Organizing Development and Adaptation Studies for Relevance and Impact

William R. Penuel, Carol Pazera, Zoë Buck Bracey, Daniel C. Edelson Kate Henson, Brian J. Reiser, Tamara Sumner, Nicole Vick, James Ryan Douglas Watkins

Abstract

Calls for research to be more relevant to practice persist, despite extensive efforts over the past two decades, because most research efforts remain unresponsive to the concerns of decision makers and educators and fail to anticipate the diverse goals and contexts of practice. We present five heuristics for organizing development and adaptation studies to improve relevance and impact in the context of a project that developed and tested a set of free, standards-aligned instructional materials in high school science with associated professional development. The five heuristics are: (1) be adaptive while also attending to history; (2) anchor design activities in a vision for equitable disciplinary teaching and learning; (3) maintain continuous attunement to interest holders' concerns; (4) design for productive adaptation; and (5) develop evidence of changed relations and of multiple potential future uses of the products and findings from the research and development effort. We argue that more studies are needed that adhere to these heuristics and we consider conditions that would need to be in place for other kinds of development and adaptation studies to succeed. We conclude by articulating the kinds of research questions that might follow design and adaptation projects such as ours.

Introduction

An enduring theme in education research is the call to make it more relevant to practice. In the early 2000s, a broad coalition of scholars, advocates, and policy makers argued that to be relevant to policy and practice, research needed to use research designs that could accurately estimate the impacts of potential interventions (Coalition for Evidence-based Policy, 2002; National Research Council, 2002; Whitehurst, 2003). Policies and research funding justified through this line of argument have been successful in generating evidence from experimental and quasi-experimental designs for a wide range of interventions, and they have expanded the capacity of the field to conduct large-scale field trials of interventions (Edovald & Nevill, 2020; National Academies of Sciences Engineering and Medicine, 2022).

Doing more research on the impacts of interventions, however, has not made education research more relevant to policy and practice. The kinds of field-initiated studies that agencies like the Institute of Education Sciences (IES) in the United States fund do not directly address many of the ongoing concerns of teachers (Schneider, 2018; Willingham & Daniel, 2021). Further, many of the current Development and Innovation studies funded

by IES maintain a high degree of researcher control and are conducted in a narrow range of settings, with little opportunity to discover the kinds of problems that emerge once innovations are implemented in more heterogeneous contexts (Farley-Ripple et al., 2018; National Academies of Sciences Engineering and Medicine, 2022). The emphasis on researcher control and "doing what works," moreover, has greatly restricted the scope of practitioners' say in decision making regarding the design and development of interventions, wherever these policies have been promoted (Biesta, 2007).

This paper makes the argument that we need to organize a higher percentage of research and development projects in ways that allow for research to be in a deep sense "responsive to this moment" (National Academies of Sciences Engineering and Medicine, 2022, p. 160) in history. Our argument is aimed at designers and scholars seeking to develop innovations that can achieve broad reach across heterogeneous school contexts. Research that is responsive to the moment is necessarily interdisciplinary, collaborative, and focused on changing systems in ways that make them more equitable and just (Spencer Foundation, 2023). Further, it is fundamentally about changing relations—among people, organizations, and institutions—both as a means of promoting systems change and as an end unto itself (Bang & Vossoughi, 2016).

Synchronization as Key to Relevance and Impact

Making research relevant to practice requires a deeper *synchronization* in time and place between research and practice. That is to say, becoming more relevant will require researchers to develop practices that allow them to attune in an ongoing way to the people who are involved in, and hold interest in, the outcomes of research before, during, and after research, that is, in the future (Akkerman et al., 2021). Further, those intervening in systems need to collaborate with others to anticipate the potential futures – of the research, of schools, and of the relationships of schools to broader social, political, and economic systems – toward which the research might build or contribute (Philip & Sengupta, 2021; Stilgoe et al., 2013). This anticipation involves consideration of both potential benefits and harms to individuals and groups furthest from power, as well as those to whom as a nation we owe an "education debt" (Ladson-Billings, 2006). It also requires anticipation of a wide range of productive adaptations of interventions in varied, contested spaces of implementation (Gutiérrez & Penuel, 2014; Penuel & Fishman, 2012).

Just how such synchronization can be accomplished in research is only beginning to be documented in educational research. In part, this is because current research and development infrastructures are organized around the goal of developing evidence for potential impact through a linear, research-to-practice pipeline (Peurach, 2016; Thomas et al., 2024). At the same time, a family of collaborative research approaches has emerged alongside—and funded partly through—the current infrastructure, and this family of approaches has yielded a number of examples of research that is both relevant and impactful (Penuel, Riedy, et al., 2020). Further, a recent National Academies of Sciences, Engineering, and Medicine (2022) report called for a new type of research named "Development and Adaptation" studies. This calls for studies that develop interventions that are responsive to real needs and that create high-leverage, multicomponent strategies to address inequities within and across schools. Such studies, the report argued, should also investigate the ways educators adapt interventions in a wide variety of contexts, recognizing that adaptable interventions are "more likely to be adopted, supported, and sustained, thus improving educational outcomes" (p. 79).

In this paper, we present heuristics for planning and carrying out these kinds of development and adaptation studies, illustrating these heuristics within a single project. The heuristics are based on three bodies of existing research. We developed the heuristics first by synthesizing evidence and recommendations from three key sources: (1) research on how education leaders perceive, access, and use research in practice; (2) research and evaluation of research-practice partnerships in education; and (3) research on supporting the design and enactment of equitable STEM learning at the scale of a school district. As we present the heuristics, we present key ideas from these areas of research. In selecting from among the many possible ideas to put forth as heuristics, we have focused on elaborating ideas that are largely missing from existing guidance developed for design and development studies in education (e.g., IES and NSF, 2013).

Though the terminology we use to describe the project is centered on a U.S. context, there are similar kinds of opportunities for funding for development and adaptation studies outside the U.S, as well as a similar focus on incentivizing research that can inform practice. In the United Kingdom, for example, the Education Endowment Foundation (EEF), established in 2011, funds both evaluation studies of interventions and syntheses of evidence related to interventions for use by educators. Some of the projects funded by Horizon Europe, a funding program for research and innovation, include several projects focused on designing and testing innovations in education. In the Netherlands, the National Regieorgaan Onderwijsonderzoek has a competitive grant program focused on "research into promising approaches and innovations" in education; several funded studies use design-based approaches to research, such as those used in the OpenSciEd project. In addition, improving the use of research in policy and practice is a priority globally. In the UK, for example, the Economic and Social Research Council (ESRC) Education Research Programme has funded projects focused on exploring "new ways of working in partnership across the boundaries between education research, policy and practice" (OECD, 2025, p. 71). At the same time, in countries beyond the US, there is also little consistent or clear guidance about how best to synchronize research, policy, and practice (van Atteveldt et al., 2019).

The OpenSciEd High School Development Project

OpenSciEd High School is a development and adaptation project focused on improving student outcomes for all students in science. OpenSciEd High School consists of instructional materials intended to address high school standards for life, physical, and Earth and space sciences articulated in the Next Generation Science Standards (NGSS; NGSS Lead States, 2013). The project to develop the materials was grounded in research-based ideas for how best to support student learning in science, building on a storyline instructional model (Reiser et al., 2021). This model organizes instructional materials around the pursuit of questions that students identify when presented with a phenomenon from the natural world or with a problem that calls for engineering design. In the storyline model, student questions motivate and guide collective sensemaking activities that lead students to develop explanatory models of phenomena or solutions to problems (Penuel, Reiser, et al., 2022). Such classrooms are organized around students' figuring out key science ideas with guidance from their teachers, rather than simply learning definitions and facts (Schwarz et al., 2017).
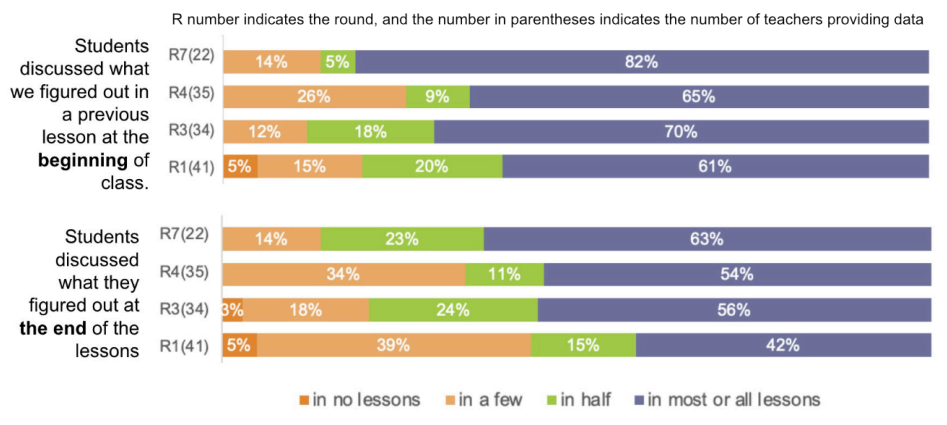
The project was led by a consortium of developers from four different institutions which was selected through a competitive process. The development team included science educators, researchers, and former classroom teachers from across the country. In addition, the consortium included a school district partner to support more intensive collaboration and a group dedicated to promoting education, social justice, and diversity in science education. The consortium was contracted by a nonprofit organization to carry out the project, which provided overall guidance and was responsible for distributing materials and certifying professional development providers upon project completion. A consortium of private foundations funded the initiative, a step that would likely be necessary for future projects of this size, given the limitations on individual project amounts for both federal and foundation grants.

