would require less coordination with external partners (in fact, some partners were inside the school), it still would require selection of free or low-cost digital tools and resources for supporting students in creating infographics that could be relatively easy to sustain.

2016-2017 Design Cycle

In our third year, the team focused on revisions to both the ecosystems unit and to evolution, and we developed a first draft of a third unit on genetics. The primary collaborators for revision were University of Colorado Boulder researchers, Denver teachers, and the Northwestern University team, now known as the Next Generation Storylines project. This design cycle saw many changes to our design process focused on our approach to representing the domain and to developing storylines. The Northwestern team helped the Colorado-based group develop more fine-grained, student-centered storylines, and the University of Colorado team created representations to help facilitate the movement from storyline to lesson. The team also began to create three-dimensional assessments as a new kind of task representation, and we developed a third type of design challenge for the new unit, which involved the design and enactment of a process, a

World Café, that would connect students across the city to one another in an event for students, their families, and the community.

Changes to Domain Representations

Two innovations with respect to domain representations to facilitate the design and revision process were to adopt an approach to analyzing focal DCIs the Northwestern team was using and to construct student explanations of anchoring phenomena. Where our approach in the second design cycle for analyzing the domain of evolution represented in the *Framework,* it became more formal and collaborative, at least between the University of Colorado and Northwestern teams. Meetings became focused on the task of "unpacking" or analyzing individual sentences in the *Framework* to identify facets or component parts of the DCIs that were either to be developed or might be prerequisites for student mastery. The analysis also included the team's open questions about the boundaries for assessment that might be implied, which we checked against the performance expectations of the NGSS. Figure 3 below shows a sample of our analysis of three sentences from the *Framework*. In the figure, the boldface text in quotations comes from the *Framework,* and the text below represents the researchers' elaboration of what students would need to understand and be able to apply to explaining phenomena and solving problems.

Natural selection leads to adaptation—that is, to a population dominated by organisms that are anatomically, behaviorally, and physiologically well suited to survive and reproduce in a specific environment.

- Note: Focus is on organisms that survive, not on the distribution of traits here, in contrast to text related to natural selection.
- Natural selection is the cause, adaptation is the effect.
    - Tie to crosscutting concept: Cause and effect.
- Organisms that can have survival advantage due to different kinds of variation.
    - They might have anatomical features that allow them to survive better.
    - They might engage in behaviors that allow them to survive better.

    - They might have physiological (e.g., pertaining to regulation and homeostasis) traits that allow them to survive better.
- Characteristics that are an advantage to survival and reproduction in one environment may not be advantageous in another.
    - Asking students to predict whether organisms with particular characteristics and that dominate one environment would do as well in another could be a good scenario, something akin to an invasive species. (Another LS2 connection here)

That is, the differential survival and reproduction of organisms in a population that have an advantageous heritable trait leads to an increase in the proportion of individuals in future generations that have the trait and to a decrease in the proportion of individuals that do not.

- There are connections here to natural selection, focusing on the changes to proportions of individuals over time with organisms with different traits.
- Focus is on proportion: Tie to crosscutting concept: Rate, proportion, and scale.
- Emphasis is both on the increase in proportion of individuals with traits that are advantageous and decreased in organisms without the trait.
- Here, over time is operationalized as being over successive generations.

Adaptation also means that the distribution of traits in a population can change when conditions change.

- Different conditions can change what is selected for
- A change in what traits are selected for results in different adaptations.
- That change also shifts the relative distribution of traits in a population.

Figure 3. Sample Analysis of *Framework* Text (LS4.C)

The analysis was painstaking, but we found that by undertaking it as a team, we gained a deeper understanding of where our current units fell short, in terms of providing student opportunities to learn. We realized, for example, the need for greater support for students to generalize across two phenomena of evolution related to bacteria and juncos, in order to help students gain a more complete understanding of survival advantage due to different kinds of trait variation within a population (e.g., physical characteristics and behavior). For the new unit, analysis of the domain jumpstarted the phenomenon brainstorm process. As we conducted the analysis, we made notes of candidate phenomena and investigations where students might have opportunities to figure out a core idea. We recognized that the domain analysis yielded statements similar to those found in the Evidence Statements for the NGSS on the Achieve website (https://www.nextgenscience.org/evidence-statements); however, the process of undertaking the analysis ourselves required us to engage in a sensemaking process that helped us tack back and forth between unit content and the domain. Importantly, it provided us with a better sense of where our analysis of the domain for evolution had fallen short, in terms of capturing the details of what students needed to know and be able to do.

A second innovation was the development of student explanatory models for existing phenomena (ecosystems, evolution) and candidate phenomena (genetics). The Northwestern team, before beginning the process of developing a storyline, constructed an elaborate student explanatory model for each anchoring phenomenon. Whereas we had engaged only a small number of team members in the process of analysis of the *Framework,* we constructed explanations for evolution as part of our summer design workshop, when educators and other team members were present. We thought it important to do so, so that members of our team might develop confidence in our own abilities to select anchoring phenomenon based on a careful analysis of the science behind our selections, something we thought we had not given sufficient attention to in our first two design cycles. We developed student models on large pieces of butcher block paper, working initially in small teams and then comparing our models to one another to build a synthetic one. The process entailed each small group spending time writing and then developing our own understandings of the science behind each phenomenon. We realized in comparing models to one another that both we and students would likely need some specific prompts to elicit aspects of the model that were critical to include without giving away the science.

The construction of these explanatory models served the design process in several ways. First, it helped the team members develop the requisite understandings of the science ideas that we would need to support students in doing as part of the units. The process of writing explanations helped us begin to imagine the kinds of investigations that would help students build, incrementally, an understanding of the phenomenon that made use of core ideas (e.g., natural selection, adaptation). In addition, the insight that students would need prompts to elicit what they had learned helped us make a step toward constructing a form for embedded assessment within the units. The embedded assessment would entail the construction of an explanatory model of the anchoring phenomenon, in which students put the pieces together they had learned across the lessons, supported by prompts designed by curriculum writers to do so. There was already the idea from storylines that pieces of understanding might contribute to a class consensus model, but the practice of writing a student model and prompts to elicit it specified more clearly how. Last, for the new unit we were developing, as we elaborate below, writing a student model helped us assess the viability of candidate anchoring phenomena.

