EDUCATIONAL DESIGNER

JOURNAL OF THE INTERNATIONAL SOCIETY FOR DESIGN AND DEVELOPMENT IN EDUCATION

Balancing Design Goals for Supporting Students' Work with Extant Data in Science

William R. Penuel, Andee Rubin, Gillian Puttick, Kate Henson, Clarissa Deverel-Rico

Abstract

This paper examines design decisions of a team seeking to support students' working with data in a standards-based high school biology curriculum. The team's decisions required them to balance four goals that often came into tension during development: (1) helping students meet performance expectations specified in the targeted standards; (2) engaging students with extant datasets; (3) supporting student sensemaking; and (4) supporting coherence from the student point of view. Efforts to balance these goals in design revealed the limitations of existing science standards for adequately supporting students' work with extant datasets and for developing students' skill in covariational reasoning. Achieving the goals of supporting student sensemaking in science requires more intensive support for building the conceptual foundations of statistical concepts when developing a grasp of the practice of using mathematics in science.

Introduction

Increasingly, designers are considering how to help students learn to work with data within scientific disciplines. One reason for this is that working with data within a discipline can provide ways to engage students with the practices of that discipline. In science, for example, exploring real-world phenomena is a powerful context for engaging students in meaningful disciplinary practices, including analyzing and interpreting data (National Academies of Sciences Engineering and Medicine, 2019). Anchoring science teaching in phenomena is useful for teaching data practices because the role of data in investigations helps students see that data are "numbers with a context" (Cobb & Moore, 1997, p. 801).

When designing materials to help students learn to work with data in science classrooms design teams face several important decisions. These include choosing which disciplinary standards involving data to focus on (Krajcik et al., 2014) and then choosing phenomena that will require students to use data and mathematics to support sensemaking as emphasized in those standards (Penuel, Allen et al., 2022). Design teams must also decide whether to have students collect their own data or use extant datasets. Working with extant scientific datasets presents additional issues, such as how well the available data fits the questions students are pursuing (Rubin, 2020). Choosing among datasets requires a careful consideration of the different goals of design team members and other interest holders, as well as students' prior experiences working with data.

In this paper, we present design rationales from a project in high school biology, focusing on how we attempted to balance different design goals as the project progressed. As elaborated below, the project developed instructional materials to build students' capacity to work with data as part of a unit that adhered to the Next Generation Science Standards (NGSS; NGSS Lead States, 2013). While these standards are specific to the US context, they are based on *A Framework for K-12 Science Education* (National Research Council, 2012) that has inspired reforms in science across the globe that seek to support conceptual learning through inquiry (see Oliver et al., 2021, for examples). We describe how we addressed challenges presented by how learning goals are described in the standards, as well as challenges that come with the decision to use extant datasets, to develop one lesson in a ninth-grade ecology unit. We present implications both for how educational designers might develop materials for helping students work with extant data and for how the standards might be revised to better support this development.

The Need for Instructional Materials for Working with Data in Science

Within science education, a focus on data is not new, but the integration of data with science education still has far to go. Since the 1990s, citizen science projects have engaged students in collecting, reporting, and sometimes helping to make sense of large data sets related to phenomena such as annual monarch butterfly migrations (Howard & Davis, 2004; Nugent, 2017), changing patterns of when trees grow and shed leaves in spring and fall (Mayer, 2010), and urbanization's effects on plant life (Gazal et al., 2008). Research found that students had difficulties in working with such data, such as formulating questions, interpreting different representations of data, and generating evidence-based claims from data (Coleman & Penuel, 2000; Feldman et al., 2000). Further, the emphasis of many citizen science projects on following complex protocols for data collection written by scientists made it less likely that students would have a chance to pose their own questions, design protocols, or interpret data (Penuel & Means, 2004; Penuel et al., 2005).

Another line of research led by Lehrer, Schauble, and colleagues (Lehrer, 2012; Lehrer & English, 2018; Lehrer et al., 2007; Lehrer & Schauble, 2011) explored how elementary-aged students gathered, interpreted, and used data to build models of natural systems. This work was the foundation of learning progressions for working with data (Lehrer et al., 2020). In addition, it has led to a re-articulation of what is a science investigation (Manz et al., 2020), where learning to construct and critique data models plays a central role in developing students' grasp of the science practice of planning and carrying out investigations.

Research by Puttick and colleagues (Drayton & Puttick, 2016; Puttick et al., 2015) centered on the development of a high school capstone course focused on Biocomplexity using extant socio-ecological data. This work suggested that to make sense of data, students need information about data sources and methods of collection, the ecological or social context, and information on the purpose and design of the original study in which the data were collected.

Simultaneously, advances in statistics education have increased our understanding of how students reason with data, and research teams have developed several student-oriented computer tools for visualizing and manipulating data. This scholarship has emphasized the central importance of variation and variability in data analysis, as well as the value of constructing models of distributions of data (e.g., Konold et al., 2007). As Rubin (2020) writes, "the process of data analysis can generally be regarded as an attempt to parcel out the variability in a set of measurements, attributing some of the differences in values to discernable causes, while other differences are unexplained" (p. 157). In addition, this literature points to the need to engage students in considering the contexts under which extant data were produced and how variation or distribution among cases can be represented as an aggregate value (Konold et al., 2015). The tools that the statistics education community has developed are intended to help students make sense of variability in data. For example, the Common Online Data Analysis Platform (CODAP; Finzer et al., 2019) was developed to support data analysis for datasets in any discipline. Tools like CODAP enable students to deftly visualize and manipulate data with an intuitive drag-and-drop interface (Rosenberg et al., 2020), helping them discern patterns in data and gain insight into their own questions (Lehrer & English, 2018).

Efforts to integrate data analysis tools and processes systematically into science instructional materials focused on making sense of phenomena are only in the early stages of development. Some of these efforts hearken back to earlier citizen science initiatives, in that they involve students collecting and analyzing data on community matters, such as air quality (e.g., Noushad et al., 2022). Others, like the Data Nuggets project, engage students in working with extant data from ongoing or recently completed scientific investigations (Schultheis & Kjelvik, 2015). As these projects have progressed, the complexities of integrating data and science have become evident. For example, the Data Clubs project (Sagrans et al., 2022) attempted to prepare teachers to develop data experiences for students using CODAP but found that many of them struggled to find a dataset of the appropriate size, complexity, and content for their students.

There remains a need for interdisciplinary research that draws on these insights to support learning goals in science education. Few students in science classrooms engage with curricular materials that provide them with exposure to the purposes, actions, choices, and tools of scientists working with data (Sagrans et al., 2022). Instead, students are expected primarily to read and interpret the data they are handed, without the opportunity to delve deeper into the decisions that led to the creation of the data. To gain a feeling for the intricacies of working with data in science requires students to see how data can be used as evidence to support answers to their own questions about phenomena and problems, as we elaborate below. Further, students need opportunities to make decisions like those that scientists make while studying phenomena, such as decisions about what data to use and what to discard, and how to represent data in ways that make visible the relationships among variables.

Goals for Our Instructional Materials Design

Our aim was to enhance an existing set of instructional materials for high school science, as part of the first, fourth, and fifth authors' efforts to develop the freely available OpenSciEd biology instructional materials. The design team included developers of the original materials, learning scientists with expertise in biology, and an expert in statistics education. Teachers in the project served as co-designers and as field testers of the unit.