A State Steering Committee for the project provided input on the design of materials at regular intervals, and it was responsible for recruiting teachers to participate in a field test of the materials. Members of the committee were either state education agency staff or their delegates. The ten states represented all regions of the United States: they included states with large urban centers and rural areas, as well as states with more conservative populations and states with more liberal politics.
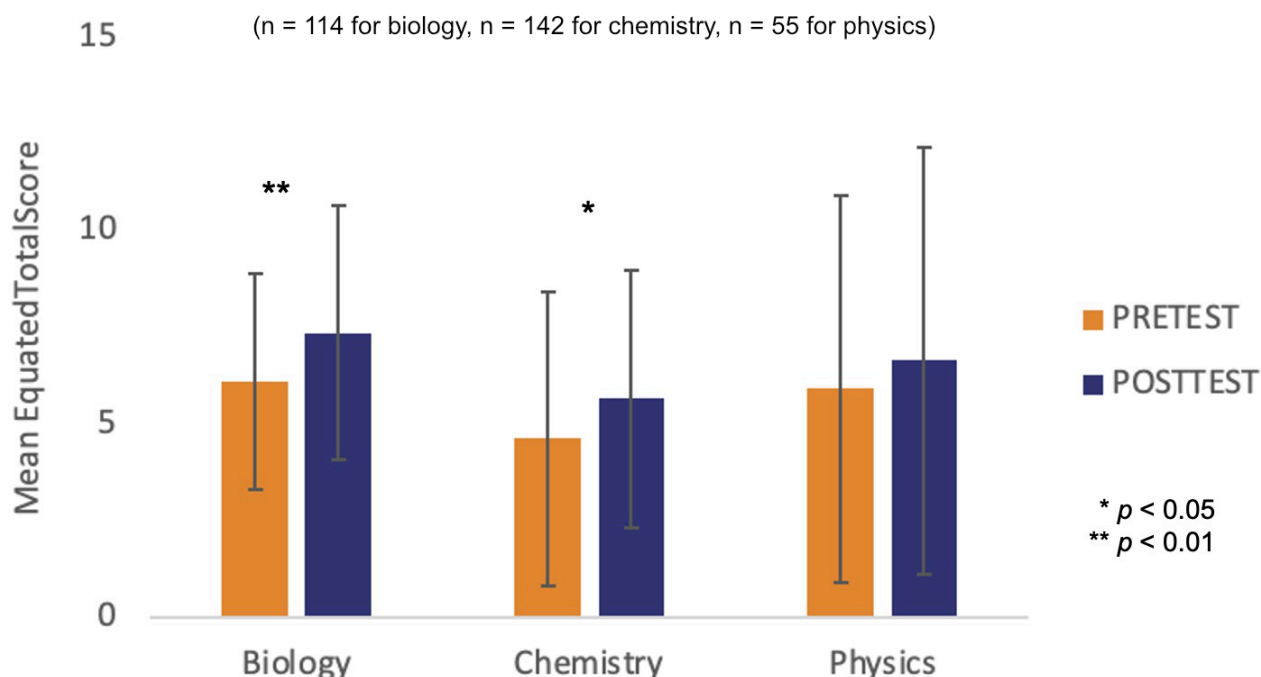
According to the common guidelines first articulated more than a decade ago by the Institute of Education Sciences and National Science Foundation (2013) for design and development studies, the project has been a success. For one, the project developed a complete intervention, including all materials needed for implementation. The OpenSciEd materials that the consortium developed are freely available. By making the materials free, the consortium intended to improve access to high-quality instructional materials and transform relationships among developers, publishers, and school districts that adopt materials. The Creative Commons license for high school allows any school district or teacher to use the materials freely and to adapt them as they see fit. The license specifies that commercial textbook companies may adapt, distribute, and sell the materials, but they must pay a fee to the nonprofit organization, OpenSciEd. The nonprofit maintains the materials and contracts with 1) providers of kits associated with units, and 2) a network of certified providers of professional learning that schools and districts can hire to prepare educators to use the materials. The result is a large distribution and support system, in which anyone can obtain a version of the materials for free, but where an ecology of publishers can sell derivative versions with their own enhancements to schools and districts.

Consistent with U.S.-based guidelines for intervention studies at this stage of development, the project developed evidence of feasibility of implementation and promise for improving student outcomes. In their responses to survey questions, teachers indicated that they able to implement the key instructional routines of the storyline model, such as connecting the day's lesson purpose, activities, and learning to that of the larger arc of the unit. Additionally, their facility with routines increased over time, that is, from early rounds of testing to later rounds (see Figure 1). Second, a test of students' ability to apply knowledge developed in units showed significant gains in both biology (n = 57) and chemistry (n = 82); in physics, where the sample size was smaller (n = 22), the trends were positive but not statistically significant (Figure 2).

Figure 1 – Teachers' use of navigation routine across rounds over time

R number indicates the round, and the number in parentheses indicates the number of teachers providing data

Students discussed what we figured out in a previous lesson at the beginning of class.

	in no lessons	in a few	in half	in most or all lessons
R7(22)		14%	5%	82%
R4(35)		26%	9%	65%
R3(34)		12%	18%	70%
R1(41)	5%	15%	20%	61%

Students discussed what they figured out at the end of the lessons

	in no lessons	in a few	in half	in most or all lessons
R7(22)		14%	23%	63%
R4(35)		34%	11%	54%
R3(34)	3%	18%	24%	56%
R1(41)	5%	39%	15%	42%

■ in no lessons ■ in a few ■ in half ■ in most or all lessons

Figure 2 – Mean (with SD error bars) pre and post-test scores

(n = 114 for biology, n = 142 for chemistry, n = 55 for physics)

PRETEST
POSTTEST

* $p < 0.05$
** $p < 0.01$

One motivation for developing this article has been the fact that the project's organization does not reflect a typical development and innovation project in science or any other subject area. Instead of being developed by a small team of researchers and tested in a handful of classrooms, this project involved a large group of people that included education researchers, educators, content experts in science, state education leaders, and professional development providers. Instead of focusing on testing in a few classrooms, the field test reached 20,000 students in 300 classrooms. Further, after gathering evidence of feasibility of implementation and promise, thousands of educators have registered to download the materials, as an indicator of its reach. In many ways, the project's impact—in terms of its reach—has already been felt. But, we argue below, that is because we have followed a set of heuristics for supporting *synchronization* between research and practice, and thereby intervened in a way so as to effectively promote the relevance of the project.

Heuristic 1: Attend to the Historical Moment While Adapting to Changing Circumstances

What we mean by this heuristic is that project teams need to be particularly attuned to the historical circumstances in which they find themselves and to recognize that such circumstances evolve in ways that require adaptation of initial research plans and compromises among partners. As a historical event, the OpenSciEd initiative began four years after the publication of the Next Generation Science Standards (NGSS; NGSS Lead States, 2013). By the time an initial consortium of developers had been selected through a competitive process led by the Carnegie Corporation of New York, a private foundation, many states had adopted the NGSS or standards based on them. Yet there were few instructional materials aligned to the new standards, and many in the field were clamoring for materials that could be used to propel changes to practice (National Academies of Sciences Engineering and Medicine, 2018). By the time developers were selected to develop high school materials, seven years had passed. Our project design reflected the sense of urgency among members of the State Steering Committee as leaders for standards implementation in their states, and we were given a contract to develop, test, and revise three full-year courses that addressed all NGSS standards in high school within a three-year time period. This was possible only because many members of the consortium had worked together in the past—and with the State Steering Committee members as well—and because an existing set of materials for two of the courses developed by two consortium members could be revised as part of the project.

As a consortium, one way we attended to history while also adapting to the present was in how we selected phenomena to anchor units of study. A key goal in selecting anchors was to consider those groups of students who have historically been excluded from science and engineering and whose interests, experiences, and concerns are rarely included in U.S. curriculum materials. So to select those anchors, we gathered survey data from students about lists of phenomena connected to specific standards, and on each survey, we gathered information (anonymously) about their race/ethnicity, gender, home language, and ZIP code. The samples included students from ZIP Codes across the country, in urban, suburban, and rural areas. In selecting phenomena to anchor units, we sought to give preference to students from systemically marginalized communities to promote equity.

We could not simply select any phenomena for students to rate, however, given the complicated circumstances politically that emerged during the period of our project. A "conflict campaign" began that resulted in several states passing laws restricting teaching about subjects that addressed matters of racial equity and gender and sexual identity (Pollock & Rogers, 2022). This required the team to be attentive to new political divides existing across states in our field test. Some phenomena we were interested in pursuing to anchor were not politically viable, and we were not able to gather data on them to know how interested students might be in them.

Our approach was to adapt to the political moment in ways that still supported teachers in engaging students in thinking about how human social, political, and economic systems shape phenomena in science and problems in engineering. The team emphasized gathering multiple perspectives and putting students in the driver's seat when it came to proposing solutions to problems. The team also had to compromise with states about specific language used in the materials, to ensure that access to them would not be unnecessarily restricted by new state laws. We came to an agreement on those

compromises through a variety of means: through conversations among developers, through feedback from State Steering Committee members, and by reviewing feedback forms from teachers. The ability to achieve these compromises underscores the importance of relying in concert on both Heuristic 3 (maintaining continuous attunement to interest holders' concerns) and Heuristic 2, to which we turn next.

Heuristic 2: Anchor Design Activities in a Vision for Equitable Teaching and Learning

Our design activities were anchored in a particular vision for science that was centered on equity. By *vision,* we mean an image of what teaching and learning ought to look like, that is, a normative characterization of how students relate to each other, to the teacher, and to tasks in the classroom. A vision is grounded both in research and in a set of values. Our team's values center on a vision of ensuring that all students have an opportunity to experience science and engineering as relevant to their lives and as contributing to the flourishing of their communities (National Research Council, 2012). That is what we mean by equity. Equity also demands that we consider for whom we are designing, paying particular attention to communities that have been excluded from science and engineering in the past, to ensure that they can see their own interests and concerns reflected in instructional materials (National Academies of Science, Engineering, and Medicine, 2024).