Our representation of candidate phenomena and process for selecting phenomena for the new genetics unit became more elaborate and specified during this design cycle. For each candidate phenomenon, we created a representation that included a possible driving question, a short description of the phenomenon, and a paragraph-length explanation of the phenomenon. In the explanation, we embedded codes from the NGSS where we saw connections between components of the explanation and specific performance expectations. We also integrated into the descriptions evidence from the student interest survey and created a new metric to help us decide what candidate phenomenon might be best: the ratio of the mean rating to the standard deviation of the ratings. We reasoned that a high score in this metric would indicate a phenomenon that was both interesting *and* interesting to a wide variety of students. In the design workshop, we engaged teams in a time-limited exercise involving analysis or unpacking of the standards, looking at student interest data together, and then writing a more extended explanatory model for four different phenomena that we thought were most promising. We then used these explanatory models to help us decide on two anchors for genetics, muscle atrophy among boys suffering from Duchenne Muscular Dystrophy and the use of CRISPR-Cas9 gene editing to search for cures for genetic diseases.

Changes to Task Representations

In this design cycle, we made significant changes to multiple *task representations* and developed new representations. We abandoned the student workflow and cascade of practices diagram, and the Colorado-based team let the Northwestern team take a large part of our evolution storyline related to antibiotic resistance and re-write it prior to the summer evolution re-design workshop. The Northwestern team introduced a new representation of storylines to us, which would become the form that we would use for this and a subsequent iteration of the design. They also, in re-writing a part of a storyline for the Colorado-based team, helped us to see ways that our own storylines were neither sufficiently driven by the student perspective nor detailed enough to identify places in the storyline where there were "leaps" between lessons (i.e., the next row in the storyline did not flow from the previous one). Figure 4 below shows a row in a storyline, using the new representation introduced by the Northwestern team.

Teacher Guide

NGSS High School Evolution Storyline

Why Don't Antibiotics Work Like They Used To? (Bend 1 - Bacteria)

This Lesson....What we are doing now: This is the first lesson in the series. Up until this moment, students may never have considered why antibiotics today don't work as well as they used to . In this lesson you will draw upon their experiences with antibiotics and introduce an anchoring event (Addie's case) to which you'll refer throughout the unit. Students will analyze a Frontline video about a pan-resistant bacterial infection case in a little girl, named Addie, Students will ask questions regarding why aren't the antibiotics helping Addie get better. You'll guide them to decide to start a fact timeline with what happened to Addie in chronological order and separate out the differences between species of bacteria "kinds" and strains "types" with each kind.

Lesson Question	Phenomena	Lesson Performance Expectation(s)	What We Figure Out (CCCs & DCIs), *New Questions* and **Next Steps**
L1: How did this little girl (Addie) get so sick? (1 period) S *Building toward* *NGSS PEs:* *HS-LS4-2 & HS-LS4-4*	A Frontline video clip introduces us to the case of a little girl (Addie) who came into the hospital with a bacterial infection. After several weeks of antibiotic treatment she ends up with life threatening pan-resistant bacteria:. We share our own experiences and related cases with sickness and antibiotics.	**Ask questions** that arise from careful observation of unexpected results, to clarify and seek additional information *about how bacteria caused this little girl (Addie) to become and stay so sick despite receiving antibiotics.*	We have lots of experiences related to bacterial infections and taking antibiotics. For example, some of us have had to take antibiotics from a doctor, some of us have taken other family member's antibiotics. We know family members who had to take antibiotics because they had pneumonia, while others took antibiotics due to a minor infection from a cut. We noticed some important patterns in Addie's case and organized these events into a timeline: • Addie was cut while playing on the playground in 2011. • She came to the hospital sick from an infection caused by one type of bacteria from this cut. • Doctors gave her an antibiotic that worked for awhile, but then stopped working. • She ended up getting another infection from a different type of bacteria while in the hospital. • Doctors then gave her a new antibiotic that worked for awhile, but then it stopped being effective and Addie got sicker. They tried a third antibiotic and Addie started to improve but this antibiotic also stopped being effective and she got worse once again. • They repeated this process trying every antibiotic available to them. We identified some important differences in bacteria that the doctors refer to: • Addie had a type of bacteria in her lungs that was not the type they expected (pan-resistant). • Pan drug-resistant bacteria have "armor" that the antibiotic can't penetrate. • There are different types of bacteria, resistant vs. susceptible (non-resistant). • And there are different kinds of bacteria: Staph (*Staphylococcus*) and *Stenotrophomonas*. We decided it was important to pay attention to the different types and kinds of bacteria and kept track of this information in a summary chart. She seems to have had infections from two different "kinds" of bacteria: (a) *Staphylococcus* and (b) *Stenotrophomonas*. Some of both kinds of bacteria appear to have been killed by antibiotics. However, some of both kinds of bacteria were not killed by those antibiotics. It appears that there are two types of bacteria within each kind of bacteria—those that were resistant to one antibiotic (and lived) and those were not resistant to an antibiotic (and died). *We have a ton of questions! Why is that happening? Can this happen to me? From where can you get bacteria? How can a substance that helps wipe out the bacteria work for a bit, then stop working? How do antibiotics even work?* **After making a record of our questions, we identify some next steps to pursue. Because we are really concerned about whether this can happen to us, we want to know if cases like Addie's are common, or if this is a pretty isolated case.**

nextgenstorylines.org

These materials were developed with funding through a grant from the Gordon and Betty Moore Foundation to Northwestern University and the University of Colorado Boulder.

5

Figure 4. Storyline Representation for 2016-17

Several features distinguish this single row from the schematic used in the previous year, beyond the detail that is evident. A domain marker is represented in a gray box on the bottom left, namely the performance expectations that the lesson "builds toward," to signal that it is focused on this performance expectation but does not as a standalone lesson "meet" that standard. Second, the phenomenon includes a hyperlink to a video—in this case, one that is used to launch the evolution unit. Third, there is a lesson level performance expectation in which color coding is used to signal the disciplinary core ideas (orange), science and engineering practices (blue), and crosscutting concept (green) for that lesson. Finally, there are conventions for developing the "what we figure out column" for representing in scenario, first-person plural form, what students experience during the lesson from start to finish. Students use the crosscutting patterns to name things they see in the video, and they name some differences in bacteria that are making the girl with MRSA sick. Finally, at the bottom of the row are questions written in purple that indicate the curriculum writers best thinking about the questions that students might have at the end of the lesson.

This first person plural account of the student experience is akin to what human-computer interaction designers might call a formal design scenario. Designers in that field use scenarios to imagine how people might use technologies that do not yet exist (Alexander & Maiden, 2004; Carroll, 1995). They help designers attend specifically to how technology tools support practice and how users interact with tools to accomplish their goals. As such, they help designers attend to the social practices in which technologies are embedded in ways that a list of user requirements cannot. In many ways, the scenario development served a similar function for our team, helping us keep the student perspective focal—their questions, their purposes and plans, their activities, and their conclusions—all of which had been easy for us to lose both in developing storylines and writing lessons.

