Engaging Young Learners With Data: Highlights From GAISE II, Level A

Leticia Perez¹ Denise A. Spangler² Christine Franklin^{3,4}

¹Graduate School of Education and Information Studies, University of California Los Angeles, Los Angeles, California, United States of America,

²Department of Mathematics, Science, and Social Studies Education, Mary Frances Early College of Education, University of Georgia, Athens, Georgia, United States of America,

³Department of Statistics Franklin College of Arts and Sciences, University of Georgia, Athens, Georgia, United States of America,

⁴American Statistical Association, Alexandria, Virginia, United States of America

Published on: Apr 30, 2021

DOI: <https://doi.org/10.1162/99608f92.be3c2ec8>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Column Editor’s Note: *What are effective ways to introduce our youngest learners to data science principles? How can we encourage the natural curiosity that children have about data and patterns in the world? In this column—part of an ongoing series on the new GAISE II guidelines for statistics and data science education—Leticia Perez, Denise Spangler, and Christine Franklin present activities that teachers in the elementary-level classroom can use with their students. We hope these activities will inspire school teachers, both to incorporate into their classes and to devise new ones of their own.*

Keywords: elementary school, classroom activities

For many people, the term ‘statistics’ conjures up images of mean, median, and mode and the struggle to keep track of what each summary represents and how it is determined (see [Chen, 2020](#)). But there is so much more to statistics or, more generally, data science, the interdisciplinary practice of extracting meaning from data to answer questions about the world (see e.g., [Wing, 2019](#)). Confident engagement with data is essential to being an educated member of society today, which compels us to start data science education as early as possible. The American Statistical Association and the National Council of Teachers of Mathematics recently released [Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education II \(GAISE II\): A Framework for Statistics and Data Science Education](#). (See the [previous column](#) for an overview). Here we highlight the key recommendations of *GAISE II* for young learners and use science-based lessons to illustrate the ways that they can powerfully engage with data.

The recommendations in *GAISE II* are divided into three levels—A, B, and C—corresponding roughly to elementary, middle, and secondary school. *GAISE II* recommends that Levels A students

develop data sense—an understanding that data are information. Students should learn that data are generated about particular contexts or situations and can be used to answer statistical investigative questions about the context or situation. They begin to learn how to interrogate data. (Bargagliotti et al., 2020, p. 22)

GAISE II describes the statistical problem-solving process, consisting of four elements: 1) formulate statistical investigative questions, 2) collect or consider already collected data, 3) analyze the data, and 4) interpret the results. Young students are full of questions about each other, the environment, and the world around them. They are capable of participating in the entire statistical problem-solving process—particularly formulating questions and interpreting the results.

Young children need opportunities to participate in statistical inquiry projects in which they interact with and reflect on different aspects of data, and it is important that they have experience with all four phases of the statistical inquiry process even if they only discuss them and do not actually carry them out. For example, in the current context of the COVID-19 pandemic, students might be curious about what is typical for students in

their class regarding the learning format (e.g., remote, in person, hybrid), number of people working remotely in their residence, type of remote learning device, or where they are learning (e.g., home, community center, coffee shop parking lot). The teacher then selects the question to investigate for which she has already prepared the lesson. Next, students can discuss how to collect data (e.g., video conferencing polling tools, applications that enable data collection, or sticky notes in person) or where to look for data collected by others. If students are going to collect data, they should discuss how to refine the data collection question to remove ambiguity or overlapping categories, the pros and cons of open-ended and multiple-choice questions, and how to efficiently collect data from everyone in the class. Again, the teacher can enact a preplanned data collection strategy, but there is value in having the students think through options initially. For the analysis stage of the framework, the teacher can guide students through a discussion of the various ways of organizing and representing the data before they enact one. The interpretation stage is where teachers can expand students' engagement with data and introduce the notion of inference. For instance, students can speculate how the data might be different if the data were collected from a different class of students (e.g., older, younger, at another school in the area, or a school elsewhere in the world) or at a different time (e.g., before/after the pandemic). These questions help students consider sources of variability that might influence the data. Students can draw inferences that go beyond the data set at hand. This context, for example, lends itself to explorations of equity issues, such as who has access to the internet and devices at home and who can more easily learn from home when schools are closed. Students could use this statistical investigation to debate whether schools will ever need to close due to weather or natural disasters in the future.

The notion of variability can be introduced and explored at each stage of the statistical problem-solving process. Whether it is the investigative questions posed, data collection methods, variation within the distribution of data, or the conclusion drawn, we want students to develop skills and approaches to anticipate and make sense of variation throughout the process. By framing investigations that leverage students' curiosity and require data as a part of the explanation, we provide opportunities for students to engage in statistical thinking. Science and social studies topics provide rich opportunities for students to make sense of data in meaningful ways. The [*Next Generation Science Standards*](#) (NGSS Lead States, 2013) identify the following science and engineering practices, all of which directly connect with data science:

- Asking questions (science) and defining problems (engineering)
- Planning and carrying out investigations
- Analyzing and interpreting data
- Using mathematics and computational thinking¹
- Constructing explanations (for science) and designing solutions (for engineering)

We next provide two examples of lessons that integrate science and data science to illustrate how the statistical problem-solving process can be carried out with Level A students.