To promote equity in both the design and implementation of materials, our team followed several equity guidelines informed by research and by the design specifications of OpenSciEd, which were developed by a different group of stakeholders primarily from the education research community. The design specifications for OpenSciEd emphasize the value of recognizing multiple ways of knowing and being, as well as considering how science and engineering are implicated in matters of justice related to race, socioeconomic class, gender, educational sovereignty, Indigenous rights, immigration history, land and water rights, sexual orientation, gender expression, abilities, and other dimensions of social difference related to justice (OpenSciEd, 2019).

We used several strategies to address these guidelines. For one, we involved diverse groups of stakeholders in the design of materials (Bang & Vossoughi, 2016). That meant constructing teams that were diverse with respect to race, ethnicity, gender, and sexual identity. In addition, we consulted regularly with community members and elders in places where units were anchored to embody within the context of disciplinary learning an appreciation for heterogeneity in ways of knowing and being (Warren et al., 2020). For example, in a unit where students investigate what causes lightning and why some places are safer than others when it strikes (the structure and properties of matter), the opening lesson invites students to engage with different lightning stories involving engineers and storytellers from many different traditions as sources for learning about lightning. These include a reading about engineer Hertha Ayrton, known for her work on electric arc lamps, and a video in which the chemistry team's collaborator, Toyin, shares Oyo oral history about the storm deities Ṣango, Ọya, Oṣun, and Ọba. Another reading featuring a Navajo historian describing Navajo knowledge about lightning supporting life is returned to again in a later lesson as students consider lightning's effects, including the production of atmospheric nitrogen.

Another example of how we addressed these guidelines comes from the physics course. In a unit where students explore gravity, orbits, and meteors, students spend some time investigating the meteor that most likely led to the extinction of the dinosaurs. The resulting crater is discernible through a pattern of water-filled caverns on the Yucatan peninsula, which were historically, and continue to be, sacred to the Mayan people. The physics team worked with Professor of Chicana/o Studies Gerardo Aldana, who speaks Mayan and has spent extensive time in modern Mayan communities, to better understand the Mayan people's relationship to these caverns and communicate about that relationship respectfully in the materials.

In addition, the developers consortium was structured to be answerable to two different groups, in terms of adhering to a vision for equitable teaching and learning. One group was an independent group that contracts with anonymous reviewers to provide both formative feedback and summative evaluation of units using a set of guidelines jointly developed by them and another independent curriculum materials review group (NextGen Science & EdReports, 2021), another independent group that evaluates curriculum materials. For state- and district-level decision makers, the evaluations of these groups figure in what materials are selected for lists of high-quality instructional materials (Doan et al., 2022). A second group to which the consortium was answerable was a non-profit organization committed to researching, building, and sustaining transformative educational opportunities in science with all students. The consortium set up a grant program that this organization administered, which supported the involvement of educators and researchers in both reviewing units and contributing to unit design. Their participation was key to ensuring a focus on the equity specifications.

To ensure that the equity specifications were front-and-center within the ongoing work of writing teams, we held regular project retreats and meetings where equity and justice were a focus. In some of these retreats, we reviewed the equity design specifications together and discussed how they were being embodied in both instructional and professional learning materials. We also had presentations and discussions with leading scholars of equity in science education, some of whom also consulted directly with the project. In addition, an in-person mid-project writers retreat focused on equity and justice as a central theme. At this retreat, we catalogued the different strategies and approaches we were using within each course and shared them with each other to make visible across teams the different strategies each course was using to address the design specifications related to equity. We also spent time together reflecting on and synthesizing our approaches to promoting equity and justice, writing about them for a practitioner audience (Penuel, Henson et al., 2024).

An ongoing challenge for us as a team was addressing two different meanings of equity that are emphasized within science education, namely embracing heterogeneity and learning and using science and engineering to promote justice. Briefly, these framings of equity emphasize diversity and divergence in classrooms (embracing heterogeneity) and engaging students directly in remedying injustices experienced in their communities (promoting justice) (National Academies of Sciences Engineering and Medicine, 2024). Our instructional model emphasizes students working toward consensus together on explanatory models of phenomena, even as what students bring from their own experiences and identities are valued (Reiser et al., 2021). But there can also be value in seeking to name differences and to organize science and engineering learning around "heterogeneity-seeking" rather than consensus-building (Morales-Doyle, 2024; Pierson et al., 2023). And while we present students with many opportunities in materials to grapple with injustices, we did not encourage students to take specific action to address those

injustices, [1] nor did we incorporate many opportunities to engage with local issues. In seeking to be useful across the country, our materials do not in themselves address the need for materials to be localized; we take this point up in the context of discussing Heuristic 3, support for adaptation. The need to avoid encouraging students to take up stances within movements toward justice has much to do with the heterogeneity of our interest holders' concerns for developing materials that could be used in states across a wide political spectrum, as discussed above under Heuristic 1. The strategies we used to attune to that heterogeneity is what we take up next in discussing Heuristic 3.

Heuristic 3: Maintain Continuous Attunement to Interest Holders' Concerns

A key goal of design and development research as defined by federal grant makers is that this early phase work should produce measures of good technical quality for assessing implementation and student learning outcomes that can be used in subsequent studies, as well as pilot data that provides evidence of the intervention's promise in improving student outcomes. As part of OpenSciEd, we did just that: we developed a range of measures for studying teachers' practice and teachers' learning from professional development (for published articles related to the middle school field test, on which measures for high school were based, see Krumm et al., 2020; Penuel et al., 2023; Penuel, Krumm, et al., 2024). In addition, we followed an evidence-centered design process (as described in Penuel et al., 2019) to develop phenomenon-based tasks to measure students' growth in their ability to use science ideas, practices, and crosscutting concepts targeted in units. The results of those studies showed the promise of units in promoting student learning, as noted above.

But if that were all the evidence gathered as part of our study, we would have little evidence to support the claim that our endeavor was relevant to and synchronized with the activities and concerns of key interest holders in the project. A key requirement for promoting relevance is that evidence gathering within design and development studies should systematize an "ongoing process of engaging, listening, and responding with the willingness to listen and respond again" (Akkerman et al., 2021, p. 421). Further, within projects, the relations between researchers and participants should be organized as much as possible as *subject-to-subject* relationships, that is, as people interacting with purposes, not just research subjects or objects of research (Akkerman et al., 2021). In this project, we prioritized the concerns of three groups of interest holders: educators, state education leaders, and students.

One of the ways that we engaged *educators* throughout the project was through the co-revision of units. The team was committed to engaging practicing teachers as designers of these materials, because past research has shown co-design builds commitment and understanding of key design principles and because it supports teachers' own agency (Severance et al., 2016; Voogt et al., 2015). It also helps researchers learn: by engaging in co-design, researchers develop a sensitivity to problems of practice that arise in implementation, along with practical design principles for balancing competing design goals (Goldman et al., 2022; Manz et al., 2022). But co-design is time-intensive, and the considerable constraints placed on the consortium with respect to time and standards made it impossible to replicate an approach used to develop initial versions of two courses. Thus, the team adapted its process to focus on inviting teachers from the field test to be

involved in co-revision meetings, held online, that were aimed at addressing feedback teachers had provided when first implementing units and feedback from external reviewers. Co-revision did give teachers less say in the overall design, but it provided space for educators to see that developers were being responsive to their concerns and to influence how specific concerns would be addressed.

Educators who served as field test teachers also had multiple opportunities to provide feedback on the structure and content of the professional development and curriculum materials. Prior to teaching a field test unit, teachers participated in a professional development workshop. Following their participation, teacher feedback related to both the content and organization of the workshop was collected via surveys. During the field test of each unit, field test researchers gathered data from teachers on implementation and experience via survey and interview. Teachers also provided direct feedback via a form monitored by the unit writers. All of this feedback informed the revision of the final public facing materials

We also stayed attuned to *state education leaders* through regular meetings with them. There were two types of meetings: monthly meetings with all leaders, and design meetings between unit leads and state leaders. At monthly meetings, the developers consortium presented on and solicited feedback on multiple key aspects of design: the course scope and sequence, the professional development goals and sequence, Earth and space science integration, how to address emergent issues related to the pandemic, changing political environments, and teachers' responses to the materials and professional learning activities. In this way, state leaders were design consultants and were full partners in implementing professional learning, in that they were responsible for recruiting and supporting a cadre of local leaders for workshops who attended facilitator workshops led by developers with expertise in professional learning. In addition, data from the field test were presented at multiple meetings, and state leaders were given opportunities to make sense of the data and to make recommendations to developers about how to address concerns identified in the field test. As part of these sessions, state leaders followed a routine of first listening to a presentation, then discussing their own noticings and interpretations of field test data in small groups, which they then brought back to developers with concrete recommendations for change. Unit leads presented to state team members information on unit-specific plans, as well as feedback from teachers and planned revisions. In these meetings, they solicited input from state leaders. In some cases, the input resulted in major shifts to units. For example, based on unpacking of standards, the physics unit 1 team originally chose an anchoring phenomenon about a young boy in Malawi named William who brought electricity to his village by building a windmill. Despite interest expressed on student surveys, the State Steering Committee felt that this phenomenon was too far from the experiences of American students to anchor a whole unit and instead suggested that we look to the then-recent widespread power outages in Texas. The team pivoted to this phenomenon, and the field test showed that it was a success, capturing the interest of students and driving learning toward the chosen standards.