Our project evolved to focus on a single lesson that spanned two classroom periods within a unit on ecosystems. The lesson had to support the unit-level goal of helping students understand the factors that affect carrying capacities of ecosystems at different scales. Below, we elaborate on the goals that guided our design efforts: helping students meet performance expectations specified in the targeted standards; engaging students with extant datasets; supporting student sensemaking; and supporting coherence from the student point of view.

Design Goal 1: Helping Students Meet Performance Expectations Specified in Standards

The focal science standards for the unit come from the Next Generation Science Standards in the United States (NGSS; NGSS Lead States, 2013). The NGSS are a set of learning goals that aim to reflect a vision of science as both knowledge and practice grounded in research on how people learn and how science is carried out (National Research Council, 2012). The vision describes science proficiency as comprising an integrated understanding of disciplinary core ideas, science and engineering practices, and crosscutting concepts. The learning goals of the NGSS are therefore represented as "three-dimensional" performance expectations, which signal one way those understandings might be integrated in a student's performance. Students demonstrate proficiency when they apply their three-dimensional understanding to explain phenomena and solve problems (National Research Council, 2014; Penuel et al., 2019).

Several of the science practices call for students to work with data in the service of sensemaking about phenomena and problems. In high school, this includes the expectation that students will "decide on types, how much, and accuracy of data needed to produce reliable measurements and consider limitations of the precision of data" (Achieve, 2013, p. 7). Students are also expected to use mathematics when analyzing data, as illustrated in this goal for high school students: "Apply concepts of statistics and probability (including determining function fits to data, slope, intercept, and correlation coefficient for linear fits) to scientific and engineering questions and problems, using digital tools when feasible" (Achieve, 2013, p. 9). Similar standards emphasizing the need for students to be able to apply statistical concepts to analyze data in service of scientific understanding are found in international tests (OECD, 2022). These practices of analyzing and interpreting data and using mathematics were focal within the standards targeted in our development effort, presented in Table 1 below.

Design Goal 2: Engaging Students with Extant Datasets

A second goal was to engage students in working with real, extant data from scientific studies, not data that were made up for educational purposes or that were collected by the students themselves. We adopted this goal because ecological datasets often require significant resources and time to collect (Schultheis & Kjelvik, 2020); further, in the real-world practice of ecology, many scientists rely on datasets collected by other teams of scientists. Therefore, we conducted a thorough review of existing published studies that focused on the relevant ecosystem, timeframe, and phenomenon for our unit (i.e., the changing populations of wildebeest and buffalo in the Serengeti between 1960 and 2000) and identified potential datasets that students might use to answer questions they generated.

To support students in working with data, we sought to take advantage of existing data visualization and analysis tools. It is important for students to construct their own representations for answering their own questions, which often differ from those for which the data were collected (Konold & Higgins, 2003). Yet, this can mean that students spend precious time building tables and plotting points on graphs by hand and lack time to engage in meaningful sensemaking about the data. Appropriate data analysis technology can ameliorate this situation. We chose the Common Online Data Analysis Platform (CODAP; Finzer et al., 2019; https://codap.concord.org/), because it is free, browser-based, and designed specifically for educational settings, and because emerging research on the platform (e.g., Rosenberg et al., 2020; Sagrans et al., 2022) suggests it can be useful for working with the kinds of datasets students encounter in secondary science.

In this unit, we sought to engage students in data analysis that approximated—but did not replicate—scientists' own process of discovery. Our view is that it is not possible or even desirable to have students fully replicate scientific discoveries that unfold over decades. However, by focusing on a phenomenon and curating appropriate extant datasets, we aimed to engage students in developing explanations and arguments that relate to scientists' own questions about a phenomenon. In this way, students can see how science practices can be used to grapple with uncertainties in ways that approximate those of scientists (Manz, 2015).

Design Goal 3: Supporting Student Sensemaking

A key goal of science learning is to engage students in "making sense" of phenomena and problems, not just "learning about" science (Schwarz et al., 2017). We draw on Odden and Russ' (2019) definition of sensemaking as the "process of building an explanation to resolve a perceived gap or conflict in knowledge" (p. 187). Focusing on sensemaking requires developers to take care in choosing phenomena. Specifically, phenomena need to lead students to pose questions that will support their developing understandings of core ideas, crosscutting concepts, and science and engineering practices in the targeted standards (Penuel et al., 2022).

Students' initial questions about phenomena, as well as their efforts to plan and carry out investigations using data, all necessarily interact with their everyday ways of making sense of the world. Although sometimes these everyday ways of thinking are treated as "misconceptions" that impede learning, our stance—supported by research in the learning sciences—is that such ways of thinking can be productive stepping stones for learning disciplinary ideas and practices (Campbell et al., 2016). Students' everyday ways of making sense of the world can highlight and promote different ways of knowing in the science classroom (Bang et al., 2017; Warren et al., 2020). In this way, our goal of promoting sensemaking is intended to make learning equitable, by creating a curriculum that "allow[s] for, invite[s], and build[s] on learners' and families' diverse sensemaking and cultural and linguistic resources" (National Academies of Sciences, Engineering and Medicine, 2022, p. 24).

Design Goal 4: Supporting Coherence from the Student Point of View

Our approach to developing units is to organize them coherently around an anchoring phenomenon that makes visible a gap or conflict in students' knowledge and thus invites them to carry out investigations to fill in the gap or resolve the conflict (Reiser et al., 2021). While many curricula use a phenomenon as a "hook" that is soon left behind once students' interest has been captivated, we present a phenomenon as an "anchor." An anchoring phenomenon serves as a focal point for sensemaking to which students return multiple times over the course of a unit, as they develop understandings that help them create a more complete explanatory model of the phenomenon. Presenting the anchor at the beginning of the unit creates a "need to know" that guides students' sensemaking. When a "time for telling" (cf., Schwartz & Bransford, 1998) in a unit arrives, students see how what they are being told will help them figure out something they themselves have decided was important to explain. Finally, in a unit that is coherent in this way, students feel as though they have two kinds of epistemic agency: they are responsible both for developing knowledge as a class to help explain phenomena and for having some impact on the direction of the class's investigations (Alzen et al., 2023).

Multiple strategies can help designers maintain coherence from the student point of view. First, the designers can construct an outline for the unit called a storyline, organized around a set of questions that students might generate about the anchoring phenomenon (Reiser et al., 2021). To anticipate these questions, designers rehearse introductions of phenomena with students to discern if the anticipated questions indeed arise (Penuel et al., 2022). Pilot tests of materials from multiple classrooms are then used to determine how many actual student questions were addressed in the unit, leading to revisions that address unanticipated student questions. Ultimately, the goal is to have each lesson be organized to answer a question that students might plausibly generate, judge to be important to pursue, and be able to investigate.

Table 1 provides additional information about the four overall design goals we had set (i.e., which specific performance expectation, which extant data set, which anchoring phenomenon, which typical student questions). These details provide a frame for the design decisions we discuss below when our four main goals came into tension.

As additional contextual information, we describe the content of the overall unit, which focused on ecosystem interactions and dynamics. Students take up the broad issues of how ecosystems work and how understanding them can help us protect them. At the beginning of the unit, students are introduced to a major conservation initiative in the U.S., then turn to the Serengeti National Park, where they learn ecosystem and conservation principles. Finally, they apply these principles to regional conservation dilemmas in the U.S. Along the way, students work with data and simulations to investigate the carrying capacities of ecosystems, how group behavior impacts survival, and how biodiversity can support ecosystem resilience. A current, freely available version of the unit is available at https://www.openscied.org/instructional-materials/b-1-ecosystem-interactions-dynamics/.