Of note is another type of routine we created that is not well represented in Brown's (2002) typology of resources for design, and that was for making use of available implementation evidence to inform the re-design of the ecosystems and evolution units. We organized presentations in a systematic fashion so as to allow for time for the Northwestern team to present ideas based on their expert judgment, but we also made time to review evidence from student exit ticket data on which lessons they found most exciting and most boring, as well as which ones students felt were most and least connected to the anchoring phenomenon or design challenge. Third, we gave time for teachers to present their experiences of teaching the units, particularly what they felt worked and what didn't. Then, on the basis of this set of presentations, the group set some priorities for revising the units. In this discussion, we purposefully positioned three types of expertise as equal—research evidence, expertise in curriculum development, and teacher experience—to arrive at this synthesis. Thus, implementation evidence gathered by researchers did not by itself inform or determine the course of re-design.

In our genetics workshop, which focused on design of a new unit and involved the member of the team based in Colorado, we stumbled upon an innovation for beginning storyline development once we had chosen an anchoring phenomenon. As part of that workshop, we had planned to immerse participants—some of whom were new to the design process—in an introductory lesson for a phenomenon-based unit, using a genetics phenomenon that was one of our candidate phenomena. We found a video introduction to the CRISPR-Cas9 system, which presented to participants in an activity of in which we asked them to write down and discuss what they noticed and wondered about, with respect to how scientists were using new gene editing technologies, how the gene editing technologies work, and whether or not they should be used. We saved these questions on a large sheet of sticky paper, and when we chose a second phenomenon, we found a video introduction to Duchenne Muscular Dystrophy that proved to be powerful launch point for that phenomenon. We repeated the process of eliciting questions from the group, this time to explain what we saw in the video, acting in "student mode," imagining questions we might have if we were high school students. We then engaged in a group activity to prioritize the questions, deciding as a design team which ones we would need to address first, because answers to them were necessary for answers to other questions we generated. The result was a really rough draft of some questions that we anticipated could constitute rows in our storyline. Though these questions would not correspond exactly to the rows of our first draft of a complete storyline, many were indeed questions we had generated as a group from what we noticed and wondered about from the two presentations of the phenomenon.

A new task representation developed for both units was a lesson plan structure, organized around a set of questions that the Colorado-based team noticed were implicit in analyzing the Northwestern University team's writing in the "What we figured out" column. We noticed that most lessons began with looking back at what was done the day before, and having students recall what they figured out. Then, students said what question they needed to answer next, and brainstormed ideas about how to pursue the question. In the storyline version, the students would come up with ideas that the curriculum design team anticipated the students might generate for investigations, and they would carry out those investigations. (In real classrooms, we understood this would require guidance from the teacher in most instances.) Toward the end of the lesson, the students would come back to the question for that day and discuss what they figured out, and at the conclusion of class, they would identify some next questions they needed to pursue, either that arose from the investigation or that students had generated when the phenomenon was first presented. Figure 5 shows how we translated this observation into a basic sequence into a formal lesson structure.

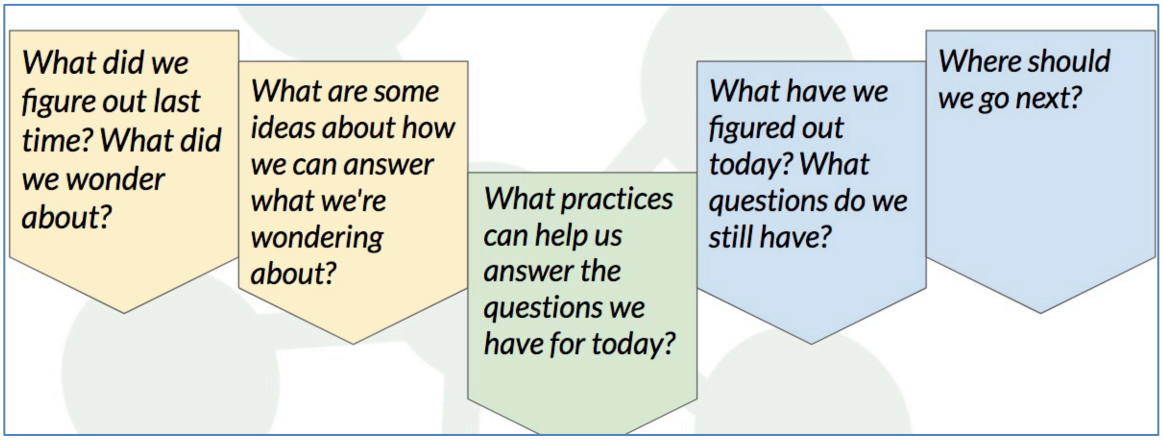


Figure 5. Lesson Structure for Storyline Units, 2016-17

We also created a template to help curriculum writers move from storyline to lesson writing, which was intended to help improve the alignment between the storyline and the lessons. As noted above, we had noticed a number of ways that lessons were not consistent with the storyline. This included a tendency for writers to slip into "present content first-then apply" lesson structures, which is inconsistent with the approach to teaching that is reflected in Figure 5, in which students and teachers partner to set the direction for the day's lesson, and students engage in science and engineering practices to answer questions they have generated or helped to generate. This template was only partly taken up in the design workshop. The process of developing a storyline is so intensive, it proved difficult for teams to tack back and forth between the storyline representation and the template. Some sub-teams did not use the template at all for their part of the storyline.

New Approaches to Material Coordination of Design Challenges

In this design cycle, we attempted a different approach to a design challenge. As in the previous year, we brainstormed different ideas for a culminating challenge, once we had chosen an anchoring phenomenon. A sub-team worked on developing different ideas for the challenge, and a new criterion we established based on our ongoing challenges in coordinating community connections in our ecosystems unit was that it needed to be feasible for a teacher to implement. There was some debate over the principle of the design challenge needing to connect to the community, but we held fast to that principle. However, we chose a challenge that could be implemented within a single classroom or school or as part of a larger district initiative. We also chose a challenge that did not involve designing an object but rather focused on the design of a process, specifically, a type of deliberative dialogue involving science called a World Café. One reason we chose this challenge was that the design team members were themselves divided over the choice of CRISPR-Cas9 as a phenomenon. Many worried explicitly about the ethics of gene editing technologies and about the kinds of controversial conversations they might have to facilitate in their classrooms. One district leader was worried about this, too. Rather than treat this as a "bug" of the phenomenon, we decided to treat it as a core feature of the phenomenon, namely an opportunity to help students learn how to engage sociopolitical and ethical issues that arise from new scientific discoveries and technologies. Further, we decided to emphasize explicitly that design could be about a practice, and part of the design students would need to consider are what questions to ask and what science to bring to a dialogue. For this year at least, though, adult concerns led to a hybrid design, in which a team comprised of a district leader and two members of the Colorado research team planned keynotes for a city-wide World Café, and students acted as table facilitators, leading discussions of questions they had generated but that had been vetted by adults.