Pumpkin Investigation for a Lower Elementary Class

As teachers seek to anchor curriculum within the context of student's everyday lives, they often orient lessons around family life, student preferences, and seasons, so we illustrate a lesson in which first graders are exploring pumpkins inspired by the children's book *How Many Seeds in a Pumpkin?* by Margaret McNamara (2007).

Formulate Statistical Investigative Questions

The teacher begins the lesson by showing photographs of several pumpkins growing in a field and encouraging student observations and questions about pumpkins and their seeds. After students share their experiences with pumpkins and other foods with seeds, the teacher presents his preplanned hypothesis to explore: Do bigger pumpkins have more seeds than smaller pumpkins?

Collect or Consider Data

He asks the students to brainstorm ways they could test the hypothesis, and together they refine their methods, including discussing how to determine pumpkin sizes and how to count the seeds, which reinforces the mathematical concepts of counting and place value. Students are put into teams and given a pumpkin out of which to scoop and count seeds, recording each seed using tally marks. The tallies are shared on a class chart labelled with columns for small and large pumpkins.

Analyze Data

During whole-class discussion, the teacher demonstrates counting the tallies by fives to get a total seed count for each pumpkin. She then constructs a pictograph to represent the number of seeds from each pumpkin (Figure 1). The graph is used to support student discussion around quantity as many first graders have not counted above 100. By exploring the pumpkins, counting the seeds, examining the photographs, and describing the tallies and the pictograph, students discover that there is no clear relationship between the size of the pumpkin and the number of seeds.

Figure 1. Pumpkin seed pictograph where one image of a seed represents 50 seeds from the students' pumpkins. Partial images show amounts of seeds less than 50.

Interpret Data

To help the class generate a list of variables that might affect the number of seeds in a pumpkin, the teacher directs their attention to the photograph of pumpkins in a field and asks the students to describe ways that pumpkins vary. The teacher then provides a photograph of a large and small pumpkin split in half. Students notice that the seeds grow along the rib of the fruit, so the teacher guides students to consider whether all pumpkins have the same number of ribs. Students count the ribs for their team's pumpkin and make a new chart to see whether the number of ribs predicts which pumpkins have more seeds. The teacher then reads *How Many Seeds in a Pumpkin?* by Margaret McNamara (2007), which reveals that generally, pumpkins with more ribs have more seeds, especially for pumpkins of similar size. Students conclude the task by comparing their team's pumpkin data to that of the main character in the book by sharing how their pumpkin varied in terms of size, seed number, and ribs. This investigation allows students to construct hypotheses and collect their own data to investigate the hypotheses, something that is rarely done with and by young children.

Salmon Investigation for an Upper Elementary Class

The context for this lesson is a third-grade classroom in the state of Oregon where students are learning about a key species in their local ecosystem by sharing the responsibility of growing salmon eggs until they are large enough to release in a nearby stream.

Formulate Statistical Investigative Questions

After the topic is introduced, students form many questions about the salmon and their environment:

- *How many fish will survive?*
- *Why are the rocks a certain size?*
- *I wonder if there has to be a certain water temperature?*
- *Once they get to teenager stage, do they leave their home?*
- *I wonder if any of these eggs are going to die?*
- *How many eggs can you find in a nest?*

Not all questions serve the data science and science lesson goals of the teacher, but it is valuable to engage the students in generating the questions as they help students explore what data are, where they might come from, and how data can be used to address real-world questions. The teacher's science content goal is to help students learn about the varied life cycle and survivorship (NGSS Lead States, 2013, 3-LS4-2), so she selects the question "How many eggs are found in a nest?" for the class investigation.

Collect or Consider Data

This investigation provides students with the opportunity to engage with data collected from an experiment or field study. To expose students to this type of data, the teacher provides teams of students with a nontraditional data source—a small set of photographs of scientists examining salmon nests—and asks them to brainstorm different types of information (data) the scientists could collect while in the field. The teacher then identifies the data available from four nests for analysis and interpretation: a close-up photograph of each nest, the number of eggs the scientist counted (378, 624, 843, 468), and a photograph of the nest location. The teacher provides an intentionally simplified but realistic data set so students can focus on data representation. Students are given question prompts to help them consider how to organize and keep track of the data from multiple nests.

Analyze Data

Student teams consider the data in the photographs and represent the data using methods of their choice. See Figure 2 for a case-value plot created by the teacher and note that this representation engages students in rounding to the nearest hundred. Next, she asks, "What do you notice? What do you wonder?" to help students identify convergent thinking around patterns and identify misunderstandings.

Figure 2. Case-value plot (often called a bar graph in curriculum materials) for number of eggs in the first four nests A-D (created using Google slides).