Table 1. Specific goals and constraints for instructional materials design

Overall Goal	Specific Goal / *Constraints*
Help Students Meet Performance Expectations Specified in Standards	HS-LS2-1. Use mathematical and/or computational representations to support explanations of factors that affect carrying capacity of ecosystems at different scales. / Clarification Statement: Emphasis is on quantitative analysis and comparison of the relationships among interdependent factors including boundaries, resources, climate, and competition. Examples of mathematical comparisons could include graphs, charts, histograms, and population changes gathered from simulations or historical data sets. Assessment Boundary: Assessment does not include deriving mathematical equations to make comparisons.
Engage Students with Extant Datasets	Engage students with data to answer a question about whether an increase food availability caused a large increase in an herbivore population in a region / *No data available on food availability for the period of time in question*
Support Student Sensemaking	Anchoring Phenomenon: Students are tasked with explaining how populations of large herbivores could have grown so large so quickly on the Serengeti Plain between 1960-1970, after hearing an ecologist, Dr. Tony Sinclair, who spent his career studying the region, declared, “large mammals just don’t just do that sort of thing!”
Support Coherence from the Student Point of View	The sequence of lessons in the unit is organized around anticipated student questions and investigations of different plausible explanations for the growth of the herbivore population. Rehearsals and pilot data suggest that students want to explore food availability first.

Analytic Methods for Studying Goal Balancing in Design

The current paper focuses on documenting and analyzing a set of design dilemmas we conjecture are relevant to any project seeking to engage students in working with real, messy data to make sense of scientific phenomena. All designers face design dilemmas whenever they seek to balance multiple goals and values in design (Tatar, 2007). Our aim in writing this paper is to identify the challenges we faced in balancing design goals and describe the strategies other curriculum designers might use in managing these dilemmas themselves.

The primary data sources we relied on to identify design dilemmas were our project notes and artifacts, including meeting slides with which we engaged educators and advisors to the project in conversations about design decisions. Our project notes include a record of conversations among the research design team members over the course of three years. Of the 149 meetings in the project notes, we identified 73 meetings in which we discussed the design of either the focal lesson for our project (about changes to food availability on the Serengeti) or a transfer task (assessment) using extant data. Six of these meetings were co-design meetings with teachers, and two were with our project advisors. Of these, 31 involved explicit discussion of the four project goals identified in the previous section. The focus on specific goals ebbed and flowed; for example, we focused on science learning goals related to sensemaking about ecosystems early in the project, but later we turned more intensively to supporting students' sensemaking about covariation and how best to facilitate their learning about measures of correlation. Coherence issues were raised each time we focused on lesson design, sometimes implicitly but often explicitly.

These meetings became the focus for identifying design dilemmas, that is, moments where project goals came into conflict or needed to be balanced. Design dilemmas are not identical to moments when teams make decisions or solve problems. For example, decisions that did not involve design dilemmas pertained to how to phrase specific questions in the lesson to students, or how to introduce students to CODAP. By contrast, when a decision involves a design dilemmas, it is because there is competition between goals, given the resources that are available (especially classroom time). An example from our study (discussed in more detail below) was the dilemmas between the "using extant data" goal and "supporting coherence from the student perspective" goal because the extant data available could not be used to answer the question that we anticipated students would ask. We focused our analysis on meetings where goals were discussed and where they came into tension with one another.

Project slides for co-design sessions with teachers and/or advisory board meetings were another source of data for identifying the dilemmas as we experienced them, because we presented multiple goals (e.g., designing the lesson to be coherent from the student point of view) explicitly for our work with teachers, both for lessons and the overall unit.

After identifying the dilemmas, team members working alone or in pairs built short design narratives of the dilemmas and their resolutions (Hoadley, 2002). These narratives describe our deliberations and decisions, particularly as they relate to how we addressed design dilemmas. We then compared these narratives to one another to build the account that is presented here.

Findings

Below, we highlight four design dilemmas that involved balancing different design goals, and the implications they had for the kinds of learning opportunities for working with data the curriculum would include. For each, we describe what necessitated a design decision, what alternatives we considered, what decision we made, and why.

Dilemma 1: Finding a Dataset that Answers a Question Students Have

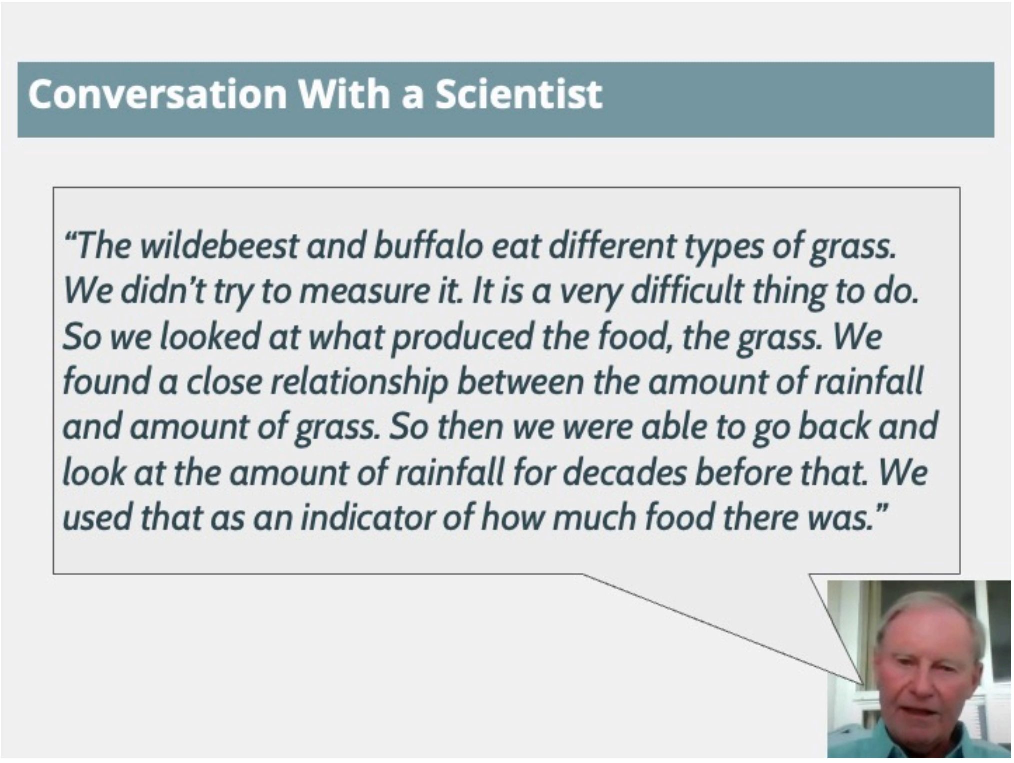
At the very beginning of the Serengeti unit, students learn that there was a marked explosion of the buffalo and wildebeest populations in the Serengeti between 1960 and 1975. In initial field tests, many students conjectured that this increase was primarily due to a significant increase in the amount of food available to the animals. This posed an initial challenge to developers because scientists had not collected data about food during that time and, thus, the relevant data to investigate this possibility were not available.