2017-18 Design Cycle

As the two research teams at Colorado and Northwestern began to plan for the summer workshop in 2017, we thought it would be a good idea to bring Denver teachers together with Illinois teachers who were writing units, so that we could broaden the base of designers on our team. The Illinois teachers who came to Colorado to support the redesign of the ecosystems unit and who also joined for the evolution redesign in Chicago had experience in developing multiple units using the storyline approach with the Northwestern team. In Colorado, we invited two teachers who had piloted units to join the design team.

By this fourth design cycle, for the Colorado-based team, all of our units had been piloted at least once, and we had established processes for making use of a range of implementation evidence in revision. We decided to replace one anchor in ecosystems, and we followed the process we had developed over the previous two design cycles for doing so. The changes in the design process were to create a new routine for re-organizing sequences of lessons to be more coherent from a student perspective and the development of a set of lesson routines—which presented more varied lesson structures that represented the diversity of lesson-types that the Northwestern team had helped create across a range of storyline units developed with our team and others in Illinois, Michigan, and Connecticut. We continued to struggle with and explore options for sustaining our ecosystems design challenge, identifying a new internal partner at the University of Colorado that was already engaged in a neighborhood tree planting initiative in Denver.

New Routines to Support Task Representation

Once a storyline has been developed, lessons written, and even after implementing lessons, it is sometimes challenging to see where there are likely discontinuities from the student perspective. As one of the groups re-designing what was now referred to as a "bend" in the ecosystems storyline—a coherent sequence of lessons organized around a single anchor—was struggling with this problem, the group decided to re-construct several of the lesson's storyline elements into a new format—driving questions, the investigative phenomenon and what students do in the lesson, what they figure out, and what questions they likely have at the end—using large 4"x6" sticky sheets. The team decided to distribute responsibility for reviewing individual lessons and then as a group reconstructed the storyline components for that lesson/row in the storyline on the wall (Figure 6). As each row was posted, discontinuities in the storyline became obvious to group members, especially when a question at the end of a row did not match up at all with the question guiding the next investigation. And, because the stickies could be moved around, the team could work as a group in the space to rearrange them easily and add rows as needed.

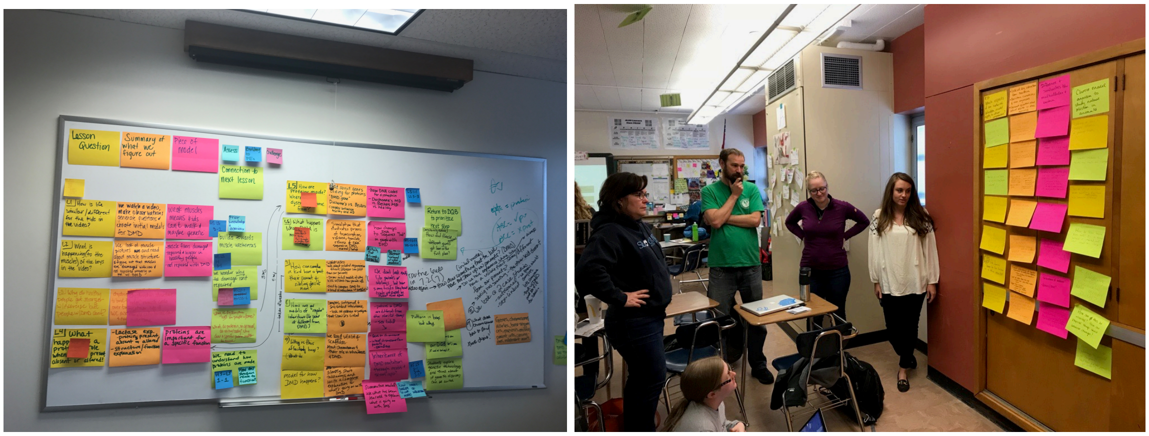


Figure 6. Reconstructing a Storyline Using Sticky Notes, Lesson by Lesson, in Design Workshop and in PD

The Northwestern team also introduced in the workshop a set of teaching routines that corresponded to different lesson structures. In one respect, these routines were simply reifications of lesson structures that had begun to emerge through many efforts by that team to develop units. But they also served a useful tool for redesign in our workshop, to make visible some aspects of lessons that were partly implicit. The Anchoring Phenomenon Routine (schematic shown below in Figure 7) is one example described in detail below. Other routines include the Navigation Routine, which is a version of our lesson structure developed in the previous design cycle, the Problematizing Routine, in which the teacher seeds a phenomenon with the express purpose of eliciting disagreement and push students to design a way to resolve it, and the Putting Pieces Together Routine, which is used to support students in integrating knowledge they have developed over a series of lessons. This routine can culminate in an extended assessment task in which students apply what they have learned to a new, related phenomenon.

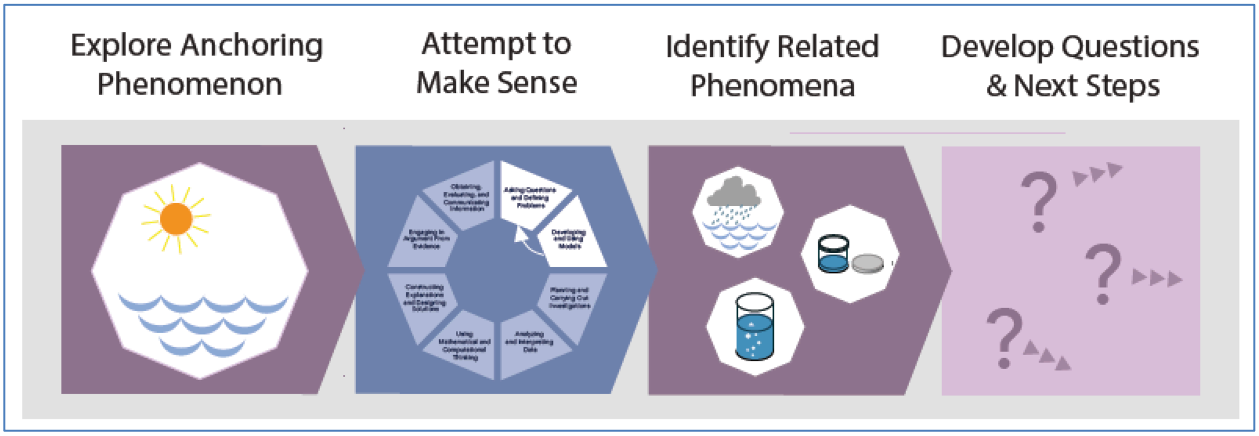


Figure 7. Anchoring Phenomenon Routine

The Anchoring Phenomenon Routine is used in the first set of lessons, when the teacher introduces the phenomenon to students. In it, the teacher first presents the phenomenon for the unit to students in the form of a demonstration, an investigation, or a video. Students write down what they observe or notice, and they write down and discuss their questions as well. The teacher then creates a public record of what individual students noticed and the questions they had. Next, students work in small groups or individually to generate initial explanations for what they saw. Note that they do so well before they could be expected to construct a complete explanation for the phenomenon; this step both provides the teacher with a sense of what knowledge students bring and helps students generate a sense of what they do not yet understand about the phenomenon at hand. The process of building a public record of areas of consensus and disagreement that follows this activity further facilitates question generation.