Student interest and experience were attended to in several ways. First, as described above, we surveyed students recruited both from field test classrooms and via the internet about the problems and phenomena they would find most interesting. The problems and phenomena included in the survey were generated by the unit development team following a deep analysis of the standards. These data informed the selection of the anchoring

phenomenon as well as lesson level and assessment phenomena. During the unit field test, student experience data was collected via regular exit tickets and surveys. Student work samples were also collected and analyzed by the unit writing team to evaluate the alignment of student learning goals and outcomes in real time.

An example from the chemistry course illustrates how one team made use of feedback from different sources to revise units in ways that stayed attuned to different stakeholders' concerns, particularly related to ensuring relevance to students. The focal unit, Thermodynamics in Earth Systems, is organized around the question, "How can we slow the flow of energy on Earth to protect vulnerable coastal communities?" The first issue the team tackled was that teachers said they rarely revisited questions generated in the first lesson, which were questions intended to drive the learning forward and help students see how the day's learning connected to something they cared about (Weizman et al., 2008). The team addressed this feedback by building in three places in the unit where teachers were invited to revisit student questions explicitly, so that students had the chance to see how their questions would be answered. Field-test teachers also reported that while student interest in the unit was high at first, it waned across the course of the unit. Teachers recommended more images to dramatize the immediate effects of glacial melting, as well as more explicit connections between glacier melt and effects of climate change felt around the world. The team also incorporated multiple prompts for student thinking from the Ethical Deliberation and Decision-Making Framework from the Learning in Places project (Learning in Places Collaborative, 2022) throughout the unit, inviting students to consider how both the human world and other species were impacted by polar ice melt and sea level rise. For example, a lesson that previously focused solely on data analysis around glacial melt in Greenland now incorporates Inuit voices and asks students to consider how Inuit hunters and fishers and NASA scientists define systems and the problem in different, but complementary, ways.

We did not involve the key interest holders of students and families directly in design. Interest surveys were an indirect means of getting their input into units, but we recognized that these provided limited insight into why some phenomena were interesting to some students but not others. Our own team's limited experience with involving students and families in design made it difficult for us to see how to do so within the resource and time constraints of the project. At the same time, following models provided by the Learning in Places project, we did incorporate into both field test and final materials multiple opportunities for "home learning" engagements in which students solicited input from family and community members on the problems and phenomena they were studying, as well as potential design solutions to them.

Heuristic 4: Design for Productive Adaptation

A goal of many design and development projects is to create programs or practices that can be implemented with fidelity in a wide range of contexts, under typical conditions of schools and districts. Such a goal is grounded in an appreciation for how effects of interventions are often mediated by implementation quality (Zvoch, 2012) but also with the belief that treatments should work anywhere, regardless of context, under the routine conditions of schools (Conaway et al., 2022). But approaching development and adaptation projects with the goal of producing something that can work anywhere leaves no room for educator agency and localization. Further, the requirement that context be

treated as something that is irrelevant to design makes heterogeneity—in implementation and in outcomes—a problem to be solved rather than a resource for making interventions more relevant to local actors. Often, intervention requires *changing* the conditions of schools and districts to accommodate ambitious and equitable teaching and learning goals (Penuel, 2019).

The goal of synchronization (attuning continuously to interest holders' goals and concerns) in development and adaptation studies demands that researchers assume that implementers bring a wide variety of goals for implementation, and that implementation contexts are varied and often challenging. First, designers should anticipate the wide variety of goals educators bring and the variety of contexts in which they work (Björgvinsson et al., 2012; Escobar, 2017). Second, implementing teachers regularly experience discontinuities that need to be anticipated between administrators and teachers, for example, in terms of their goals and systems for evaluation and guidance (Elmore & Forman, 2011; Hannan et al., 2015; Penuel, 2019). Third, implementing teachers experience discontinuities between families and schools, with respect to how home knowledge is valued for science learning (e.g., Ishimaru & Bang, 2022) and between what is expected regarding performance, grading, and feedback and accommodations that are necessary in classrooms characterized by neurodiversity. The degree to which people are able to anticipate these pluralities and discontinuities is likely to depend both on the composition of the design team, as well as situational and institutional constraints on design goals (Le Dantec & DiSalvo, 2013; Penuel, Allen, et al., 2022).

The OpenSciEd High School Developers Consortium anticipated several goals high school teachers might bring and the variety of their contexts, based on our earlier curriculum development work. For example, some educators are motivated strongly by goals typically associated with place-based education for ensuring that science instruction focuses on phenomena that are local; others are concerned with ensuring their students are prepared for demanding college courses. At the same time, we also knew that, broadly speaking, teachers' visions for science teaching and learning were most well-aligned to the idea of science as being both knowledge and practice and that science should connect to students' interests and experience, even if teachers' ideas about equity were less well-aligned to that of the *Framework* (Penuel, Bell, et al., 2020). Further, we anticipated that to meet these goals, teachers were more inclined to rely on materials that they developed themselves than the coherent units of instruction that we were developing (Doan et al., 2023).

Our professional learning workshops with teachers reflected our understanding of these diverse goals that teachers brought to field testing new instructional materials. In anticipation of teachers' reluctance to use materials "as is," we outlined ways that teachers might adapt field test materials in ways that still maintained integrity to the design specifications. In addition, we sought to build appreciation for how the unit design process, the focal phenomena, and activities of units could facilitate students seeing how science and engineering could be relevant to their everyday lives and to their communities. In both units and in professional learning, we gave strong emphasis to building equitable learning cultures and to recognizing heterogeneous ways of making sense of phenomena, to help develop teachers' visions for equity in science teaching and learning. We supported teachers in developing and revisiting a set of "community agreements" in their classroom for fostering respect, equity, and a shared commitment to collaborative knowledge building (Affolter et al., 2022). For example, designers of the first chemistry unit for the year incorporated intentional moments to return to community agreements into every lesson in the first lesson set, as well as at other key moments in the unit. In the fifth lesson,

students explicitly consider how they can work throughout class to be "Committed to Our Community." They also discuss as a class which community agreements might be most important to keep in mind during a consensus decision and take time at the end of class to reflect on how community agreements could support the investigations needed to answer students' key questions.

With respect to addressing discontinuities teachers encountered, we gathered data from educators on their persistent challenges, which helped state leaders and facilitators understand the emergent discontinuities teachers experienced. One such perceived discontinuity among teachers was between meeting the needs of students with disabilities in the curriculum materials and the goal of supporting students in making sense of phenomena using the three dimensions emphasized in the NGSS. Through interviews conducted both as part of the field test data collection and with designers of professional learning materials, we surfaced different understandings of the problem, as well as strategies that some teachers were using to adapt materials to meet requirements involving students' Individualized Education Plans (IEPs). To address these concerns, the team developed a two-day professional learning workshop that engaged educators in using the framework of Universal Design for Learning (UDL; Rose & Meyer, 2002) to identify ways to adapt materials to meet their specific students' IEP requirements. The professional learning also highlighted ways materials were designed with this framework in mind. For revised units, we built supports based on research that address the kinds of supports that benefit students with different kinds of disabilities in meeting ambitious disciplinary learning goals in science (Palincsar et al., 2001; Therrian et al., 2011). For example, the materials in all three courses suggest ways in which teachers can encourage students to communicate their ideas through gesture, drawing, and kinesthetic activity, in addition to orally, during classroom discussions. Alternative activities are suggested when students are required to walk or run, particularly outside. We also provide extensive support for students' social interactions, in the form of community agreements that are jointly negotiated, as elaborated above.

One challenge relates to how to support productive adaptation of materials beyond our project, particularly related to the ongoing press for materials to address local phenomena and problems, including those that pertain to socio-ecological justice. To that end, members of our team worked with state leaders and researchers on two related projects that were focused on localization. The team developed a set of resources for teachers to use related to five different strategies for localizing OpenSciEd materials, highlighting differences in the time and capacity needed for each. The most demanding—in terms of capacity and time—is to add or swap an anchoring phenomenon. Such an adaptation requires rewriting all the lessons that follow, for them to be coherent from the student point of view. Somewhat less demanding is the strategy of adding or swapping out an investigative phenomenon, which involves writing in a new lesson at a specific point in the storyline. Teachers can also swap out a transfer task that could be used as an assessment that ties to a local phenomenon or issue. Both these strategies still rely on the pedagogical design capacity of teachers for developing tasks that both elicit students' use of the three dimensions of science and that connect to students' interests, experiences, and identities. Other strategies rely less on teachers' design of new materials, and more on how they respond to students' questions and interest with existing materials. One of these adaptation strategies entails making use of phenomena that students identify at the beginning of each unit that they have experienced and that they think can help them explain the focal phenomenon of the unit. A second involves making use of student exit tickets to invite students to make connections between the day's lesson and something in their everyday lives or that they care about (Hulleman & Harackiewicz, 2009).

Additional challenges to fostering productive adaptation fall beyond the scope of our immediate project but depend on local actors leveraging resources and expertise within local networks developed through the project (Heuristic 5 below). These include challenges that are well-documented in the literature, such as teacher access to sustained professional learning opportunities and incoherent guidance from other components of systems such as assessments and teacher evaluation frameworks (Cobb et al., 2018; Penuel, 2019; Rorrer et al., 2008; Stosich et al., 2017). The challenge of providing access to sustained professional learning opportunities is particularly acute at this moment, since the post-pandemic funds many states used to support teachers in the past few years of field testing are no longer available.

