HOW ARE STUDENTS THINKING ABOUT DATA AND GRAPHING? FINDINGS FROM PRE- AND POST-INSTRUCTION ASSESSMENTS

Christina Siddons
University of Maine
christina.siddons@maine.edu

Franziska Peterson
University of Maine
franziska.peterson@maine.edu

Engaging with authentic data across all steps of the data investigation process – including collecting, processing, exploring, and visualizing data – is a critical skill set to build in K-12 students. Understanding the challenges students encounter throughout this process is key to offering instructional support to deepen their understanding and prepare students for navigating a data-rich world. Results from a pre-post instructional Data Assessment survey (n=202 from 7 classrooms) along with follow-up interviews (n=13) indicate students who have experience working with sensors appear to have a better approach to processing data. However, choosing the appropriate visualization for a given dataset poses challenges to students across classrooms.

Keywords: Middle School Education, High School Education, Integrated STEM/STEAM, Data Analysis and Statistics

Purpose of the Study

Data literacy is an essential competency to cultivate in K-12 students for them to be able to navigate a data-rich world in their future academic and career paths. Grounding data literacy skills in authentic, place-based learning experiences can be one way to increase student interest and buy-in as they are provided with opportunities to see how these skills translate to real-world scenarios within their local communities. We established a research-practice partnership (RPP; Henrick et al., 2023) with teachers and community members from rural coastal communities in the Northeastern U.S. to promote integration of ocean science, data, and technology-related competencies into STEM (Science, Technology, Engineering, and Math) subjects to explore the ways in which data literacy competencies can be developed authentically. This study investigates how students' conceptualizations around working with data changed after being involved in an integrated marine science project, guided by the following research question: How do students reason about the steps or processes involved in data investigations?

Theoretical Framework

Students have more meaningful learning experiences with data when it is grounded in relevant, real-world contexts (Kjelvik & Schultheis, 2019; Langen et al., 2014). Previous studies have emphasized the importance pedagogical practices that investigate STEM subjects authentically, such as using large, messy datasets and emphasizing the integrated nature of STEM subjects (Kelley & Knowles, 2016; NGSS Lead States, 2013; Roehrig et al., 2021). To translate these goals to classroom practices, our work is grounded in the six processes of data investigations proposed by Lee et al. (2022): Frame Problem, Consider and Gather Data, Process Data, Explore and Visualize Data, Consider Models, and Communicate and Propose Action. We supplemented this framework with attributes of "messy data" (Kjelvik & Schultheis, 2019) and opportunities to embed quantitative reasoning (Mayes, 2019) while teachers were developing data-rich, place-based learning experiences during Summer 2024.

Grappling with data across all six processes of the Lee et al. (2022) framework can be challenging for many students (Boaler et al., 2021). For example, "sanitized" datasets often used in classroom exercises do not offer opportunities for students to engage with noisy data, denying them the practice of engaging with the variability inherent in authentic data (Wilkerson et al., 2022). Additionally, when tasked with a graphing exercise, students can approach data with a "mold-in-mind" mindset where graph choice is dictated by students' preconceived visions of which graph is appropriate, regardless of data type (Keenhold, 2019).

Methods

Participants

Student participants were recruited from grade 7-12 classrooms enrolled in the project following Institutional Review Board approval for all recruitment and data collection activities. Matched pre- and post-instruction survey responses for this analysis were collected from 202 students (n = 12 middle school students, n = 190 high school students) from 7 participating teachers' classrooms at the start and end of the 2024-2025 academic year. Additionally, 13 students participated in follow-up interviews. Among survey respondents, 83 self-identified as male, 92 as female, 4 as non-binary or transgender, and 23 left this item blank. Students were enrolled in a variety of courses covering diverse STEM content areas including Earth Science, Biology, Ocean Science, Chemistry, Engineering, Algebra I, and Honors Precalculus. Classrooms were coded as sensor versus non-sensor classrooms depending on whether lessons developed by teachers included building and/or collecting data using sensors.