The next phase of this routine presents students with the opportunity to make use of their own experiences, focused on the question, “Where else does something like this happen?” The teacher asks students to think of experiences they have had that might be related to the phenomenon at hand and how they might support their making sense of that phenomenon. The class builds a public record of these experiences, adding questions students might have about these experiences that they might pursue—either individually or as a class—over the course of the unit. This step supports the teacher in helping connect the phenomenon to students' own experiences in that particular class.

In the final phase of this routine, the class creates an artifact called a Driving Questions Board (DQB) that sets the initial direction for the unit as a whole. The DQB is a jointly constructed artifact that reflects the questions that the class believes will help them make sense of the anchoring phenomenon and answer the Driving Question (Weizman, Shwartz, & Fortus, 2008). The class generates a Driving Questions Board by asking individual students (or pairs of students) to again frame a question, this time that the class can take up as a whole to make progress in answering the Driving Question or explaining the anchoring phenomenon. Students each have an opportunity to post their question to the board and defend their question. The teacher and student then group and prioritize the questions, focusing on an order that the class agrees is logical for pursuing an answer to the Driving Question. This collective effort is intended to build ownership over the Driving Question and the associated student questions (Weizman et al., 2008). As a final step in the routine, students outline some foci for investigations they might pursue in the class to answer the questions they have generated.

Both of these sets of routines proved to be useful to introduce to teachers learning to implement the new units. The routines helped communicate curricular structures of storyline units to teachers, structures that are not part of most science curricula today. The sticky note routine helped the teachers see the overall flow of the units, including where there remained discontinuities—at least from participating teachers' perspectives. The co-design team's willingness to change these showed the teachers, moreover, that the team was committed to supporting their own productive adaptations of materials.

The team in Colorado continued to wrestle with how best to sustain the design challenges in this design cycle. We explored a new partnership with another center at the University of Colorado Boulder that was engaged in a tree planting effort in a low-income Denver neighborhood with few trees. But rather than simplifying coordination, the team discovered we had to help the staff in this center partner more effectively with schools and on turnaround times they were not used to making. While the team remains committed to the idea of each unit culminating in a design challenge that connects youth to their communities, we are planning a new strategy altogether of creating a process—not unlike our process for selecting an anchoring phenomenon—that a teacher team in a school can engage their students in to select a design challenge related to a community concern.

What Collective Design Capacities Were Needed and What Capacities Were Developed through Co-Design?

In this section, we take a step back to consider ways that tools and routines emerged from some capacities that we found we needed to develop as a design team but did not yet have, and we draw on reflections from participants to articulate some of the new capacities that routines introduced successfully supported. Though we include some individual reflections here about what participants say they learned, we focus on ideas and capabilities mentioned by multiple individuals and that—in our view at least—reflect the capabilities needed at the team level.

Capacity to Create Tasks to Support and Assess 3D Learning

From the first design cycle, one of the activities participants found most useful was the explicit guidance regarding the development three-dimensional tasks to assess student learning, using the Science and Engineering Practices Task Formats for developing tasks. In some ways, this was surprising to us, as our principal activity was on design of a storyline for the curriculum materials. But for many teachers, the activity of designing assessment tasks helped them see how the three dimensions could come together in practice and made clear to them how the standards were different. As one teacher commented in her reflections from the 2015-16 design cycle:

> The 3D assessment training and practice exercise was very helpful in being able to dissect how to get my kids to learn what I want them to learn. I believe it will make me a better teacher and have better tests and assessments for my students.

We have observed through other lines of work within our partnership and within a partnership involving a network of state-level teams in science (Penuel et al., in press) that beginning with analyzing and adapting tasks to be 3D does help groups begin to see more clearly some of the key shifts in what is expected of students to do and is consistent with the "backward design" (Wiggins & McTighe, 1998) approach to developing instructional sequences with which many educators are familiar.

At the same time, the tools and routines used in the first cycle provided insufficient support to the team for developing a shared understanding of how practices and crosscutting concepts could be integrated with understanding of the disciplinary core ideas. While they felt more confident in how to teach disciplinary core ideas, the connections between the storyline development process as we engaged participants in it were not clear to some of the co-designers. Moreover, the connections between assessments and how lessons built toward mastery of performance expectations were not yet clear—and would not become clear to us as a collaborative design team when we began to develop student explanations of phenomena prior to develop a first draft of a storyline for a unit.

Capacity to Choose Phenomena That Motivate the Need to Develop Understanding

Participants beginning in the second year began to see the value in the deliberate process through which we engaged in the process of selecting phenomena. Some had experienced directly their students' initial excitement with the tree planting project of ecosystems, as well as the waning of students' engagement as the unit unfolded. Others had found it difficult to address some key disciplinary ideas through the first version of the unit. Those teachers in particular appreciated the deepening of the approach used to select new anchors for evolution and genetics. Tools and routines cited by multiple participants that supported this process included the systematic unpacking of the *Framework* text for the Disciplinary Core Ideas bundles, and the exploration of the science behind candidate phenomena.

One teacher commented that after the second design cycle, she appreciated the need for presenting students with engaging phenomenon so that they would "feel a personal need to explore it deeper." This is one good way to capture the importance of choosing a phenomenon that not only addresses performance expectations but also captures and sustains students' interest. And while the practice of generating candidate phenomena that met both criteria was not distributed equally across members of the design team, we found teams could easily engage in discussions of the relative viability of different phenomena, after they had engaged in some analysis of standards and student interest data and constructed student explanations for phenomena. The tools and routines refined by the second design cycle, then, supported the collective work of selecting compelling anchors that could motivate the need for students to figure out core ideas.