Heuristic 5: Developing Evidence for Changed Relations and Possible Future Uses

As noted above, our development and adaptation project developed evidence related to feasibility of implementation and promise of OpenSciEd. However, for research to be relevant, evidence of generativity is needed, namely evidence that the project has yielded products which can support multiple possible future uses and which show ways that those uses could transform relations within the classroom and beyond. We anticipated several potential uses of research tools and evidence for which we designed: political uses to support adoption decisions, uses to inform adoption and piloting, and uses by other scholars studying OpenSciEd. To support these different uses, we made all instruments used in the field test Open Educational Resources, just as we did the materials. We also considered both the kind of evidence needed and formats for visualization and sharing that could readily be used by state and local leaders in multiple contexts.

Political uses of research are often contrasted negatively with instrumental uses of evidence to inform decisions regarding programs and practices (see, for example, Coalition for Evidence-based Policy, 2002; Dunworth et al., 2008). But political uses of research are inevitable, and they can be beneficial and bring about needed changes to practice when the evidence is recruited accurately to the cause (Weiss & Bucuvalas, 1980). For state leaders and district decision makers, data are necessary tools of persuasion regarding adoption decisions for new instructional materials. And while data on learning outcomes are especially prized, data related to implementation and to teacher and student perceptions are valuable, too, for learning about how educators adapt materials in ways that can inform iterative design (Penuel et al., 2014). In a world where individual teachers have considerable say over what materials they use, positive endorsements from peers matter a great deal to leaders (Torphy et al., 2020).

To support effective political uses of research by state education leaders, we began the field test by engaging state leaders in thinking about the data we might collect, and they provided feedback on what data would be most important for informing local leaders' decisions regarding the adoption of OpenSciEd. On that menu were documentation of student learning outcomes, teacher and student perceptions of the materials, implementation of the materials, and documentation of changes to practice to align better with the vision of teaching and learning in *A Framework for K-12 Science Education* (National Research Council, 2012). State leaders believed student learning outcomes would be important to informing local leaders' adoption decisions. They also highlighted the value of two other kinds of evidence to them, namely that students from systemically marginalized groups and communities felt the materials were personally relevant to them

and important to their communities, and that the materials supported changes to practice that they thought were important to make, but also difficult to make without high quality instructional materials. For state leaders, new materials were part of a broader strategy to change science teaching, not an end unto themselves, and so data on whether the materials were supporting such changes were important to them.

To accommodate their needs, we incorporated student surveys of experience into data collection like ones that we had used in the past (see, e.g., Penuel, Raza, et al., 2024). For example, we gave students survey items such as, "I found today's lesson interesting," and "Today's lesson relates to a problem we have in our city or town that needs to be solved" as part of brief surveys which students completed once per unit. In slide presentations to State Science Supervisors, data team leaders presented results, disaggregated by race/ethnicity and by gender, that could be reused in state contexts. To facilitate both leaders' and developers' own queries of the data by state, the team created a dashboard for each unit, with data from all survey items that had been administered to students and teachers.

Another potential use that we anticipated for the tools for data collection we had developed was to support adoption decisions. Most larger school districts have formal adoption processes for instructional materials (Allen & Seaman, 2017). As part of that process, many also engage in pilot testing with groups of teachers, and a few also incorporate data collection into the pilots, to understand what teachers think about the materials and how they are enacting them (EdReports, 2023). The process—as well as the instruments—for rapidly gathering and analyzing data from the field test of OpenSciEd is one that was designed to be adaptable for other kinds of pilots, supported by external partners (e.g., a research or evaluation team). If used in this way, the field test process and instruments could enhance the capacity of systems to learn from pilot tests of instructional materials and to customize teacher learning based on evidence of how teachers are perceiving and enacting materials. Unlike previous instruments and approaches (e.g., Hall & Hord, 2014), the measures are keyed to specific practices targeted in materials and focus on the student experience, not just on teacher implementation.

To assess the degree to which relationships among students and between teachers and students and classroom practice had shifted, we included questions on student surveys that asked about their perceptions of the classroom community. We incorporated scales about students' sense of belonging, care, sense of epistemic agency, and perceptions about what they thought it took to succeed in their science classrooms. We also asked them to report on classroom practices that their teachers used to support their collaborative sensemaking about phenomena. These were scales which we had first piloted and reported on as part of our study of middle school instructional materials for OpenSciEd (Penuel, Krumm, et al., 2024; Singleton et al., 2024), or that we had adapted from other studies that focus specifically on instructional shifts promoted in the NGSS *Framework* (e.g., Campbell et al., 2021; Hayes et al., 2016). And as with data on student experience, we presented these data to state leaders in a disaggregated fashion, to show the degree to which teaching and learning were perceived to be equitable by students themselves.

Researchers from the OpenSciEd developers consortium have also shared its instruments with a growing community of scholars who are interested in studying OpenSciEd. A nonprofit organization, Digital Promise, has formed the OpenSciEd Research Community that connects researchers, practitioners, policymakers and innovators who are interested in studying the use of the materials and associated professional development both to

advance scholarship in science education and build field-wide capacity for implementation. The group has awarded mini grants to scholars to develop larger research proposals for external funding that are building on the knowledge base generated by members of the OpenSciEd developers consortium.

Potential Generality of the Heuristics and Conditions for Synchronization

The heuristics named above worked together to promote synchrony and are unified by a common focus on continuous attunement in the present to multiple interest holders that is also focused on potential for promoting equity at scale in the future. The work of attunement and anticipation we maintain is likely present in any successful development effort, but is rarely the focus of guidelines or advice to investigators. In making these heuristics explicit, we hope that this more intensive focus on the relational dimensions of development and adaptation studies provides useful guidance both for individual studies and the preparation of investigators and other team members in studies. That is to say, we hope that the project has embodied the principle of generativity within the ontological conception of relevance offered by Akkerman and colleagues (2021).

Whether or not these heuristics can be readily applied to other development and adaptation projects is partly an open question. However, there are other examples of projects like OpenSciEd High School that have used many of the same approaches to synchronizing research and practice that our team has, and with similar levels of success in terms of reach. Many of these projects have also demonstrated impact on student outcomes, such as Success for All (Cooper et al., 1998) and Building Blocks (Sarama & Clements, 2004). Notably, the stories told about these projects tend to highlight the outcomes but not the ways that research and development were structured in a way to maintain continuous responsiveness to a wide range of interest holders. Peurach's (2011) account of Success for All is an exception, and it is notable for the detail he provides about how researchers and the Success for All Foundation engaged in continuous learning about and from implementation, adjusting course as needed to support the growth and spread of the intervention. What was replicated across sites was as much a set of routines and practices as it was a particular design for improving learning (Peurach & Glazer, 2012).

What is arguably different about OpenSciEd from these other projects is that, rather than starting small, the project started big, in terms of its reach, and it was designed for rapid expansion, even as evidence of its promise was still being gathered. As Clements (2007) rightly notes, these investments are resource-intensive, and investing in something not yet proven effective in a range of classrooms carries risks. OpenSciEd was about three times the size of investment made in a typical U.S.-based design and development grant. Can funders like those that invested in OpenSciEd bear such risk for other projects?

In the case of OpenSciEd, several factors helped to reduce that risk that may need to be present to answer that question affirmatively. For one, starting big was possible in part because of the specific conditions in which the project found itself—several years into a standards movement that had yet to produce high-quality instructional materials for the secondary level. There were, moreover, willing and enterprising leaders and teachers who were ready to try out the materials we were developing. And there was a research and design team with both experience in multi-institutional collaborative design and ample resources to accomplish its aims. This kind of confluence of political conditions, will, and capacity that is needed for reforms to succeed is well-established (Bryk, 2015; Cohen & Ball, 1999; Tichnor-Wagner et al., 2017).

Having these conditions in place, however, would not have been sufficient for our team to achieve success. Synchronizing research and practice to promote relevance required multiple, structured opportunities for continuous engagement across the research-practice-policy divide. Not only did these opportunities facilitate the use of data and evidence from the field test for different purposes, they also enabled researchers to stay closely attuned to different interest holders, from students to teachers to state-level leaders. There was also a need to incorporate ongoing learning opportunities in the collaboration. We accomplished this through periodic internal team retreats, whole-team meetings, and a regular newsletter that documented stories, but also through our collective commitment to orienting to one another's current concerns and hoped-for futures. Our own experience is consistent with findings about other collaboratives like ours, which have found that supporting the ongoing learning of members is key–but also a challenge (Cohen et al., 2013). As a team comprised of both researchers and designers with different backgrounds in curriculum development and teaching, to engage in research and development that is relevant to practice requires continuous learning and enhanced capacities that can only be developed through collaborative endeavors like this one, and with a great willingness to take considered risks on investing in people and networks who hold a vision for how teaching and learning could be better for all.

What Comes Next After a Development and Adaptation Project Like This One

In the report, *The Future of Education Research at IES: Advancing an Equity-Oriented Science* (National Academies of Sciences Engineering and Medicine, 2022), the committee called for a new category of research called Impact and Heterogeneity. Such studies might follow a Development and Adaptation project, if the goal were to gather causal evidence of an intervention's efficacy or effectiveness. In contrast to current practice at IES, the committee called for the category to include quasi-experimental studies to accommodate situations where systems-level change efforts make experimental designs infeasible. In addition, the committee called on scholars to pay closer attention to heterogeneity in treatment effects early on, as part of their emphasis on producing useful research that also attends to equity.

The OpenSciEd project would benefit from an efficacy study that would allow for causal inferences about the impact of using the instructional materials when supported by sustained professional learning from teachers. Such a study could employ the measures developed by the team as part of the Design and Adaptation project, alongside other, more distal measures that were developed independently. At this stage of the project, a quasi-experimental design might not just be more feasible to implement, and it might also be more desirable. The substantial changes to instruction required of most teachers for implementation makes it more desirable to include teachers who have taught the materials for at least a year before testing its impacts on students.