Data Collection and Analysis

Data was collected using a pre-post survey aimed at assessing student perceptions about data, what "counts" as data (Bargagliotti et al., 2020), and their thinking about data investigation processes. This Data Assessment survey was created by the research team in Year 2 of the grant and refined and iterated based on initial data collection for implementation in Year 3 (2024-2025 academic year). Importantly, three teachers involved in the project representing middle and high school, as well as math and science classrooms, were consulted during revisions to ensure language and concept accessibility across a wide array of students, as well as ensure broad applicability across classrooms addressing a diversity of subject matters. Although some survey items were retained from the initial version – 8 Likert statements to assess characteristics of useful or meaningful data (Peterson & Siddons, 2024) – later questions were revised to be more aligned with the six processes described by Lee et al. (2022). In this brief research report, we share our findings on selected Likert statements and two process questions, one on visualizing data (choosing between graph types) and one on processing data collected by a sensor. Summary statistics were analyzed using Excel and Tableau.

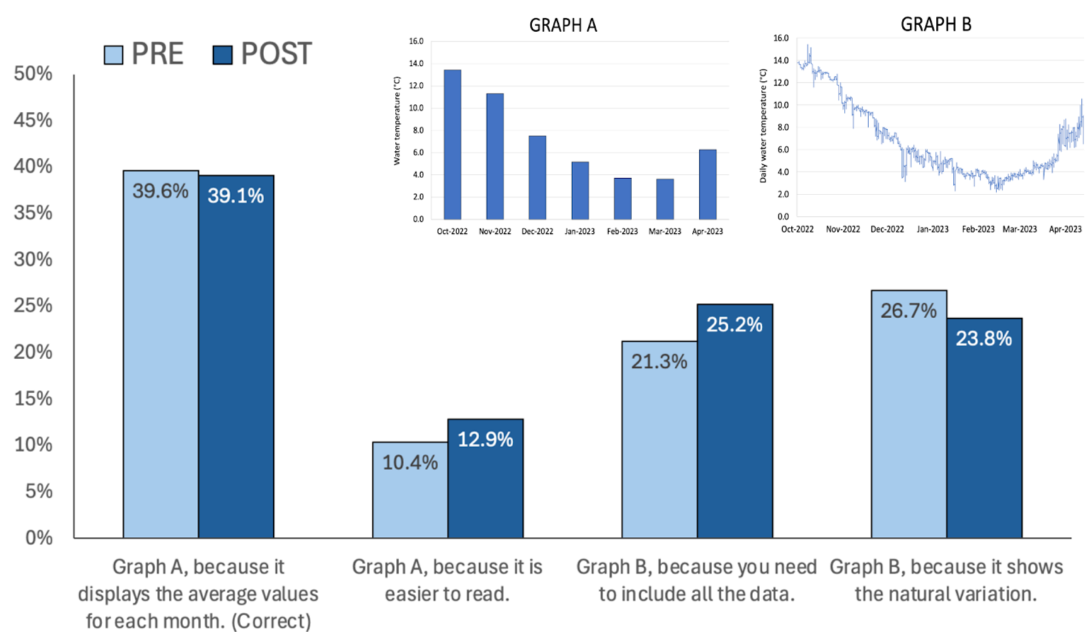
After a preliminary analysis of the pre-survey responses, the research team designed a follow-up interview protocol to further explore student reasoning in the context of broader data investigations conducted in the classroom. Student interviews were conducted across participating classrooms in Spring 2025, transcribed and analyzed through the lens of the guiding frameworks. All student names were replaced with pseudonyms.

Results

To statements about the usefulness of qualitative data, we saw no noticeable change between pre- and post-instruction responses. Students overwhelmingly agreed that qualitative sources represented useful data ("Descriptions or observations of the natural world and/or experiments can be useful data", 86.2% agreement; "Stories from people with local or historical knowledge can be useful data"; 77.2% agreement). Despite this high agreement, only 49% of students overall *disagreed* with the statement "Data are only useful if they are associated with a number (e.g., a count or a quantity)" in the pre-instruction survey shifting only slightly to an agreement of 53.5% in the post-instruction survey. More students selected *neutral* for this statement than for either of the other two Likert items mentioned above. The pre-instruction responses showed 35.6% of respondents selected neutral versus 28.7% in the post-instruction survey.