Capacity to Take the Student Point of View on Coherence

It took longer for the team to develop the capacity to take a student view on the coherence of the curriculum, even though after the very first workshop, participants began to appreciate both the novelty and value of designing with the student point of view at the center. As one teacher who started in the first year with our partnership and has worked for four years with us put it that first year:

> Ever since I finished school to be a teacher (10 years ago), I have been interested in getting away from the cookie cutter labs and trust me science and move toward true inquiry science where students are solving a real-world problem and truly uncovering the science content along the way.

Experiencing a part of a storyline as students helped participants gain more direct insight into what it means for curriculum to be designed to be coherent from the student point of view. After seeing a member of the Northwestern team facilitate a group discussion with adult design team members about how a simulation might help the group explore what was happening inside the bloodstream of a girl with MRSA, A computer science student who developed simulations to support the units in development commented:

> By having the students develop the list of needs they have for the simulation, you really start to see how the simulation is accomplishing what the students want,

> and the students aren't just being thrown in front of this magic box that has patterns they can find if they try.

A teacher from Colorado said that the re-written segment of the storyline that this simulation illustrated served as an "exemplar" she had been looking for in the first year she was involved in design, but that had been missing for her.

But neither the storyline tool or immersion experience itself was sufficient for the team to develop a coherent series of lessons or to keep the student point of view central. Time and again, the team struggled to "stay in student mode" in developing specific lesson plans. What struck the team as most needed were capacities for designing lessons in which teachers could "not only get them [students] to want to ask *any* questions, but to drive/guide them to ask the questions you plan for, in order to set up the next activity." A researcher commented that an important but challenging part of design was about students coming to an investigation within the storyline with "the right information -- and not too much or too little—to be in the position to want to know the next thing." As teachers came to see how important building knowledge incrementally over multiple lessons would be, some voiced worries about students who might lose track or fall behind, because they were not absent or did not catch on as quickly as other students. There would need to be ways for the group to design opportunities for all students to track what the class was learning and ensure that they all still felt a part of the class' endeavor to explain the anchoring phenomenon or solve the design challenge.

Capacity to Support Incremental Model Building through Teaching Routines

The capacity to design supports for students in building knowledge of phenomena—specifically, explanatory models of phenomena—has proven to be a critical task, and three tools and routines particularly helped the team to develop this capacity. The need for some of the supports we have developed arose not within design meetings but as we began to introduce new teachers to the designed units as part of professional development. Though we decided to limit the scope of this paper to supports that we developed to support the design process, it is worth noting that some of the tools to support knowledge building within lessons, such as the Putting Pieces Together Routine, have arisen from the sense among teachers of the need not just to help students recall the knowledge they have developed through a series of generalizations, but to apply it to the anchoring phenomenon and develop preliminary generalizations that can be tested on new, related phenomena. And, based on our observations of teachers using the units—including those who have been part of the collaborative design process—the capacity of the team to design supports sufficient to facilitate high quality enactment across a wide range of teachers is still needs much more development.

In addition to the Putting Pieces Together Routine, the lesson plan template introduced in the second year helped the team both build storylines in a more coherent way and write lessons that were more likely to support coherence from the student point of view. We noticed that after one team member showed how the parts of a lesson could be made visible even within the "what we figured out" part of the storyline, many writers started using different colors to highlight that part of the storyline, to guide their lesson writing. A district coordinator thought the lesson template was so important, that it should play a central role in preparing teachers to implement the new units:

> I see a way to focus first on the lesson template….If I were to just bring that, from each lesson, to the table and work through it with teachers, I think teachers would

> naturally come to identify many of the shifts we outlined at the beginning of our work this week: moving from "I taught this" to "students figured this out."

From the point of view of participants in our first genetics workshop, modeling and naming different types of discussions were also important resources. For one teacher, it impressed upon her the importance of "the art of guiding discussion and thinking on your feet." Others pointed to the need for the team to develop professional development supports for the different types of discussions embedded in the units, including videos of teachers using them with students. One said, "Teachers will need more support in the specifics of different types of discussions and the protocols around those (such as consensus building, taking stock conversations)."

Discussion and Conclusion

In this final section, we offer some perspectives on how our co-design activity has supported our collective learning and supported the development of our partnership since we began this effort to co-design NGSS-aligned curriculum materials. As others have noted, "capacity begets capacity," in that we have discovered it important to build from available resources and expertise to create new capacities, and that doing so has enabled us—even compelled us, we might say—to expand the nature and range of activities we have pursued. These, in turn, have often led us to seek out new expertise or create the kinds of tools and routines we have described here to support our co-design process.

Learning with and from One Another through Co-Design

The best way to characterize the learning by expanding process from the participants' point of view is both rewarding and frustrating. On the one hand, many involved describe it as something that renews and refreshes their commitment to their work as educators and as researchers. People have appreciated, too, the requirement that good curriculum takes diverse forms of expertise to develop, and that each person has something to contribute. As one district leader put it,

> [I]t takes a village to write curriculum. This was a large shift for me and prior to my work this week I wondered why the experts didn't just sit in a room and write all of this themselves. I now have a better understanding of why the process isn't usually completed this way. The experiences with the team were invaluable and each brought his/her own individual experience and expertise to the table.

At the same time, participants found it hard to "make it up as we went along." Participants frequently referred to the need for exemplars of units, something they could work from that could serve as a blueprint. As one teacher put it, "I personally like to have an exemplar when I'm being asked to make something. It is nice to know what the product should generally look like." Whenever we presented exemplars—of assessment tasks, of a worked out section of a storyline—team members appreciated them and commented in reflections on their value to the design process. The challenge for us is that no complete exemplars existed when we started. Moreover, we discovered again and again the need for new kinds of domain and task representations to support the design process, but these were not so much exemplars as means for creating them. Members of the research team, too, are somewhat skeptical of overreliance on exemplars, since we have extensive experience of exemplars being used not as resources but as rigid templates to be followed.

Learning from Exemplars and Routines Lead to Emergence of New Goals

With each design cycle, we developed new capacities for recognizing when phenomena were likely to be engaging to students, for writing lessons that built on one another, and for supporting incremental building of explanatory models of phenomena. But with each new cycle, we uncovered also new goals or outcomes to pursue. If in the first year (2014-15) we were satisfied to produce a complete unit on time, by the second year, our design team became concerned with the consistency across lessons, ensuring that the pacing fit within the constraints of the school year, and with ensuring that the phenomenon could sustain student engagement. Capabilities developed in the second design cycle led us to focus more fully on developing sequences of lessons that would be coherent from the student point of view, as well as a lesson structure that kept students at the center. And in the third year, we became more and more focused on the need to embed supports for teaching with the materials, not just with their coherence.