But efficacy studies are not the only research studies that are needed; several unanswered questions emerged from our work that are likely to emerge from other efforts like ours. For one, data are needed on families' or communities' feedback on interventions. Knowing something about the range of possible responses to materials might be valuable to district leaders considering adoption. They could use such information to inform local adaptation of the materials or to plan communications with caregivers about the materials. Second, it is important to gather information about teachers' and publishers' adaptations of materials, as to whether they maintain integrity to the design principles. For the middle school version of these materials, researchers are already investigating teachers'

adaptations ([McNeill et al., 2023](#)). This research, as future research on the high school program might, focuses on teachers' goals for adaptation, the form of those adaptations, and the role of different aspects of teachers' local contexts in shaping their adaptations. A third area of research might focus on sustainability. The long-term sustainability of this intervention depends on school- and district-level adoptions. While there is some research in the grey literature on adoption decisions (e.g., [Allen & Seaman, 2017](#)), understanding how OpenSciEd fares in adoption decisions and what reasons and data inform those decisions is an important question to answer. Furthermore, knowledge mobilization research may explore interventions to enhance the use of evidence from various sources (e.g., field test data, local pilot data) in decision-making processes. Research on sustainability might also explore strategies schools and districts pursue for sustaining teachers' implementation and ongoing learning from classroom implementation. At present, research on what happens "when the research ends" is rare and sobering ([Fishman et al., 2011](#)). Thus, future research on successful strategies for sustainability is needed.

Last, future research that might be conducted on Design and Adaptation projects like ours could be empirical studies of the processes themselves. We could not devote resources to contract or commission a study of our own process with an outside consultant. Had we done so, we would have been interested in documenting more systematically the development process, along with how we used feedback to inform revisions to materials. While we attempted through our newsletters to develop contemporaneous accounts of the process, they are necessarily from the developers' point of view. We think large Development and Adaptation projects like ours would benefit from a systematic, multi-perspectival account of patterns and variation in terms of how development unfolded within and across teams and courses. Such research may be crucial in demonstrating just how ambitious Development and Adaptations can be and still realize goals for quality, relevance, and impact.

At the present moment–in the U.S. at least–it is difficult to imagine when or how these next steps might take place, given the dismantling of the research and development infrastructure that happened at the beginning of 2025. But calls for more relevant research and for more responsible innovation that includes the voices of interest holders are global phenomena (e.g., [OECD, 2025](#); van [Atteveldt et al., 2019](#)). It is likely that the heuristics here could be used wherever there are significant investments in curriculum design that have commitments to being informed by evidence and guided by a vision for ensuring all students can see curriculum as a window into possible future and a mirror into their own lives and those of their communities.

References

Affolter, R., McNeill, K. L., & Brinza, G. (2022). Some of you are smiling now: Supporting trust, risk taking, and equity in your classroom. *Science Scope*, *45*(5), 26-34. [http s://www.nsta.org/science-scope/science-scope-mayjune-2022/some-you-are-smiling-now](https://www.nsta.org/science-scope/science-scope-mayjune-2022/some-you-are-smiling-now)

Akkerman, S., Bakker, A., & Penuel, W. R. (2021). Relevance of educational research: An ontological conceptualization. *Educational Researcher*, *50*(6), 416-424. https://doi.org/10.3102/0013189X211028239

Allen, I. E., & Seaman, J. (2017). *What we teach: K-12 school district curriculum adoption process*. Babson Survey Research Group.

Bang, M., & Vossoughi, S. (2016). Participatory design research and educational justice: Studying learning and relations within social change making. *Cognition and Instruction*, *34*(3), 173-193. https://doi.org/10.1080/07370008.2016.1181879

Biesta, G. (2007). Why "what works" won't work: Evidence-based practice and the deficit in educational research. *Educational Theory, 57,* 1–22. https://doi.org/10.1111/j.1741-5446.2006.00241.x

Björgvinsson, E., Ehn, P., & Hillgren, P.-A. (2012). Agonistic participatory design: Working with marginalised social movements. *CoDesign*, *8*(2-3), 127-144. https://doi.org/10.1080/15710882.2012.672577

Bryk, A. S. (2015). Accelerating how we learn to improve. *Educational Researcher*, *44*(9), 467-477. https://doi.org/10.3102/0013189X15621543

Campbell, T., Lee, H., Longhurst, M. L., McKenna, T. J., Coster, D. C., & Lundgren, L. (2021). Next generation science classrooms: The development of a questionnaire for examining student experiences in science classrooms. *School Science and Mathematics*, *121*(2), 96-109. https://doi.org/10.1111/ssm.12449

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula". *Journal for Research in Mathematics Education*, *38*(1), 35–70. https://doi.org/10.2307/30034927

Coalition for Evidence-based Policy. (2002). *Bringing evidence-driven progress to education: A recommended strategy for the U.S. Department of Education*. http://www.eric.ed.gov/PDFS/ED474378.pdf

Cobb, P. A., Jackson, K., Henrick, E., Smith, T. M., & the MIST Team (Eds.). (2018). *Systems for instructional improvement: Creating coherence from the classroom to the district office*. Harvard Education Press.

Cohen, D. K., & Ball, D. L. (1999). *Instruction, capacity, and improvement (CPRE Research Report Series N. RR-43)*. Consortium for Policy Research in Education. https://www.cpre.org/sites/default/files/researchreport/783_rr43.pdf

Cohen, D. K., Peurach, D. J., Glazer, J. L., Gates, K., & Goldin, S. (2013). *Improvement by design: The promise of better schools*. University of Chicago Press.

Conaway, C., Tipton, E., & Artiles, A. J. (2022, November 7). The value of variation: Why we need to attend to heterogeneity in intervention research. *The Public Scholarship Collaborative*. https://doi.org/10.1177/01614681241276956

Cooper, R., Slavin, R. E., & Madden, N. A. (1998). Success for All: Improving the quality of implementation of whole-school change through the use of a national reform network. *Education and Urban Society*, *30*(2), 385-408. https://doi.org/10.1177/0013124598030003006

Doan, S., Eagan, J., Grant, D., Kaufman, J. H., & Setodji, C. M. (2023). *American Instructional Resources Surveys: 2022 technical documentation and survey results*. RAND Corporation.

Doan, S., Kaufman, J. H., Woo, A., Tuma, A. P., Diliberti, M. K., & Lee, S. (2022). *How states are creating conditions for use of high-quality instructional materials in K–12 classrooms: Findings from the 2021 American Instructional Resources Survey*. RAND.

Dunworth, T., Hannaway, J., Holahan, J. and Turner, M.A. (2008) , Beyond ideology, politics, and guesswork: The case for evidence-based policy. Urban Institute.

Edovald, T., & Nevill, C. (2020). Working out what works: The case of the Education Endowment Foundation in England. *ECNU Review of Education, 4*(1), 46–64. https://doi.org/10.1177/2096531120913039

EdReports. (2023). *Access to quality curriculum is making a difference: Highlights from the field*. EdReports. https://www.edreports.org/resources/article/access-to-quality-curriculum-is-making-a-difference-highlights-from-the-field?utm_medium=email&utm_source=pardot&utm_campaign=state-district-spotlight-b2s-2023

Elmore, R. F., & Forman, M. L. (2011). *Building coherence within schools*. SERP.

Escobar, A. (2017). *Designs for the pluriverse: Radical interdependence, autonomy, and the making of worlds*. Duke University Press. https://doi.org/10.1215/9780822371816

Farley-Ripple, E., May, H., Karpyn, A. E., Tilley, K., & McDonough, K. (2018). Rethinking connections between research and practice in education: A conceptual framework. *Educational Researcher*, *47*(4), 235-245. https://doi.org/10.3102/0013189X18761042

Fishman, B. J., Penuel, W. R., Hegedus, S., & Roschelle, J. (2011). What happens when the research ends? Factors related to the sustainability of a technology-infused mathematics curriculum. *Journal of Computers in Mathematics and Science Teaching*, *30*(4), 329-353.

Goldman, S. R., Hmelo-Silver, C. E., & Kyza, E. A. (2022). Collaborative design as a context for teacher and researcher learning: Introduction to the special issue. *Cognition and Instruction*, *40*(1), 1-6. https://doi.org/10.1080/07370008.2021.2010215

Gutiérrez, K. D., & Penuel, W. R. (2014). Relevance to practice as a criterion for rigor. *Educational Researcher*, *43*(1), 19-23. https://doi.org/10.3102/0013189X13520289

Hall, G. E., & Hord, S. M. (2014). *Implementing change: Patterns, principles, and potholes*. Pearson Education, Inc.

Hannan, M., Russell, J. L., Takahashi, S., & Park, S. (2015). Using improvement science to better support beginning teachers: The case of the Building a Teaching Effectiveness Network. *Journal of Teacher Education*, *66*(5), 494-508. https://doi.org/10.1177/0022487115602126

Hayes, K. N., Lee, C. S., DiStefano, R., O'Connor, D., & Seitz, J. C. (2016). Measuring science instructional practice: A survey tool for the age of NGSS. *Journal of Science Teacher Education*, *27*, 137-164. https://doi.org/10.1007/s10972-016-9448-5

Hulleman, C., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, *326*(5958) , 1410-1412. https://doi.org/10.1126/science.1177067

Institute of Education Sciences, & National Science Foundation. (2013). *Common guidelines for education research and development*. Authors.

Ishimaru, A. M., & Bang, M. (2022). Solidarity driven co-design: Evolving methodologies for expanding engagement with familial and community expertise. In D. J. Peurach, J. L. Russell, L. Cohen-Vogel, & W. R. Penuel (Eds.), *The foundational handbook of improvement research in education* (pp. 383-402). Rowman & Littlefield.