Because our design team was committed to providing opportunities for students to answer their own questions (fulfilling our goal of coherence from the student point of view), we searched for real data that could help answer their question about food availability. The two populations of herbivores in question feed primarily on grasses and forbs, but no scientists had been studying how much of these foods were available during the time of rapid population increase. The population increase was already underway when scientists began to wonder about this question themselves. While current global satellite data offers us ways to measure chlorophyll content and thus understand how much vegetation is in an area, this kind of data collection technology was not available then. Instead, researchers at that time had to rely on rainfall data as a *proxy variable* for food availability. Rainfall data had been collected in the past, albeit for different reasons.

As a design team, we decided that we should rely, as the scientists had, on a proxy variable —rainfall—for which there were publicly available datasets for the years 1960-75. We had not previously considered including the idea of a proxy variable, common in science and social science investigations, in the curriculum, but it was made necessary by our dual commitment to coherence and the use of extant data. One dataset involved monthly populations of wildebeest and buffalo and monthly rainfall in three different locations. This dataset showed that local populations of wildebeest fluctuated during the year based on monthly rainfall, thus providing students an opportunity to figure out how and why wildebeest migrate. A second dataset provided population data and rainfall by year to support analysis of students' question of whether increased food availability was a possible cause of the dramatic rise in the buffalo and wildebeest populations.

However, the rainfall data did not fully resolve the tension between the goal of promoting coherence from the student point of view and our goal of having students work with real data. We still needed a way to help students see the *need* for data on rainfall to answer their conjecture about increased food availability. While it made logical sense to look at rain to understand grass, it required a bit of careful design work to motivate students to ask questions about rain when they really wanted to know about grass. We decided the best approach would be to hear about the dilemma from an interview we conducted with the scientist who confronted it directly (Dr. Sinclair). This interview briefly explained why the scientists needed to study rainfall and how it was logically connected to the quantity of available grass, the animals' main food (see Figure 1). In this way, we were able to balance the goal of promoting coherence more productively with the goal of having students work with real data, using the video to motivate the need for students to look at rainfall data. However, we made sure the video did not give away the bigger conclusion from the investigations conducted by Dr. Sinclair's team: that an increase in food availability was not the cause of the big increase in population.

Figure 1 – Slide with interview excerpt explaining use of proxy data

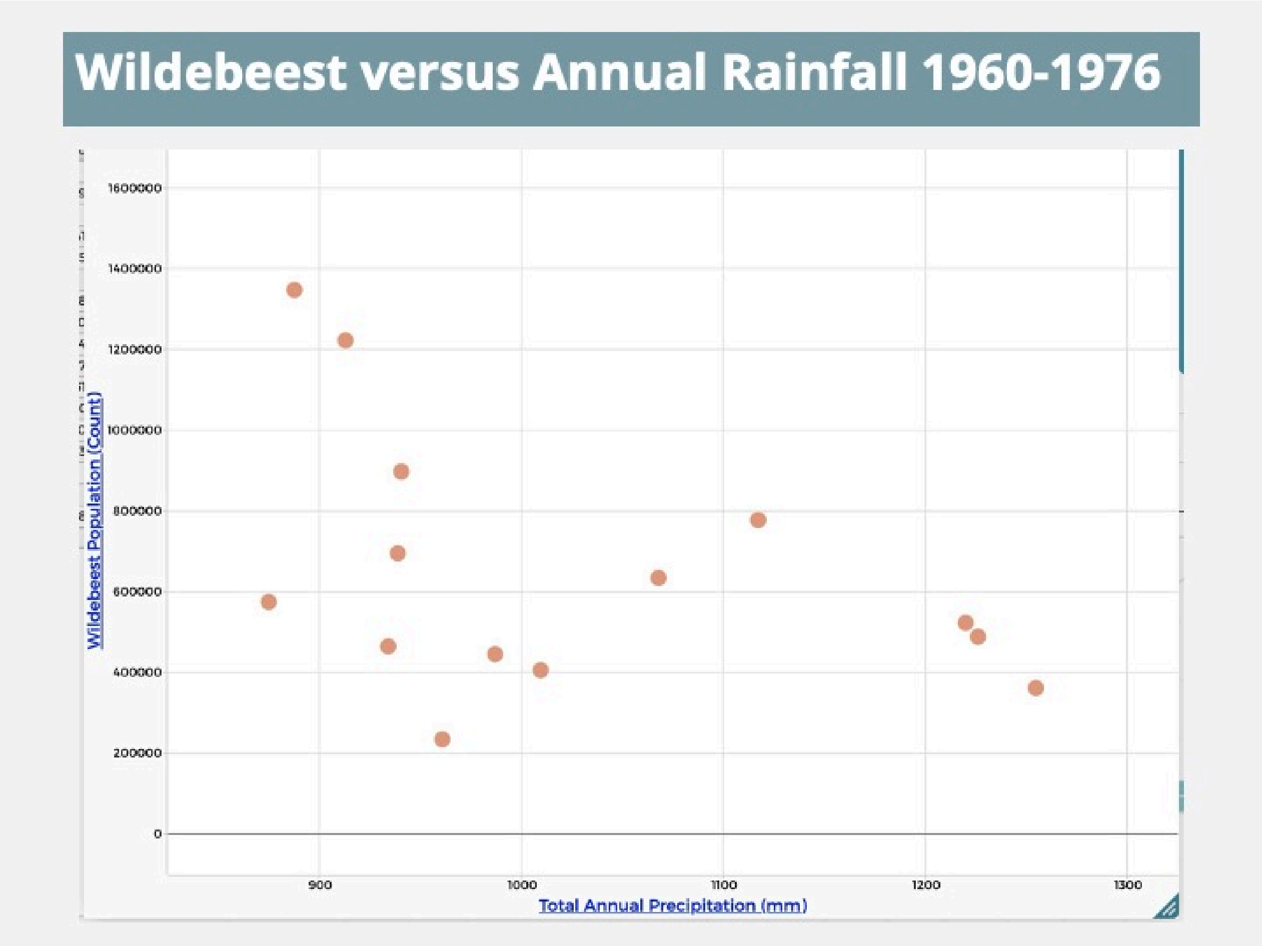


Dilemma 2: How Best to Support Students' Understanding of Covariation in Data

Deciding whether a change in rainfall would help explain the increase in population required students to engage in covariational reasoning, so we needed to decide how best to support their reasoning about patterns in the data. In this context, the team needed to balance the goal of supporting students' sensemaking with the goal of meeting the grade-level expectations for both the practice of analyzing and interpreting data and the crosscutting concept of cause and effect in the Next Generation Science Standards. In our field tests, we discovered that many students struggled to use the graphs of rainfall and animal populations to answer their initial questions. For example, in her implementation logs, a teacher noted that students struggled with "how to interpret graphs to determine what, if any, correlations exist."

Specifically, students struggled to make sense of a graph that showed a limited, if any, relationship between rainfall and population growth (see Figure 2). Although many students correctly concluded that a sudden increase in food availability did not cause the increase in populations of wildebeest and buffalo, their answers rarely pointed to graphs as evidence; if they did mention the graphs, they described them in ways that showed limited understanding of covariation. For example, one student said, "The points never overlapped on the graphs," while another wrote, "they [the data points on the scatterplot] are completely opposite over each other on the grid." These comments were made in spite of the fact that students had plotted best-fit lines on scatterplots in math class. We conjecture that in math class, students generally encounter scatterplots with relatively clear linear trends, so they have had little experience with non-relationships and may interpret their task with a scatterplot as finding the best-fit line, even if there is little evidence of a linear relationship.