In part, these new goals followed from our use of selected tools and routines we introduced to support the process. Whenever exemplars of products and enacted routines enabled someone to "see" what a finished product might look like—such as an assessment, a row in a storyline written from the student point of view, or a sequence of lessons that built on one another—a new goal came into focus for the team. But wasn't the tools and routines that we introduced to facilitate co-design per se, that led us to focus on new and more ambitious goals for ourselves, it was the opportunity we had to experience applying tools to solve a particular curriculum design problem or to experience what teaching and learning might look like in the classroom that brought new goals into focus.

These experiences—along with our observations from classrooms, not reported here—also made apparent to our team some significant gaps between the aims of the curriculum and current teaching practices. A desire to close those gaps also contributed to the emergence of new goals, especially to focus on how to support teaching with new units that have emerged in the past year.

Shifting Research Agendas Follow from Evolution of Goals and Opportunities in the Environment

The need for proposal development in a partnership is ongoing; fully one third of the project meetings included discussion of a proposal in development in the past year. We have discovered that this need, while challenging to our team, presents opportunities for the partnership to revisit research opportunities in light of what we are learning together. Proposal development provides a chance to clarify new goals that arise from what we now perceive as a "seed of possibility" and not just as a problem of practice—that is, something we believe can realize not just within a small design team but as for all students across a large school district.

The wider environment, though, figures prominently in our agenda setting. Because partnership work is heavily dependent on external funding, we have to be responsive as a partnership to the particular demands of funders at any given time. Over the past year, our partnership has expanded its efforts to build storyline-based curriculum units that integrate computational thinking practices into middle school STEM classrooms, bring storyline unit implementation to scale in the district, and develop an infrastructure for collecting data on student perceptions of coherence and relevance. Each of these efforts resulted from a successful grant proposal that framed the particular line of work we pursued. And though our own commitments to building materials that are coherent and relevant from the student point of view are reflected in these efforts, the lines of research are shaped by the funding streams that support the work.

How Co-Design Supports Our Partnership Development

A key goal of any partnership is to enhance the capabilities of participating researchers, practitioners, practice organizations, and research organizations to engage in partnership work (Henrick, Cobb, Penuel, Jackson, & Clark, 2017). One indicator that partnerships have succeeded in this aim is that team members identify with and value mutual engagement in inquiry to address problems of practice. Evidence from participant workshop reflections, the continued participation of a core of educators, leaders, and researchers, as well as the willingness to and success in pursuing additional funding for co-design within the partnership suggests that our co-design activities support this aim. The willingness to explore new participant structures and routines for coordinating work is another indicator of development along this dimension. Each year, the different routines for organizing curriculum evolved with our new capacities. A third indicator in which we have demonstrated progress is in providing "capacity-building opportunities" to members of the partnership, by providing teachers and educational leaders with opportunities to engage with experts in curriculum development, and by providing researchers with expertise in different areas with opportunities to engage in exchange with one another and with opportunities to encounter different ways of collaborating with district leaders and teachers.

As a partnership, we are concerned with other outcomes typical of research-practice partnerships, such as improving student outcomes. We are engaged in ongoing studies of students' experience of coherence and relevance of the curriculum, of teachers' capabilities with respect to recognizing three-dimensional tasks, and of student learning as part of our curriculum. We have focused here on the capabilities developed through co-design, though, because it is a leading activity of our partnership and aimed at an important goal of any partnership, to build capacity to work productively across the research-practice "divide." As a leading activity for us, it is a practice that is continuously evolving, revealing of new goals and challenges, and supportive of our individual and collective development.

References

Alexander, I. F., & Maiden, N. (2004). *Scenarios, stories, use cases*. New York, NY: Wiley.

Borman, G. D., Gamoran, A., & Bowdon, J. (2008). A randomized trial of teacher development in elementary science: First-year achievement effects. *Journal of Research on Educational Effectiveness, 1*(3), 237-264.

Bransford, J. D., Sherwood, R. D., Hasselbring, T. S., Kinzer, C. K., & Williams, S. M. (1990). Anchored instruction: Why we need it and how technology can help. In D. Nix & R. J. Spiro (Eds.), *Cognition, education, and multimedia: Exploring ideas in high technology* (pp. 115-141). Hillsdale, NJ: Erlbaum.

Brown, M. W. (2002). *Teaching by design: Understanding the intersection between teacher practice and the design of curricular innovations.* (doctoral dissertation), Northwestern University, Evanston, IL.

Brown, M. W. (2009). Toward a theory of curriculum design and use: Understanding the teacher-tool relationship. In J. T. Remillard, B. A. Herbel-Eisenman, & G. M. Lloyd (Eds.), *Mathematics teachers at work: Connecting curriculum materials and classroom instruction* (pp. 17-36). New York: Routledge.

Bruner, J. S. (1960). *The process of education*. Cambridge, MA: Harvard University Press.

Carroll, J. M. (1995). *Scenario-based design.* New York: Wiley.

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.

Couso, D. (2016). Participatory approaches to curriculum design from a design research perspective. In D. Psillos & P. Kariotoglou (Eds.), *Iterative design of teaching-learning sequences* (pp. 47-71). Dordrecht, the Netherlands: Springer.

Davis, E. A., Palincsar, A. S., Smith, P. S., Arias, A. M., & Kademian, S. M. (2017). Educative curriculum materials: Uptake, impact, and implications for research and design. *Educational Researcher, 46*(6), 293-304.

DeBarger, A. H., Penuel, W. R., Harris, C. J., & Kennedy, C. A. (2015). Building an assessment argument to design and use NGSS assessments to evaluate the efficacy of curriculum interventions. *American Journal of Evaluation, 37*(2), 174-192.

Engeström, Y. (1987). *Learning by expanding: An activity theoretical approach to developmental research*. Helsinki: Orienta-Konsultit.

Engeström, Y. (1996). Development as breaking away and opening up: A challenge to Vygotsky and Piaget. *Swiss Journal of Psychology, 55*, 126-132.

Engeström, Y., Pasanen, A., Toivainen, H., & Haavisto, V. (2005). Expansive learning as collaborative concept formation at work. In K. Yamazumi, Y. Engeström, & H. Daniels (Eds.), *New learning challenges: Going beyond the industrial age system of school and work* (pp. 47-77). Osaka, Japan: Kansai University Press.

Engeström, Y., & Sannino, A. (2010). Studies of expansive learning: Foundations, findings and future challenges. *Educational Research Review, 5*, 1-24.

Engeström, Y., Sannino, A., & Virkkunen, J. (2014). On the methodological demands of formative interventions. *Mind, Culture, and Activity, 21*(2), 118-128.