Krumm, A. E., Penuel, W. R., Pazera, C., & Landel, C. (2020). Measuring equitable science instruction at scale. In M. Gresalfi, I. S. Horn, N. Enyedy, H.-J. So, V. Hand, K. Jackson, S. E. McKenney, A. Leftstein, & T. M. Philip (Eds.), *Proceedings of the International Conference of the Learning Sciences* (Vol. 4, pp. 2461-2468). International Society of the Learning Sciences.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher*, *35*(7), 3-12. https://doi.org/10.3102/0013189X035007003

Le Dantec, C. A., & DiSalvo, C. (2013). Infrastructuring and the formation of publics in participatory design. *Social Studies of Science*, *43*(2), 241-264. https://doi.org/10.1177/0306312712471581

Learning in Places Collaborative. (2022). *Ethical deliberation and decision-making in socio-ecological systems framework*. Learning in Places. https://learninginplaces.org/frameworks/ethical-deliberation-and-decision-making-in-socio-ecological-systems-framework/

Manz, E., Heredia, S. C., Allen, C. D., & Penuel, W. R. (2022). Learning in and through researcher-teacher collaboration. In G. Jones, J. A. Luft, & T. R. Tretter (Eds.), *Handbook of Science Teacher Education* (pp. 452-464). Routledge.

McNeill, K. L., Fine, C. G., Lowell, B. R., & Affolter, R. (2023, April). *Teachers' descriptions and rationales of customizations of storyline science curriculum: Adapting for their classroom contexts* NARST International Conference, Chicago, IL.

Morales-Doyle, D. (2024). *Transformative science teaching: A catalyst for justice and sustainability*. Harvard Education Press.

National Academies of Sciences Engineering and Medicine. (2018). *Design, selection, and implementation of instructional materials for the Next Generation Science Standards: Proceedings of a workshop*. National Academies Press. https://doi.org/10.17226/25001

National Academies of Sciences Engineering and Medicine. (2022). *The future of education research at IES: Advancing an equity-oriented science*. National Academies Press. https://doi.org/10.17226/26428

National Academies of Sciences Engineering and Medicine. (2024). *Equity in preK-12 STEM education: Framing decisions for the future*. National Academies Press. https://doi.org/10.17226/26859

National Research Council. (2002). *Scientific research in education*. National Academies Press. https://doi.org/10.17226/10236

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press. https://doi.org/10.17226/13165

NextGen Science, & EdReports. (2021). *Critical features of Instructional materials design for today's science standards*. WestEd and EdReports.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. National Academies Press. https://doi.org/10.17226/18290

OECD. (2025). Everybody cares about using education research sometimes: Perspectives of knowledge intermediaries. OECD Publishing. https://doi.org/10.1787/5ef88972-en.

OpenSciEd. (2019). *OpenSciEd Design Specifications*. https://openscied.org/why-openscied/design-specifications/

Palincsar, A. S., Magnusson, S. J., Collins, K. M., & Cutter, J. (2001). Making science accessible to all: Results of a design experiment in inclusive classrooms. *Learning Disability Quarterly*, *24*(1), 15-32. https://doi.org/10.2307/1511293

Penuel, W. R. (2019). Infrastructuring as a practice of design-based research for supporting and studying equitable implementation and sustainability of innovations. *Journal of the Learning Sciences*, *28*(4-5), 659-677. https://doi.org/10.1080/10508406.2018.1552151

Penuel, W. R., Allen, A.-R., Deverel-Rico, C., Singleton, C., & Pazera, C. (2023). How teachers' knowledge of curriculum supports partnering with students in their science learning. *Journal of Science Teacher Education*, *34*(8), 861-882. https://doi.org/10.1080/1046560X.2023.2167508

Penuel, W. R., Allen, A.-R., Henson, K., Campanella, M., Patton, R., Rademaker, K., Reed, W., Watkins, D. A., Wingert, K., Reiser, B. J., & Zivic, A. (2022). Learning practical design knowledge through co-designing storyline science curriculum units. *Cognition and Instruction*, *40*(1), 148-170. https://doi.org/10.1080/07370008.2021.2010207

Penuel, W. R., Bell, P., & Neill, T. (2020). Creating a system of professional learning that meets teachers' needs. *Phi Delta Kappan*, *101*(8), 37-41. https://doi.org/10.1177/0031721720923520

Penuel, W. R., & Fishman, B. J. (2012). Large-scale intervention research we can use. *Journal of Research in Science Teaching*, *49*(3), 281-304. https://doi.org/10.1002/tea.21001

Penuel, W. R., Krumm, A. E., Pazera, C., Singleton, C., Allen, A.-R., & Deverel-Rico, C. (2024). Belonging in science classrooms: Investigating its relation to students' contributions and influence in knowledge building. *Journal of Research in Science Teaching*, *61*(1), 228-252. https://doi.org/10.1002/tea.21884

Penuel, W. R., Phillips, R. A., & Harris, C. J. (2014). Analysing curriculum implementation from integrity and actor-oriented perspectives. *Journal of Curriculum Studies*, *46*(6), 751-777. https://doi.org/10.1080/00220272.2014.921841

Penuel, W. R., Raza, A., Salinas del Val, Y., Salinas-Estevez, R., Williamson, E., Smith, J., & Gill, Q. (2024). Making formative use of student experience data to promote equity. *Science Scope*, *47*(2), 34-39. https://doi.org/10.1080/08872376.2024.2314675

Penuel, W. R., Reiser, B. J., McGill, T. A. W., Novak, M., Van Horne, K., & Orwig, A. (2022). Connecting student interests and questions with science learning goals through project-based storylines. *Disciplinary and Interdisciplinary Science Education Research*, *4*(1), 1-27. https://doi.org/10.1186/s43031-021-00040-z

Penuel, W. R., Riedy, R., Barber, M., Peurach, D. J., LeBoeuf, W., & Clark, T. L. (2020). Principles of collaborative education research with stakeholders: Toward requirements for a new research and development infrastructure. *Review of Educational Research*, *90*(5), 627-674. https://doi.org/10.3102/0034654320938126

Penuel, W. R., Turner, M. L., Jacobs, J. K., Van Horne, K., & Sumner, T. (2019). Developing tasks to assess phenomenon-based science learning: Challenges and lessons learned from building proximal transfer tasks *Science Education*, *103*(6), 1367-1395. https://doi.org/10.1002/sce.21544

Peurach, D. J. (2011). *Seeing complexity in public education: Problems, possibilities, and Success for All*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199736539.001.0001

Peurach, D. J. (2016). Innovating at the nexus of impact and improvement: Leading educational improvement networks. *Educational Researcher*, *45*(7), 421-429. https://doi.org/10.3102/0013189X16670898

Peurach, D. J., & Glazer, J. L. (2012). Reconsidering replication: New perspectives on large-scale school improvement. *Journal of Educational Change*, *13*(2), 155–190. https://doi.org/10.1007/s10833-011-9177-7

Philip, T. M., & Sengupta, P. (2021). Theories of learning as theories of society: A contrapuntal approach to expanding disciplinary authenticity in computing. *Journal of the Learning Sciences*, *30*(2), 330-349. https://doi.org/10.1080/10508406.2020.1828089

Pierson, A. E., Brady, C. E., Clark, D. B., & Sengupta, P. (2023). Students' epistemic commitments in a heterogeneity-seeking modeling curriculum. *Cognition and Instruction*, *41*(2), 125-157. https://doi.org/10.1080/07370008.2022.2111431

Pollock, M., & Rogers, J. (2022). *The conflict campaign: Exploring local experiences of the campaign to ban "Critical Race Theory" in public K–12 education in the U.S., 2020–2021* Institute of Democracy, Education, and Access, University of California Los Angeles.

Reiser, B. J., Novak, M., McGill, T. A. W., & Penuel, W. R. (2021). Storyline units: An instructional model to support coherence from the students' perspective. *Journal of Science Teacher Education*, *32*(7), 805-829. https://doi.org/10.1080/1046560X.2021.1884784

Rorrer, A. K., Skrla, L., & Scheurich, J. J. (2008). Districts as institutional actors in educational reform. *Educational Administration Quarterly*, *44*(3), 307-357. https://doi.org/10.1177/0013161X08318962

Rose, D. H., & Meyer, A. (2002). *Teaching every student in the digital age: Universal Design for Learning*. ASCD.

Sarama, J., & Clements, D. H. (2004). *Building Blocks* for early childhood mathematics. *Early Childhood Research Quarterly*, *19*, 181-189. https://doi.org/10.1016/j.ecresq.2004.01.014

Schneider, M. (2018, November 14). How to make education research relevant to teachers. *Thought Leadership*. https://nces.ed.gov/learn/blog/how-make-education-research-relevant-teachers

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). NSTA.