Figure 2 – Scatterplot showing weak negative correlation between rainfall and wildebeest population. Each dot is one year's data



Further, students' question about food was not just about correlation but about causality—that is, did increased food availability (as measured through the proxy variable of rainfall) *cause* an increase in buffalo and wildebeest populations? Even if students were able to determine the presence or absence of correlation, a judgment of correlation alone cannot answer this question directly. Formulating a complete explanation would require students to link the correlation to an understanding about the role of rain in what wildebeest and buffalo eat, and an understanding about the role of food in the survival and growth of populations of organisms. In other words, to make a causal inference from these data, the data would need to exhibit a strong positive correlation, and students themselves would need to supply the causal mechanism from what they had learned about ecosystems. While some students correctly identified the potential for a causal relationship between food availability and population (e.g., "The more rain there is, the more food there is; the more food, the higher the population"), few used the data to reflect on the plausibility of this relationship explaining the population increase.

The standards for data analysis and causation provided us with a challenge to meeting our own goal of building on students' three-dimensional sensemaking. Here, students needed to make sense of a scatterplot to help them evaluate a causal claim. Statistics education research suggests that students first need to build a qualitative understanding of covariation—that is, different ways in which two variables might covary and how they might be reflected in a graph—before being able to interpret measures such as correlation coefficient (Cobb & Moore, 1997; Confrey & Smith, 1995). Judging the strength of a relationship between two variables is particularly challenging for students when they do not already have a conception of how the variables might be causally related (Wright & Murphy, 1984). These findings suggest that students need experience considering what a graph indicating covariation looks like and how variability and trend can coexist, especially in the absence of a strong knowledge base for answering a causal question. In our case, the need to use a proxy variable was an additional complication.

Yet the standards by which our unit would be evaluated (NextGen Science & EdReports, 2021) call for high school students to use more sophisticated data-related skills than the research indicates they are likely to bring. For the practice of analyzing and interpreting data, students should be able to "apply concepts of statistics and probability (including determining function fits to data, slope, intercept, and correlation coefficient for linear fits)" (NGSS Lead States, 2013, p. 391). In addition, the high school level element for cause and effect requires that students should be able to "understand that empirical evidence is required to differentiate between cause and correlation and to make claims about specific causes and effects" (NGSS Lead States, 2013, p. 414). For our unit to meet the high school standard would require us to build toward an understanding of a correlation coefficient, how it reflects the strength of correlation, and how to interpret a particular value. In addition, we would need to build students' understanding of when and how to use a correlation coefficient as evidence in support of a causal claim, as well as how a small correlation coefficient might be used to rule out an explanation. Because of the amount of time these would take, we opted not to teach correlation coefficients within this unit, instead privileging our sensemaking goal over fully addressing the standard to meet external review criteria.

Dilemma 3: When the Solution to Supporting Students' Covariational Reasoning Caused New Problems

To support students in developing a better understanding of covariation that could serve as a foundation for exploring correlation coefficients, we decided to have them work with data that were not part of the unit storyline. Our goal was to help students develop concepts about the strength and direction of covariation, as well as to explore the criteria for moving from covariation to causation. This departure from the storyline was in service of helping students make sense of the data on rainfall and herbivore populations. We chose a dataset (ESTEEM, 2022) in which relationships between variables were clear and for which the presence or absence of a causal mechanism was relatively obvious. The dataset comprised characteristics of a set of rollercoasters, including their top speed, length, and maximum height. It included one relationship that was both correlational and causal (between top speed and maximum height) and one that was not correlational (between top speed and length), so that students could begin to explore the complex relationship between correlation and causation.

However, this solution resulted in two of our goals coming into tension, namely that of supporting coherence from the student point of view and that of supporting student sensemaking. In selecting the rollercoaster dataset, we helped students develop a language for talking about covariation and looking at graphs to infer the strength of correlations, thus supporting their sensemaking capabilities. However, it wasn't clear to them how they would use that information to answer questions about the Serengeti, so we sacrificed a sense of coherence from the student perspective. A dataset related to the Serengeti – but with patterns as clear and causal mechanisms as obvious to students as the rollercoaster data set—would have been necessary to better balance these two goals.

Dilemma 4: Finding a New Phenomenon to Test for Transfer

A fourth challenge that we confronted was finding an analogous phenomenon to assess students' ability to apply their understandings of ecosystems dynamics in a new situation. In practice, it is often challenging to find a "parallel" phenomenon for a transfer task; finding one that also involves an appropriate data set is even more difficult. Working in the context of ecology added additional challenge, as even common ecological mechanisms function differently across different ecosystems. Finding a dataset for a performance task that takes only a single class period and meets the requirements of a given performance expectation is a very difficult task.

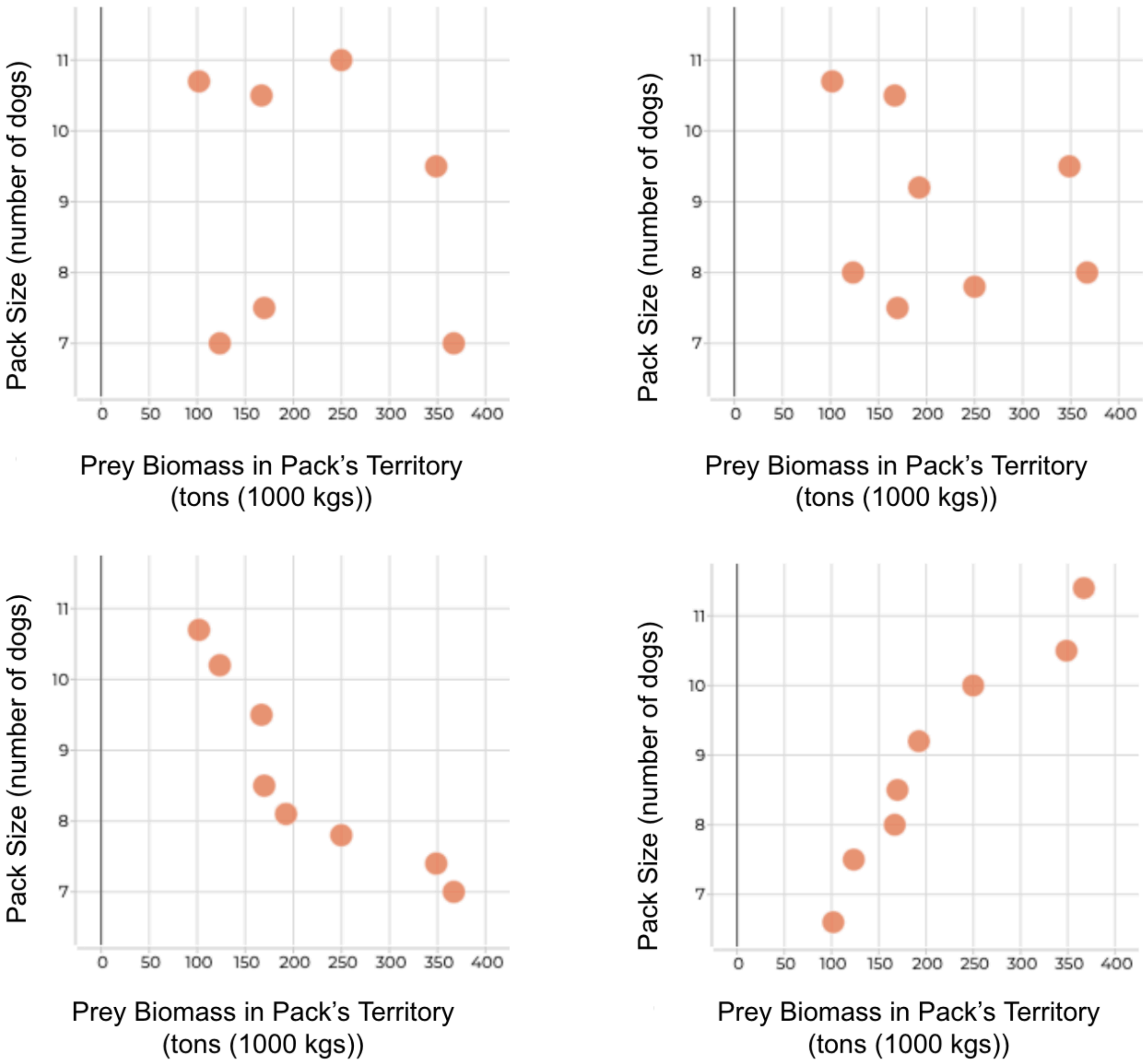
In the construction of a transfer task for the Serengeti unit, we began the search for phenomena by looking in the scientific literature for studies related to the key standard or performance expectation, in our case examining factors that could affect carrying capacity. We found a study that examined "factors likely to be affecting the distribution and density of wild dogs living in the Kruger National Park, South Africa" (Mills & Gorman, 1997, p. 1397). The study was framed around a puzzle that motivated biologists to study the phenomenon: the relationship between wild dog pack size and food abundance in their territory. Our challenge was to create a problem-solving context for students where they could explore the correlation between pack size and food abundance, while also exploring data that might provide evidence in support of a different set of factors influencing carrying capacity.