Fishman, B. J., Penuel, W. R., Allen, A.-R., Cheng, B. H., & Sabelli, N. (2013). Design-Based Implementation Research: An emerging model for transforming the relationship of research and practice. 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), 136-156.

Fortus, D., & Krajcik, J. (2012). Curriculum coherence and learning progressions. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds.), *Second International Handbook of Science Education* (pp. 783-798). Dordrecht, the Netherlands: Springer.

Fortus, D., Sutherland Adams, L. M., Krajcik, J., & Reiser, B. (2015). Assessing the role of curriculum coherence in student learning about energy. *Journal of Research in Science Teaching, 52*(5), 1408-1425.

Henrick, E. C., Cobb, P., Penuel, W. R., Jackson, K., & Clark, T. R. (2017). Assessing research-practice partnerships: Five dimensions of effectiveness. New York, NY: William T. Grant Foundation.

Lemov, D. (2010). *Teach like a champion: 49 techniques that put students on the path to college*. San Francisco, CA: Jossey-Bass.

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.

National Science Teachers Association. (2018). About the Next Generation Science Standards. Retrieved from http://ngss.nsta.org/About.aspx

NGSS Lead States. (2013). Next Generation Science Standards: For states, by states. Washington, DC: National Academies Press.

Penuel, W. R. (2014). Emerging forms of formative intervention research in education. *Mind, Culture, and Activity, 21*(2), 97-117.

Penuel, W. R., Allen, A.-R., Farrell, C., & Coburn, C. E. (2015). Conceptualizing research-practice partnerships as joint work at boundaries. *Journal for Education of Students at Risk (JESPAR), 20*(1-2), 182-197.

Penuel, W. R., Bell, P., Neill, T., Shaw, S., Hopkins, M., & Farrell, C. C. (in press). Building a Networked Improvement Community to promote equitable, coherent systems of science education. *AASA Journal of Scholarship and Practice*.

Penuel, W. R., Harris, C. J., D'Angelo, C., DeBarger, A. H., Gallagher, L. P., Kennedy, C. A., . . . Krajcik, J. S. (2015). Impact of project-based curriculum materials on student learning in science: Results of a randomized controlled trial. *Journal of Research in Science Teaching, 52*(10), 1362-1385.

Penuel, W. R., Roschelle, J., & Shechtman, N. (2007). Designing formative assessment software with teachers: An analysis of the co-design process. *Research and Practice in Technology Enhanced Learning, 2*(1), 51-74.

Reilly, R. C. (2009). Participatory case study. In A. J. Mills & G. Durepos (Eds.), *Encyclopedia of case study research* (pp. 658-660). Thousand Oaks, CA: SAGE.

Reiser, B. J. (2014, April). *Designing coherent storylines aligned with NGSS for the K-12 classroom.* Paper presented at the National Science Education Leadership Association Meeting, Boston, MA.

Reiser, B. J., Novak, M., & McGill, T. A. W. (2017). *Coherence from the students' perspective: Why the vision of the Framework for K-12 Science Education requires more than simply "combining three dimensions of science learning*. Paper presented at the Board on Science Education Workshop "Instructional Materials for the Next Generation Science Standards, Washington, DC.

Rethinam, V., Pyke, C., & Lynch, S. (2008). Using multi-level analyses to study the effectiveness of science curriculum. *Evaluation & Research in Education, 21*(1), 18-42.

Schwarz, C. V., Passmore, C., & Reiser, B. J. (2017). Moving beyond 'knowing about' science to making sense of the world. In C. Schwarz, C. Passmore, & B. J. Reiser (Eds.), *Helping students make sense of the world using next generation science and engineering practices* (pp. 3-21). Washington, DC: NSTA.

Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M., & Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others? *Teaching and Teacher Education, 27*(2), 259-267.

Severance, S., Penuel, W. R., Sumner, T., & Leary, H. (2016). Organizing for teacher agency in curriculum design. *Journal of the Learning Sciences, 25*(4), 531-564.

Takaya, K. (2008). Jerome Bruner's theory of education: From early Bruner to later Bruner. *Interchange, 39*(1), 1-19.

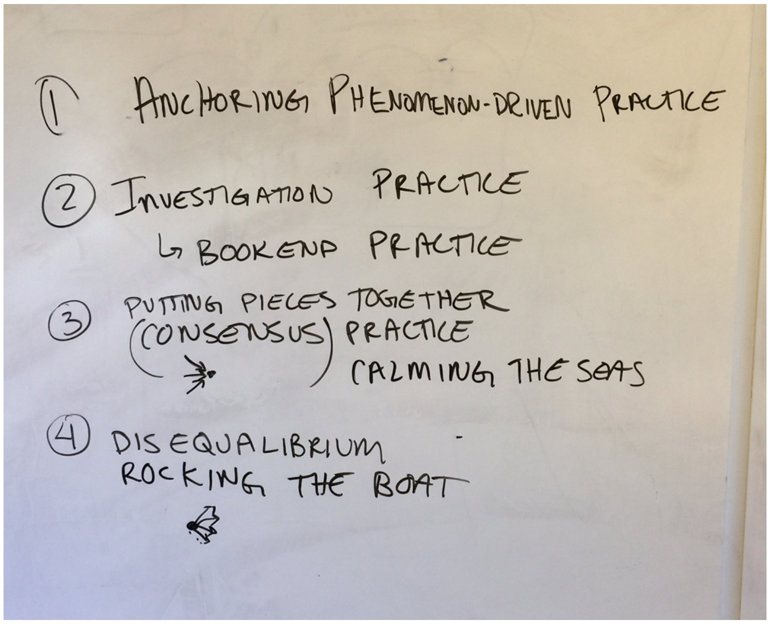
Taylor, J. A., Roth, K. J., Wilson, C. D., & Stuhlsatz, M. A. M. (2017). The effect of an analysis-of-practice, videocase-based, teacher professional development program on elementary students' science achievement. *Journal of Research on Educational Effectiveness, 10*(2), 241-271.

Voogt, J. M., Laferrière, T., Breuleux, A., Itow, R. C., Hickey, D. T., & McKenney, S. E. (2015). Collaborative design as a form of professional development. *Instructional Science, 43*(2), 259-282.

Weizman, A., Shwartz, Y., & Fortus, D. (2008). The driving question board: a visual organizer for project-based science. *The Science Teacher, 75*(8), 33-37.

Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

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


① ANCHORING PHENOMENON-DRIVEN PRACTICE
② INVESTIGATION PRACTICE
↳ BOOKEND PRACTICE
③ PUTTING PIECES TOGETHER
(CONSENSUS) PRACTICE
CALMING THE SEAS
④ DISEQUALIBRIUM
ROCKING THE BOAT