Students were asked to select an appropriate graph choice *and* appropriate reasoning for determining typical sea surface temperatures for the month of January (Figure 1, inset). Although ~40% of students chose the correct answer in both pre- and post-instruction results, nearly half chose the incorrect graph (Graph B; Figure 1). Even post-instruction, 12.9% of students chose the correct graph but incorrect reasoning ("Because it is easier to read").

Figure 1: Pre-post student survey responses to the graph choice item (*n* = 204)



Of the nine students who correctly selected Graph A during interviews, five gave similar lines of reasoning regarding the bar graph being easier to read or comprehend:

> So, if you wanted to show typical sea surface temperature, I would go with Graph A, because if you're only looking for the average slash typical, I would go with a simpler graph. It would help you visualize it better for your brain, and it also allows you to understand a bit faster. (Jordan, student interview participant)

All four student interviewees who chose Graph B mentioned that it showed more data or information than Graph A. Three of the four also mentioned that the variability of data shown in Graph B was pertinent to their selection process. These lines of reasoning reflect the reasoning in the survey answer choices for Graph B (Figure 1).

The data processing question provided students a scenario in which their class had collected salinity data for a week; students were tasked with selecting and ordering appropriate steps from a provided list. Steps were classified into six bins: Download Data (one step), Organize Data (two steps), Interrogate Data (three steps); Manipulate Data (one step); Visualize Data (one step), and finally, incorrect steps that should not be included (three steps). Amongst the matched student responses, 178 pre-instruction survey responses indicated ~52% selected Download Data as step one, with the next most choice Visualize Data (25.8%) (students who chose no steps were excluded). Post-instruction survey data had 82 matched students from sensor classrooms and 100 students from non-sensors classrooms after removing blank responses. Notably, 69.5% of students from sensor classrooms correctly identified Download Data as the first step, as compared to only 40% from non-sensor classrooms. Furthermore, only 14.6% of students from sensor classrooms chose Visualize Data as a first step, as compared to double the percentage in non-sensors classrooms (28.0%).

Interview data indicated 8 of the 13 participants agreed that downloading the data first was correct; six of these participants came from sensor classrooms. Students had a variety of reasons for their choice, including needing to assess the data first for utility or accuracy, a recognition that data may need to be manipulated or organized before graphing, or that the type of graph depends on the data. Students who agreed with graphing the data first (n = 4; 1 student agreed with both) were evenly split between sensors and non-sensors classrooms. All four had reasoning that indicated that visualizing the data made it easier to understand before anything else, as reflected in Taylor (student interview participant) saying, “basically, that's how you visualize all your data, because it's some people might be able to, but I can't just visualize if like, there's 100 different numbers. It's much easier, I think, to read a graph.”

Discussion & Conclusions

Survey results from Likert statements indicate that students’ theoretical assessments about the characteristics of useful or meaningful data did not shift by much pre- to post-instruction. However, application questions that required students to make practical decisions around how to handle data, whether in the processing or visualization steps, were more challenging. Based on our analysis from matched survey responses and supporting interview data, students continue to struggle with choosing the appropriate visualization for a given context. Even when the correct graph was chosen, interview data suggested that students selected the bar graph for the incorrect reason (i.e., it’s “easier to read”). While this was the least popular choice on the survey, it would be interesting to see if shuffling the order in which students see the answer choices for this question produces different results. Interview data also suggested that students who chose the incorrect type of graph may have been influenced by similar graphs created in class and/or because students believed that the line graph depicted “more” data and thus was better, failing to recognize the bar graph represented calculated averages, not *less* data.