In designing the curriculum, we had been struggling with balancing the goals of having the unit be aligned with standards and simultaneously coherent from the student point of view. It proved hard, for example, to write individual prompts for each of the targeted elements in the performance expectation that made sense from a student point of view. These difficulties were repeated as we designed the transfer task. It turned out that predation was a factor that limited the size of packs in particular ecosystems—wild dogs avoided areas where there were lions, which preyed on them—and this was a factor called out in the element related to the Disciplinary Core Idea in the targeted learning standard. However, writing a prompt that required students to use the idea that the "significance of a phenomenon is dependent on the scale, proportion, and quantity at which it occurs," which was part of the Crosscutting Concept of Scale, Proportion, and Quantity, proved difficult to do in a way that would be motivated by a question students might pose themselves. In addition, the presentation of data in the study was not in a format that students had seen, and working with the data as presented would have likely been a test of far rather than near transfer (cf., Ruiz-Primo et al., 2012).

To balance the goals of meeting standards and creating a task that would be coherent from the student point of view, we adopted two strategies. First, we transformed the data presented in the paper ourselves to mirror the kinds of graphs that students had already seen in CODAP as part of the unit, so that the task did not test graph-reading along with understanding scientific concepts (see Figure 3). We also activated students' newly acquired knowledge about factors that affect animal populations in ecosystems by asking them to predict what factors might affect wild dog populations, before sharing data from the study. To elicit the targeted Crosscutting Concept, we presented students with a table showing information about pack size, territory size, and number of dogs per unit area, and asked them to estimate how large an area would be needed for a successful reintroduction of a pack of 10 wild dogs. To answer this question, we believed that students would need to use the ratio of pack size to territory size—that is, they would need to apply the concept of ratio (one dimension of the crosscutting concept Scale, Proportion, and Ratio) to help solve the conservation problem.

The data in the transfer task echoed the data on rainfall in that there was no correlation between food abundance and pack size. This made the task harder for students, as they looked for relationships that were not there, as one might expect from students taking a test. This repetition of the need to help students recognize and make sense of the absence of a relationship made this issue more salient. At this point, however, our design and development effort had come to an end, making this a question that we do not yet have a sense how to tackle.

Figure 3 – Sample of how graphs were represented within a transfer task prompt to better resemble CODAP graphs encountered in the unit



Discussion

In this study, we focused on documenting design decisions that involved attempts to balance different goals we had for instructional materials in science that could support students' meaningful engagement with data. Like past efforts to engage students in working with data (Feldman et al., 2000), our team found that students struggled to interpret and analyze data to answer a question of their own. We used prior research from statistics education on ways to support students' covariational reasoning as inspiration when we encountered students' difficulties with data (Cobb et al., 2003). Like other more recent projects that have incorporated tools like CODAP into the classroom, we found that such tools facilitated students' attention to relationships in data (e.g., Rosenberg et al., 2020), while allowing them to produce graphs easily, thus preserving time for the teacher and students to focus on sensemaking.

At the same time, it became apparent to us that there is a strong bias in US science education standards for requiring students to analyze data they have collected themselves from investigations they have planned themselves, a bias that is evident in teachers' practice as well (Rosenberg et al., 2022). This bias is understandable, in that students can better see how data practices from planning to analyses can work together when they carry out their own investigations (Manz et al., 2020). At the same time, many foundational ecological phenomena unfold over spatial and temporal scales that cannot be observed directly by students. This leads us to suggest that more emphasis on the practice of working with extant data is needed in standards, as is the issue of encountering and using proxy variables such as the one we had students rely upon. Scientists regularly must "make do" with existing variables, and in so doing, they sometimes need to modify the questions they are asking (Pickering, 1995). Further, standards might need to be adjusted to reflect the need for students to become skilled at interrogating the proxy variables they do use. For example, students must consider whether the ways in which proxy variables were produced make them useful for answering the questions they are posing.

Curriculum design work that is organized to support and study student learning via sensemaking, and not just to write materials to address specific standards, provides opportunities to explore how standards might be revised. Standards are committees' best thinking about what students should be able to know and do at a given time (Moss & Schutz, 2001). Standards are used to *guide* curriculum development; that is, they are developed before students have the chance to encounter materials designed to help them meet standards. Because such materials did not exist when standards were written, curriculum design studies like ours could be tools in helping revise standards. The deliberations, design dilemmas, team design decisions, and observed consequences of decisions can point to limitations of standards and how they might need to change.

Our design effort further revealed that the field still has a way to go in developing standards that can support depth, and not just breadth. A major criticism of instructional materials developed for the first generation of science standards is that they did not focus on big ideas with sufficient depth (Kesidou & Roseman, 2002). This finding echoes international studies, which have criticized American classrooms for the sheer number of curricular intentions pursued, to the expense of depth (Schmidt et al., 1997). Despite the attempt to organize the NGSS in a way to promote depth through a focus on a few disciplinary core ideas, we ran into challenges with the time needed to engage students meaningfully in working with data in the context of science so that they could gain a better "grasp of practice" (Ford, 2008) related to using data. Additional time is needed for the "data moves" (Erickson et al., 2019) that support sensemaking with data, as well as to understanding key ideas related to covariation and how to model and express the strength of relationships between variables quantitatively. With three-dimensional standards, each dimension requires sufficient time to develop and build. In the case of supporting students in working with data, the time required may be more than a science teacher can afford to spend. The standards may thus need to be revised in ways that sharpen the focus further on a few key ideas to provide more time for students to develop a grasp of foundational aspects of science practices. Standards and curriculum frameworks in other countries seeking to integrate science knowledge and data analysis practice will likely need to consider these needs in developing materials.

