A Teaching Routine for Working With Existing Data in Science Classrooms

William R. Penuel, University of Colorado Boulder, william.penuel@colorado.edu
Andee Rubin, TERC, andee_rubin@terc.edu
Kate Henson, University of Colorado Boulder, kate.henson@colorado.edu
Gillian Puttick, TERC, gilly_puttick@terc.edu
Clarissa Deverel-Rico, University of Colorado Boulder, clarissa.deverelrico@colorado.edu

Abstract: Working with existing data is central to science investigations, but students and educators have generally not had experience using existing data sets to answer their own questions. We introduce a teaching routine that makes explicit critical steps in the process of working with data to gain insight into real-world phenomena. We intend the routine to support both curriculum developers and teachers in designing and enacting lessons to support students in engaging productively with scientific data, focusing on steps that are not commonly encountered in science classes.

Working with existing data in science classrooms

Exploring real-world phenomena using existing scientific data is a powerful context for engaging students in meaningful disciplinary practices in science. (NASEM, 2019). Working with already-collected data adds complexity to the already complicated task of data analysis. While research has identified many of the supports students need for working with data (Manz et al, 2020; Rubin, 2020; Feldman et al, 2000), working with data that was collected by someone else in a different context adds considerable complexity. In this paper, we report on work in progress on a teaching routine for developers and teachers seeking to support students in working with existing data in classroom investigations in science. Teaching routines are "recurring patterned sequences of interaction that teachers and students jointly enact to organize opportunities for student learning." (DeBarger et al., 2010, p. 225). Teaching routines can function as resources for curriculum writers as they construct sequences of lessons to engage students in science practices, as well as helping teachers adapt instruction to students' emergent ideas.

Project context: EMBEDS

The Exploring the Mathematics of Biological Ecosystems with Data Science (EMBEDS) project investigates the potential of integrating "data excursions" for developing high school students' competencies with data practices and data modeling into phenomenon-based instructional materials. These "data excursions" allow students to interact with datasets collected by scientists related to ecosystem dynamics to query their contexts, change the way they are aggregated and represented, and explore patterns they reveal.

An example of one such excursion we have developed takes place within a unit on ecosystems being developed for OpenSciEd, a free, phenomenon-based set of instructional materials aligned to the Next Generation Science Standards. The overarching unit question pertains to how the creation of the Serengeti National Park impacted local ecosystems. Early in the unit, students try to determine whether large increases in wildebeest and buffalo populations between 1960-75 could have been caused by an increase in available food. After learning that scientists had no data on food–but had to rely instead on rainfall as a proxy–students use CODAP (a free educational data analysis tool) to explore data on rainfall in different regions of the Serengeti during that time period.

As our project unfolded, we decided that it would be helpful to articulate a teaching routine to guide the design of our excursions and to help teachers support students' work with existing scientific data. The routine makes explicit aspects of investigating scientific phenomena that students do not typically encounter, such as the idea that a given dataset may not be able to answer the question at hand because it was collected for a different purpose or that measurements may have been made differently by different scientists.

The data routine

Our project is studying how a data routine can support sensemaking about phenomena in science. Our high-level conjecture guiding this inquiry is that:

> *Lessons (excursions) organized around the elements of the data routine can support students' ability to use existing multivariate datasets to help explain complex ecosystems phenomena.*

We further hypothesize that certain elements of the routine may be particularly consequential in their impact on students' facility with data, e.g. predicting what patterns in the data would imply particular answers, deeply investigating the way in which data were collected and how measures were defined, and working with a tool that facilitates the creation of multiple linked representations that support students' data fluency. We intend the Data Routine to be a resource that other design research projects might explore, particularly those connected to curriculum efforts in science, social science, and the emerging field of data science. An important set of questions pertains not only to its flexibility in other disciplinary contexts, but also what kinds of outcomes might be supported by its use, and what other embodiments might be necessary to achieve those outcomes.

Table 1
Elements in the Data Routine

Element	Student Actions
Framing Questions	● Students come to a consensus on question(s) to address related to the anchoring phenomenon and possible answers to their question(s). ● Students decide what kinds of data they need to answer their question(s). ● Students identify multiple plausible answers, and for each, students make predictions about what the pattern in the data would look like, if each answer were true.
Orienting to the Data	● Students orient to the data (what is each case? what are the variables?), discuss the source of the dataset, query who collected the data, by what methods and why, and evaluate its reliability, ● Students evaluate whether the dataset might help answer their questions, if it can answer a different related question instead, or is not relevant to their question.
Exploring the Data	● Students discuss ways to explore the data to help answer their questions. ● Students create initial representations and notice and record patterns they see. ● Students make initial claims based on patterns in the data, including how confident they are in the presence of variability.
Sensemaking about the Data	● Students present their claims and relevant representations to support them to others. ● The class engages in a discussion to decide what claims can be supported from the data and articulates limitations of the data. ● The class discusses whether they can reach consensus on an answer to the question(s) and identifies remaining and additional questions that arise from their analysis.

References

DeBarger, A. H., Penuel, W. R., Harris, C. J., & Schank, P. (2010). Teaching routines to enhance collaboration using classroom network technology. In F. Pozzi & D. Persico (Eds.), *Techniques for fostering collaboration in online learning communities: Theoretical and practical perspectives* (pp. 222-244). IGI Global.

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

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

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

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

Acknowledgments

This material is based in part upon work supported by the National Science Foundation under Grant Number DRL-2031468. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.