Additionally, we saw an interesting divide in post-instruction survey responses when comparing sensors to non-sensors classrooms and what students identified as a first step in sensor data processing. Interview data revealed, however, that while the majority of students from sensor classrooms correctly identified downloading data as the first step, there were still those who thought graphing first was correct. Students also struggled to make meaning from data when presented in tabular or spreadsheet form. Importantly, while we recognize that visualizing data *can* improve understanding, these students still missed that the goal of the prompt was to think about *when* that visualization should occur. Thus, the “Processing Data” step (Lee et al., 2022) between collection and visualization seems to pose a challenge for many students, particularly as this would likely entail examining data in a table or spreadsheet format first.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 2148520. 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.

References

Bargagliotti, A., Franklin, C., Arnold, P., Johnson, S., Perez, L., & Spangler, D. A. (2020). *Pre-K-12 guidelines for assessment and instruction in Statistics education II: (GAISE II): A framework for statistics and data science education*. American Statistical Association. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf

Boaler, J., LaMar, T., & Williams, C. (2021). Making sense of a data-filled world. *Mathematics Teacher: Learning and Teaching PK-12, 114(*7), 508–517. https://doi.org/10.5951/mtlt.2021.0026

Henrick, E., Farrell, C. C., Singleton, C., Resnick, A. F., Penuel, W. R., Arce-Trigatti, P., Schmidt, D., Sexton, S., Stamatis, K., & Wellberg, S. (2023). *Indicators of research-practice partnership health and effectiveness: Updating the five dimensions framework*. OSF. https://doi.org/10.17605/OSF.IO/YVGCN

Keenhold, B. W. (2019). Assessing quantitative reasoning in a ninth grade science class using interdisciplinary data story assignments. (Publication No. 2963) [Master's Thesis, University of Maine]. https://digitalcommons.library.umaine.edu/etd/2963

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education, 3*(1). https://doi.org/10.1186/s40594-016-0046-z

Kjelvik, M. K., & Schultheis, E. H. (2019). Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. *CBE Life Sciences Education, 18*(2). https://doi.org/10.1187/CBE.18-02-0023/SUPPL_FILE/COMBINEDSUPMATS.PDF

Langen, T. A., Mourad, T., Grant, B. W., Gram, W. K., Abraham, B. J., Fernandez, D. S., Carroll, M., Nuding, A., Balch, J. K., Rodriguez, J., & Hampton, S. E. (2014). Using large public datasets in the undergraduate ecology classroom. *Frontiers in Ecology and the Environment, 12*(6), 362–363. https://doi.org/10.1890/1540-9295-12.6.362

Lee, H. S., Mojica, G. F., Thrasher, E. P., & Baumgartner, P. (2022). Investigating data like a data scientist: Key practices and processes. *Statistics Education Research Journal, 21*(2). https://doi.org/10.52041/serj.v21i2.41

Mayes, R. (2019). Quantitative reasoning and its role in interdisciplinarity. In B. Doig, J. Williams, D. Swanson, R. Borromeo Ferri, & P. Drake (Eds.), *Interdisciplinary mathematics education: The state of the art and beyond* (pp. 113–133). Springer International Publishing. https://doi.org/10.1007/978-3-030-11066-6_8

Next Generation State Lead States. (2013). *Next generation science standards: For states, by states*. Next Generation Science. https://www.nextgenscience.org/get-know-standards

Peterson, F., & Siddons, C. (2024). Student thinking and attitudes when working with data in an integrated, place-based curriculum. In K. W. Kosko, J. Caniglia, S. A. Courtney, M. Zolfaghari, & G. A. Morris (Eds.), *Proceedings of the forty-sixth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1652-1656). Kent State University.

Roehrig, G. H., Dare, E. A., Ellis, J. A., & Ring-Whalen, E. (2021). Beyond the basics: a detailed conceptual framework of integrated STEM. *Disciplinary and Interdisciplinary Science Education Research, 3*(1). https://doi.org/10.1186/s43031-021-00041-y

Wilkerson, M. H., Lanouette, K., & Shareff, R. L. (2022). Exploring variability during data preparation: a way to connect data, chance, and context when working with complex public datasets. *Mathematical Thinking and Learning, 24*(4), 312–330. https://doi.org/10.1080/10986065.2021.1922838