Our design process also revealed a set of opportunities and challenges to integrating perspectives from mathematics and statistics education fully into a science classroom. At multiple points, we found ourselves wrestling with science standards written in a way that did not reflect our understanding of the challenges students face when analyzing data and when interpreting covariation. And yet taking time to fully address these challenges by engaging students directly with the mathematical ideas in the context of a science class proved challenging. For science teachers faced with having to teach to many standards, this situation presents a complex dilemma: try to cover all the required standards or ensure students have the chance to make sense of data in ways that draw on the resources they bring to the classroom. The dilemma is similar for the curriculum designer faced with developing materials that embody science standards, when those standards do not provide any indication themselves of the underlying *mathematical* ideas required to make sense of relevant data. These goals cannot easily be balanced unless science standards directly address the mathematical ideas and practices embedded within the science standards.

Conclusion

This paper focused on the efforts of a design team to develop instructional materials to support students in working with data to investigate phenomena in service of science learning goals. By focusing on design dilemmas, we identified misalignments between expectations presented in science standards for what students should know and be able to do and the supports students need to be able to make sense of data in science. We revealed a gap, too, in the field's understanding of how best to support students in working with extant datasets in science to understand covariation between variables of interest, especially recognizing and interpreting a lack of covariation. Our study shows that analysis of dilemmas in curriculum materials design is a potential context for revising standards.

Acknowledgments

This material is based in part upon work supported by the National Science Foundation under Grant Number DRL-2031469. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Achieve (2013). *Appendix F: Science and Engineering Practices in the NGSS*. https://www.nextgenscience.org/sites/default/files/Appendix%20F%20%20Science%20and%20Engineering%20Practices%20in%20the%20NGSS%20-%20FINAL%20060513.pdf

Alzen, J., Edwards, K., Penuel, W. R., Reiser, B. J., Passmore, C., Griesemer, C., Zivic, A., Murzynski, C., & Buell, J. Y. (2023). Characterizing relationships between collective enterprise and student epistemic agency in science: A comparative case study. *Journal of Research in Science Teaching*, *60*(7), 1520-1550. https://doi.org/10.1002/tea.21841

Bang, M., Brown, B. A., Calabrese Barton, A., Rosebery, A., & Warren, B. (2017). Toward more equitable learning in science: Expanding relationships among students, teachers, and science practices. In C. Schwarz, C. Passmore, & B. J. Reiser (Eds.), *Helping students make sense of the world using next generation science and engineering practices* (pp. 33-58). NSTA.

Campbell, T., Schwarz, C. V., & Windschitl, M. (2016). What we call misconceptions may be necessary stepping-stones toward making sense of the world. *Science and Children*, *53*(7), 69-74. https://doi.org/10.2505/4/ss16_039_07_19

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, *104*(9), 801-823. https://doi.org/10.1080/00029890.1997.11990723

Cobb, P. A., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction*, *21*(1), 1-78. https://doi.org/10.1207/S1532690XCI2101_1

Coleman, E. B., & Penuel, W. R. (2000). Web-based student assessment for program evaluation. *Journal of Science Education and Technology*, *9*(4), 327-342. https://doi.org/10.1023/A:1009432509448

Confrey, J., & Smith, E. (1995). Splitting, covariation, and their role in the development of exponential functions. *Journal for Research in Mathematics Education*, *26*(1), 66-86. https://doi.org/10.2307/749228

Drayton, B., & Puttick, G. (2016). Digital design of smart images: A design story. *Educational Designer*, *3*(9), 1-30. https://www.educationaldesigner.org/ed/volume3/issue9/article32/index.htm

Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data moves. *Technology Innovations in Statistics Education*, *12*(1), 1-25. https://doi.org/10.5070/T5121038001

ESTEEM (2022). https://fi.ncsu.edu/projects/esteem/

Feldman, A., Konold, C., & Coulter, B. (2000). *Network science: A decade later*. Lawrence Erlbaum Associates.

Finzer, W., Damelin, D., Heller, J., Wong, N., Konold, C., & Sandoe, J. (2019). *Common Online Data Analysis Platform (CODAP)*. In https://codap.concord.org

Ford, M. J. (2008). 'Grasp of practice' as a reasoning resource for inquiry and nature of science understanding. *Science & Education*, *17*(2-3), 147-177.https://doi.org/10.1007/s11191-006-9045-7

Gazal, R., White, M. A., Gillies, R., Rodemaker, E., Sparrow, E., & Gordon, L. (2008). GLOBE students, teachers, and scientists demonstrate variable differences between urban and rural leaf phenology. *Global Change Biology*, *14*(7), 1568-1580. https://doi.org/10.1111/j.1365-2486.2008.01602.x

Hoadley, C. (2002). Creating context: Design-based research in creating and understanding CSCL. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community* (Vol. 453-662). Routledge.

Howard, E., & Davis, A. K. (2004). Documenting the spring movements of Monarch butterflies with Journey North, a citizen science program. In K. S. Oberhauser & M. J. Solensky (Eds.), *The Monarch butterfly: Biology and conservation* (pp. 105-114). Cornell University Press.

Kesidou, S., & Roseman, J. E. (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review. *Journal of Research in Science Teaching*, *39*(6), 522-549. https://doi.org/10.1002/tea.10035

Konold, C., Finzer, W., & Kreetong, K. (2015). Modeling as a core component of structuring data. *Statistics Education Research Journal*, *16*(2), 191-212. https://doi.org/10.52041/serj.v16i2.190

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning*, *12*(3), 217-230. https://doi.org/10.1007/s10758-007-9123-1

Konold, C., & Higgins, T. L. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 193-215). National Council of Teachers of Mathematics.

Krajcik, J. S., Codere, S., Dahsah, C., Bayer, R., & Mun, K. (2014). Planning instruction to meet the intent of the Next Generation Science Standards. *Journal of Science Teacher Education*, *25*(2), 157-175. https://doi.org/10.1007/s10972-014-9383-2

Lehrer, R. (2012). *A learning progression emerges in a trading zone of professional community* and identity.

Lehrer, R., & English, L. D. (2018). Introducing children to modeling variability. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), International handbook of research on statistics education (pp. 229-260). Springer. https://doi.org/10.1007/978-3-319-66195-7_7

Lehrer, R., Kim, M.-j., & Schauble, L. (2007). Supporting the development of conceptions of statistics by engaging students in measuring and modeling variability. *International Journal of Computers for Mathematical Learning*, *12*(3), 195-216. https://doi.org/10.1007/s10758-007-9122-2

Lehrer, R., & Schauble, L. (2011). Designing to support long-term growth and development. In T. Koschmann (Ed.), Theories of learning and studies of instructional practice (pp. 19-38). Springer. https://doi.org/10.1007/978-1-4419-7582-9_2

Lehrer, R., Schauble, L., & Wisittanawat, P. (2020). Getting a grip on variability. *Bulletin of Mathematical Biology*, *82*, 1-26. https://doi.org/10.1007/s11538-020-00782-3

Manz, E. (2015). Representing argumentation as functionally emergent from scientific activity. *Review of Educational Research*, *85*(4), 553-590. https://doi.org/10.3102/0034654314558490

Manz, E., Lehrer, R., & Schauble, L. (2020). Rethinking the classroom investigation. *Journal of Research in Science Teaching*, *57*(7), 1148-1174. https://doi.org/10.1002/tea.21625

Mayer, A. (2010). Phenology and citizen science: Volunteers have documented seasonal events for more than a century, and scientific studies are benefiting from the data. *BioScience*, *60*(3), 172-175. https://doi.org/10.1525/bio.2010.60.3.3

Mills, M. G. L., & Gorman, M. L. (1997). Factors affecting the density and distribution of wild dogs in the Kruger National Park. *Conservation Biology*, *11*(6), 1397-1406. https://www.jstor.org/stable/2387366