After this discussion, the teacher introduces the challenge question, “What is the typical number of eggs found in a nest?” to encourage students to go beyond summarizing the data at hand. To better reflect messy data in the real world, students are given data from six more nests (with 747, 509, 635, 701, 684, and 577 eggs). The teacher lets teams grapple with the larger data set and discuss what their case-value plot would look like if these nests were added. To help students delve deeper into “the typical number of eggs” question, she provides an interactive histogram, showing the data from the four original nests in grey, for students to complete (see Figure 3). Students classify the nest-count data into groups or bins representing a range of values. It is a big shift for students to move from thinking about individual observations, in this case the number of eggs in a nest, to the number of nests with a particular number of eggs in them. To help students make sense of how the case-value plots and histograms are related, the teacher asks them to look at their original case-value plot

(Figure 2) and the grey boxes on the histogram representing the first four nest counts (Figure 3) and talk about how the same data can be represented in very different ways.

Figure 3. A drag-and-drop histogram for visualizing the larger data set (created in Google slides).

After the students are comfortable with the representations, the teacher redirects their attention to the “typical number” question. Students tend to gravitate to the tallest or longest bar in a histogram to choose the typical value, but the availability of the student-generated and teacher-generated graphs provides opportunities to consider other ways of coming up with a typical number and whether the number they picked is the most ‘fair’ depiction of the typical number of salmon eggs. The teams decide whether they will report a single number or a range of numbers and practice explaining their reasoning to each other. In explorations such as this one, students learn that data are often messy, some representations may result in the loss of information, and people can work together as a community to use mathematical and statistical thinking to deal with messy data.

Interpret Data

In the final step of the statistical problem-solving process, the teacher pushes students to think about where variability played a role in this investigation and to identify possible sources of variability in future investigations. She asks students if the answer they obtained for the typical number of eggs in a salmon nest implies there are always that many eggs. When students say no, she asks what might affect the number of eggs, which ties nicely to the science objectives of the lesson (NGSS 3-LS4-3). Students conjecture that the nest location in the stream, the time of year, the number of salmon spawning in the stream, whether it rains and the resulting strength of the flow of water, and the type of salmon in the stream might affect the number of eggs in a given nest. Further, they wonder if the number of eggs in Oregon nests differs from the number of eggs in Washington, Alaska, or British Columbia nests. As part of the discussion, one student notes that it was difficult

to count the eggs in the photographs because the eggs were massed together and on top of each other, which gives the teacher an opportunity to talk briefly about another source of variability—measurement error.

This lesson snapshot demonstrates how statistics can be integrated with science content, the use of nontraditional data, and the ways elementary age children can engage with messy data.

Conclusion

To meet the goal of having data savvy adults who are able to leverage their statistical knowledge to fully participate in society, it is imperative that we engage students in data science from a young age. Although data science does not appear formally in the curriculum until upper elementary or middle grades in many states, teachers can regularly incorporate parts of the statistical problem-solving process into existing curricula, particularly in social studies and science. Across the curriculum, teachers can help students look for questions that can be answered with data and invite students to brainstorm questions and possible approaches for investigating those questions. When engaging young children with data, teachers often will need to provide curated data or construct artificial situations with data that are age appropriate mathematically. However, it is important to provide students with opportunities to realize that data are usually messy and to engage with messy data. At the heart of all data science is the notion of variability and the effort to describe or quantify variability in some useful way. It is crucial to provide students with opportunities to go beyond seeking answers to a single statistical question or a single answer to a messy data problem. Throughout a statistical investigation, students' curiosity can generate additional questions for future investigation, highlighting the role of inquiry in data science.

To help teachers and teacher educators engage children in developing data sense, *GAISE II* contains multiple examples of developmentally appropriate statistical investigations. Additional support for teachers can be found in [*Statistical Education of Teachers*](#) (American Statistical Association, 2015) and on the [American Statistical Association web site under the Education tab](#).

Disclosure Statement

Leticia Perez, Denise A. Spangler, and Christine Franklin has no financial or non-financial disclosures to share for this article.

References

American Statistical Association. (2015). *Statistical education of teachers*. Arlington, VA. <https://www.amstat.org/asa/files/pdfs/EDU-SET.pdf>

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). *Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education (GAISE) report II*. American Statistical Association and National Council of Teachers of Mathematics.

<https://www.amstat.org/asa/education/Guidelines-for-Assessment-and-Instruction-in-Statistics-Education-Reports.aspx>

Chen, A. (2020). High school data science review: Why data science education should be reformed. *Harvard Data Science Review*, 2(4). <https://doi.org/10.1162/99608f92.1e28ce9e>

Franklin, C., & Bargagliotti, A. (2020). Introducing GAISE II: A guideline for precollege statistics and data science education. *Harvard Data Science Review*, 2(4). <https://doi.org/10.1162/99608f92.246107bb>

McNamara, M. (2007). *How many seeds in a pumpkin?* Schwartz & Wade.

NGSS Lead States. (2013). *Next generation science standards: For states, by states* (Appendix F). NextGenScience. <https://www.nextgenscience.org/pe/3-ls4-3-biological-evolution-unity-and-diversity>

Wing, J. (2019). Data life cycle. *Harvard Data Science Review* 1(1). <https://doi.org/10.1162/99608f92.e26845b4>

©2021 Leticia Perez, Denise A. Spangler, and Christine Franklin. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.

Footnotes

1. Specifically, data science uses inferential thinking and computational thinking. ➞