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. https://doi.org/10.1080/10508406.2016.1207541

Singleton, C., Deverel-Rico, C., Penuel, W. R., Krumm, A. E., Allen, A.-R., & Pazera, C. (2024). The role of equitable classroom cultures for supporting interest development in science. *Journal of Research in Science Teaching*, *61*(5), 998-1031. https://doi.org/10.1002/tea.21936

Spencer Foundation. (2023). *Transformative research program*. Spencer Foundation. Retrieved April 13 from https://www.spencer.org/transformative-research-program

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568-1580. https://doi.org/10.1016/j.respol.2013.05.008

Stosich, E. L., Bocala, C., & Forman, M. (2017). Building coherence for instructional improvement through professional development: A design-based implementation research study. *Educational Management, Administration, & Leadership*, *46*(5), 864-880. https://doi.org/10.1177/1741143217711193

Therrian, W. J., Taylor, J. C., Hosp, J. L., Kaldenberg, E. R., & Gorsh, J. (2011). Science instruction for students with learning disabilities: A meta–analysis. *Learning Disabilities Research & Practice*, *26*(4), 188–203. https://doi.org/10.1111/j.1540-5826.2011.00340.x

Thomas, M. S. C., Howard-Jones, P., Dudman-Jones, J., Palmer, L. R. J., Bowen, A. E. J., & Perry, R. C. (2024). Evidence, policy, education, and neuroscience—The state of play in the UK. *Mind, Brain, and Education, 18*(4), 461–473. https://doi.org/doi.org/10.1111/mbe.12423

Tichnor-Wagner, A., Wachen, J., Cannata, M., & Cohen-Vogel, L. (2017). Continuous improvement in the public school context: Understanding how educators respond to plan-do-study-act cycles. *Journal of Educational Change*, *18*, 465-494. https://doi.org/10.1007/s10833-017-9301-4

Torphy, K. T., Liu, Y., Hu, S., & Chen, Z. (2020). Sources of professional support: Patterns of teachers' curation of instructional resources in social media. *American Journal of Education*, *127*(1), 13-47. https://doi.org/10.1086/711008

van Atteveldt, N. M., Tijsma, G., Janssen, T. W. P., & Kupper, F. (2019). Responsible research and innovation as a novel approach to guide educational impact of mind, brain, and education research. *Mind, Brain, and Education, 13*(4), 279–287. https://doi.org/10.1111/mbe.12213

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. https://doi.org/10.1007/s11251-014-9340-7

Warren, B., Vossoughi, S., Rosebery, A., Bang, M., & Taylor, E. (2020). Multiple ways of knowing: Re-imagining disciplinary learning. In N. i. S. Nasir, C. D. Lee, R. Pea, & M. McKinney de Royston (Eds.), *Handbook of the Cultural Foundations of Learning* (pp. 277-293). Routledge.

Weiss, C. H., & Bucuvalas, M. J. (1980). *Social science research and decision-making*. Columbia University Press.

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.

Whitehurst, G. J. (2003, April). *The Institute of Education Sciences: New wine, new bottles* Annual Meeting of the American Educational Research Association, Chicago, IL.

Willingham, D. T., & Daniel, D. B. (2021). Making education research relevant: How researchers can give teachers more choices. *Education Next*, *21*(2), 28-33.

Zvoch, K. (2012). How does fidelity of implementation matter? Using multilevel models to detect relationships between participant outcomes and the delivery and receipt of treatment. *American Journal of Evaluation*, *33*(4), 547-565. https://doi.org/10.1177/1098214012452715

End Notes

[1]We would characterize our stance as one of political *restraint* rather than neutrality, and we acknowledge that such a position is itself a political position. We did attempt to expose students to perspectives of interest holders who are farthest from power when presenting issues; in addition, we directly encouraged students to consider what values they wanted to use in recommending design solutions to problems that pertain to injustices and where science and engineering ideas and practices might have something to offer. Nonetheless, we acknowledge this approach is different from what some equity- and justice-oriented scholars in science education would advocate.

About the Authors

William R. Penuel (william.penuel@colorado.edu) is Distinguished Professor of Learning Sciences and Human Development in the Institute of Cognitive Science and School of Education at the University of Colorado Boulder. He designs and studies curriculum materials, assessments, and professional learning experiences for teachers in STEM education, primarily in science. He also investigates how contemplative practices and critical inquiry can support educators in cultivating more compassionate learning environments and schools. A third line of his research focuses on how long-term research-practice partnerships can be organized to address systemic inequities in education systems linked to race, gender and sexual diversity, and language.

In each of his projects, Penuel works in partnership with educators and education leaders to explore how to attenuate inequities in school systems by: (1) creating equitable classroom cultures that attend to student experience; (2) testing strategies for address epistemic injustices in whose knowledge is elicited and valued; and (3) connecting teaching to the interests, experiences, and identities of learners, particularly those to whom our society owes an education debt. He uses a wide range of research methods, including one Penuel and colleagues developed called design-based implementation research, to test what they co-design.

Penuel is an author of two books on research-practice partnerships, *Creating Research-Practice Partnerships in Education* (Harvard Education Press, 2017), and *Connecting Research and Practice for Educational Improvement* (Routledge, 2018) and co-edited a book on improvement research, The *Foundational Handbook on Improvement Research in Education* (Rowman & Littlefield, 2022). He is also co-author on a book released by John Wiley and Sons in late April 2025, *Creating Compassionate Change in School Communities: Leading Together to Address Everyday Suffering in Schools.*

Carol Pazera (Carol.pazera@austin.utexas.edu) serves as evaluation lead at the Charles A. Dana Center at The University of Texas at Austin. Shemanages and conducts research and evaluation for projects and initiatives that facilitate a successfultransition for students from high school to college; build the capabilities of constituents to identify andaddress barriers to student access and success in math and science; and introduce educators to research andpractice to support students to engage, persist, and succeed in math and science.

Carol coordinates multistate evaluation efforts and manages the work of the Center's external research andevaluation personnel. She liaises with external organizations to bring a variety of expertise to impactmeasurement and to ensure that the voices of students and families are centered in the work.Before joining the Dana Center, Carol served as a research analyst at the Austin Independent School District,where she designed and conducted evaluations of programs. Carol led the evaluation of the district's Careerand Technical Education program and conducted research on access to postsecondary education, especiallyfor Hispanic graduates, and strategies to address disparities. She was executive director and vice president ofevaluation at True Brilliance, a nonprofit that helped teens experiencing disciplinary issues succeed in highschool.

Carol holds a B.A. in Secondary Education/Social Studies from University of Illinois, Urbana-Champaign, aswell as an M.S. in Community and Regional Planning and an M.A. in Latin American Studies from TheUniversity of Texas at Austin.

Zoë Buck Bracey (zbuck@bscs.org) is Senior Science Educator and Director of Learning for a Just and Sustainable Future at BSCS Science Learning. She works in collaboration with educators, administrators, students, and scholars to lead cross-functional teams in the design of comprehensive resources to support the teaching and learning of science, particularly in physics, astronomy, and Earth science.

Daniel Edelson is the executive director of BSCS Science Learning, a nonprofit center for research and development. He is a curriculum and software developer, educational researcher, and advocate for educational reform. In his work, he draws on contemporary research to develop programs that support effective teaching and learning across diverse settings.

Kate Henson (Kate.henson@colorado.edu) is the Director of K-12 STEM Teaching and Learning at inquiryHub at the University of Colorado Boulder. Kate is a former high school science teacher with a passion for supporting educators with high quality instructional materials. She holds a BS and MS in Biology and a PhD in Curriculum and Instruction.

Brian J. Reiser is the Orrington Lunt Professor of Learning Sciences in the School of Education and Social Policy at Northwestern University. Reiser's work explores how to make science learning more meaningful in K-12 classrooms as students investigate questions and problems they identify. Reiser's research examines how to support students in science knowledge-building practices through storyline curriculum materials and teaching approaches, and how teachers learn as they enact these reforms. Reiser heads *NextGen Science Storylines*, a researcher-teacher collaborative developing and investigating design principles for storyline units in which students help manage the trajectory of science knowledge building. Reiser leads the Northwestern team of the *OpenSciEd* Developer's Consortium, a partnership with ten state education agencies to design and field test storyline instructional materials. The project has released a freely available middle school science curriculum in 2022, a high school curriculum in 2024, and is currently developing materials for K-5. Reiser was a member of the National Research Council's Board on Science Education from 2011 to 2018, serving on the NRC committee authoring *A Framework for K-12 Science Education*, which led to development of the Next Generation Science Standards and related three-dimensional standards across many states. Reiser is a Member of the National Academy of Education, a Fellow of the International Society of the Learning Sciences, and a Fellow of the American Educational Research Association.

Tamara Sumner sumner@colorado.edu is a Professor at the University of Colorado in the Institute of Cognitive Science and the Department of Computer Science. She leads an interdisciplinary research and development lab that studies how computational tools – combining learning sciences, machine intelligence, and interactive media – can improve teaching practice, learning outcomes and learner engagement in K12 STEM education. This research and development team combines expertise in research-practice partnerships, learning sciences, STEM education, codesign, human-computer interaction, software engineering, learning analytics, and machine learning/natural language processing. She has authored or co-authored over 240 articles on these topics.

Dr. Nicole Vick is a team member on the Next Generation Science Storylines Project at Northwestern University. Prior to joining the project, she was a high school science teacher for 18 years teaching a multitude of courses. Nicole has led professional development to support educators transitioning to three-dimensional teaching across the United States. She is active in state and national science education organizations having served as a NSTA District Director, Illinois Science Teaching Association President, and was named a NSELA 2024 VESEL Scholar. She received her Doctorate of Education from the University of Missouri St. Louis.

Jim Ryan (jryan@openscied.org) has over 30 years of experience working to improve STEM education for all students. As OpenSciEd Executive Director, Jim leads day-to-day operations, partner relationships, finance, strategic planning, and oversees Board meetings. Previously, he was the STEM Executive Director for San Francisco Unified School District (SFUSD) where he oversaw the development and implementation of core curricula in math, science, and computer science. Prior, he served in several leadership roles at Key Curriculum Press, a leading publisher of mathematics and science tools and curricula, where he led the creation of a new division focused on supporting the implementation of the Common Core State Standards (CCSS) for Mathematics. Jim also worked as an analyst, programmer, and team lead for Apple Computers. He started his education career as a high school teacher and site administrator.

Douglas Watkins (douglas_watkins@dpsk12.net) supports K-12 Science teachers in Denver Public Schools through the team of instructional specialists he manages. He is proud to have had a hand in shaping the science instructional landscape in DPS to a point now where they can claim to have officially adopted science curriculums for grades K-12 that meet or exceed the state standards for science in Colorado.