Moss, P. A., & Schutz, A. (2001). Educational standards, assessment, and the search for consensus. *American Educational Research Journal*, *38*(1), 37-70. https://doi.org/10.3102/00028312038001037

National Academies of Sciences Engineering and Medicine. (2019). *Science and engineering for grades 6-12: Investigation and design at the center*. The National Academies Press. https://doi.org/10.17226/25216

National Academies of Sciences Engineering and Medicine. (2022). *Science and engineering in preschool through elementary grades: The brilliance of children and the strengths of educators*. The National Academies Press. https://doi.org/10.17226/26215

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

National Research Council. (2014). *Developing assessments for the Next Generation Science Standards*. National Academies Press. https://doi.org/10.17226/1840

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

Noushad, N. F., Shim, J., & Yoon, S. A. (2022). Navigating the conceptual and critical demands of data literacy in high school space using first hand data. In C. Chinn, E. Tan, C. K. K. Chan, & Y. Kali (Eds.), *Proceedings of the 16th International Conference of the Learning Sciences* (pp. 195-202). International Society of the Learning Sciences.

Nugent, J. (2017). Citizen Science: Navigate classroom citizen science throughout the school year with Journey North. *Science Scope*, *41*(1), 10-13. https://doi.org/10.2505/4/ss17_041_01_10

Odden, O. B., & Russ, R. S. (2019). Defining sensemaking: Bringing clarity to a fragmented theoretical construct. *Science Education*, *103*(1), 187-205. https://doi.org/10.1002/sce.21452

Oliver, M., McConney, A., & Woods-McConney, A. (2021). The efficacy of inquiry-based instruction in science: A comparative analysis of six countries using PISA 2015. *Research in Science Education*, *51*, 595-616. https://doi.org/10.1007/s11165-019-09901-0

Organisation for Economic Co-operation and Development [OECD]. (2022). *PISA 2022 Technical Report*. OECD Publishing.

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., Shear, L., Korbak, C., & Sparrow, E. (2005). The roles of regional partners in supporting an international Earth science education program. *Science Education*, *89*(6), 956-979. https://doi.org/10.1002/sce.20079

Penuel, W. R., & Means, B. (2004). Implementation variation and fidelity in an inquiry science program: An analysis of GLOBE data reporting patterns. *Journal of Research in Science Teaching*, *41*(3), 294-315. https://doi.org/10.1002/tea.20002

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

Pickering, A. (1995). *The mangle of practice: Time, agency, and science*. University of Chicago Press.

Puttick, G., Drayton, B., & Karp, J. (2015). Digital curriculum in the classroom: Authority, control, and teacher role. *International Journal of Emerging Technologies in Learning*, *10*(6), 11-20. https://doi.org/10.3991/ijet.v10i6.4825

Reiser, B. J., Novak, M., McGill, T. A., & 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

Rosenberg, J. M., Edwards, A., & Chen, B. (2020). Getting messy with data: Tools and strategies to help students analyze and interpret complex data sources. *The Science Teacher*, *87*(5), 30-35. https://www.jstor.org/stable/27048120

Rosenberg, J. M., Schultheis, E. H., Kjvelvik, M. K., Reedy, A., & Sultana, O. (2022). Big data, big changes? The technologies and sources of data used in science classrooms. *British Journal of Educational Technology*, *53*(5), 1179-1201. https://doi.org/10.1111/bjet.13245

Rubin, A. (2020). Learning to reason with data: How did we get here and what do we know. *Journal of the Learning Sciences*, *29*(1), 154-164. https://doi.org/10.1080/10508406.2019.1705665

Ruiz-Primo, M. A., Li, M., Wills, K., Giamellaro, M., Lan, M.-C., Mason, H., & Sands, D. (2012). Developing and evaluating instructionally sensitive assessments in science. *Journal of Research in Science Teaching*, *49*(6), 691-712. https://doi.org/10.1002/tea.21030

Sagrans, J., Mokros, J., Voyer, C., & Harvey, M. (2022). Data science meets science teaching. The Science Teacher, 89(3), 64-69. https://doi.org/10.1080/00368555.2022.12293671

Schmidt, W. H., Raizen, S., Britton, E. D., Bianchi, L. J., & Wolfe, R. G. (1997). Many visions, many aims: Volume II: A cross-national investigation of curricular intentions in school science. Kluwer. https://doi.org/10.1007/978-0-306-47208-4

Schultheis, E. H., & Kjelvik, M. K. (2015). Data nuggets: Bringing real data into the classroom to unearth students' quantitative and inquiry skills. *The American Biology Teacher*, *77*(1), 19-29. https://doi.org/10.1525/abt.2015.77.1.4

Schultheis, E. H., & Kjelvik, M. K. (2020). Using messy, authentic data to promote data literacy and reveal the nature of science. *The American Biology Teacher*, *82*(7), 439-446. https://doi.org/10.1525/abt.2020.82.7.439

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*(4), 475-522.

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.

Tatar, D. (2007). The design tensions framework. *Human-Computer Interaction*, *22*(4), 413-451. https://www.tandfonline.com/doi/abs/10.1080/07370020701638814

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

Wright, J. C., & Murphy, G. L. (1984). The utility of theories in intuitive statistics: The robustness of theory-based judgments. *Journal of Experimental Psychology: General*, *113*(2), 301-322. https://doi.org/10.1037/0096-3445.113.2.301

About the Authors

William R. Penuel 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*.

Andee Rubin is a mathematician, computer scientist, and learning scientist at TERC, an educational research and development non-profit; she has an undergraduate degree in mathematics and graduate degrees in computer science and artificial intelligence. She has been studying the growth of students' and teachers' reasoning about and with data for over 40 years, particularly as it is enabled by research-based tools for statistics education. She has developed software and curriculum materials around data-based reasoning for students as young as 6, both in and out of school (including in zoos and aquariums), as well as professional development materials for teachers and out-of-school facilitators. She is a member of the consensus study committee of the US National Academies of Science, Engineering and Medicine on fundamental competencies for the future of data and computing. In all her work, she is committed to empowering learners to use data to pursue social justice.

Gillian Puttick, Ph.D., has a broad background in biology and biology education and has worked as an educational researcher and curriculum developer at TERC for over 30 years. As a PI or co-PI, she has directed many federally funded education research and development projects, and focuses her efforts on bringing fresh scientific discoveries to formal and informal settings through designing, developing and testing curriculum, programs and activities for students and teachers. Over the past 20 years, her research and development work has focused primarily on ecology and climate change education.

Kate Henson 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.

Clarissa Deverel-Rico, Ph.D., is a science educator and research scientist at BSCS Science Learning, and has worked on phenomenon- and student-driven curriculum materials, supported teachers to take up equity and responsiveness in their instructional practices, along with developing and studying classroom assessment that support current science education reform initiatives. She has also partnered with teachers and created space to learn directly from students and families how classroom assessment can be spaces that sustain students' interests and identities. Dr. Deverel-Rico is committed to studying innovative approaches for how classroom assessment can support a vision of science education that prioritizes epistemic justice, agency, and student experience.

Penuel, W.R., Rubin, A., Puttick, G., Henson, K., Deverel-Rico, C. (2025) Balancing Design Goals for Supporting Students' Work with Extant Data in Science. *Educational Designer,* 5(19). ISSN 1759-1325
Retrieved from: http://www.educationaldesigner.org/ed/volume5/issue19/article78/