Received: 1 December 2021 | Accepted: 20 May 2022

DOI: 10.1111/bjet.13245

ORIGINAL ARTICLE

British Journal of Educational Technology BERA

# Big data, big changes? The technologies and sources of data used in science classrooms

**Joshua M. Rosenberg**[1] | **Elizabeth H. Schultheis**[2,3] | **Melissa K. Kjelvik**[3] | **Aaron Reedy**[4] | **Omiya Sultana**[1]

[1]Department of Theory and Practice in Teacher Education, University of Tennessee, Knoxville, Knoxville, Tennessee, USA

[2]Michigan State University, Hickory Corners, Michigan, USA

[3]Department of Plant Biology, Michigan State University, East Lansing, MI, USA

[4]DataClassroom, Charlottesville, Virginia, USA

**Correspondence**
Joshua M. Rosenberg, University of Tennessee, Knoxville, 1122 Volunteer Blvd., Knoxville, TN 37996, USA.
Email: jmrosenberg@utk.edu

**Funding information**
National Science Foundation, Grant/Award Number: DEB1832042

## Abstract

With improving technology and monitoring efforts, the availability of scientific data is rapidly expanding. The tools that scientists and engineers use to analyse data are changing in response. At the same time, science education standards have shifted to emphasize the importance of students making sense of data in science classrooms. However, it is not yet known whether these exciting new datasets and tools are used science classrooms, and what it would take to facilitate their use. To identify opportunities, research is needed to capture the data practices currently performed in classrooms, and the roles of technology for student learning. Here, we report findings from a survey conducted in the United States of 330 science teachers on the data sources, practices and technologies common to their classroom. We found that teachers predominantly involve their students in analysing relatively small data sets that they collect. In support of this work, teachers tend to use the technologies that are available to them—namely, calculators and spreadsheets. In addition, we found that a subset of teachers used a wide variety of data sources of varying complexity. We discuss what these findings suggest for practice, research and policy, with an emphasis on supporting teachers based on their needs.

KEYWORDS

analysing and interpreting data, data literacy, data science, science education, survey research methods

**Practitioner notes**

What is already known about this topic

- Collecting and analysing data are central to the practice of science, and these skills are taught in many science classrooms at the pre-collegiate (grades K-12) level.
- Data are increasingly important in society and STEM, and types and sources of data are rapidly expanding. These changes have implications for science teachers and students.

What this paper adds

- We found that the predominant data source science teachers use is student-collected, small data sets.
- Teachers use digital tools familiar and available to them: spreadsheets and calculators.
- Teachers perceive the cost and time it would take to learn to use digital tools to analyse data with their students as key barriers to adopting new tools.
- Despite the predominance of small, student-collected data analysed using spreadsheets or calculators, we also found notable variability in the data sources and digital tools some teachers used with their students.

Implications for practice and/or policy

- Many of the changes called for in science education standards and reform documents, regarding how students should collect and analyse data, have not yet been fully realized in pre-collegiate classrooms.
- Science teacher educators and science education researchers should build curricula and develop digital tools based on which kinds of data sources and digital tools teachers presently use, while encouraging more complex data useage in the future.

# INTRODUCTION

*Landsat 9* is a NASA-developed satellite that records satellite imagery covering all of Earth's surface every 16 days (NASA, 2021). Using the imagery recorded by *Landsat 9*, *LandsatLook* (https://landsatlook.usgs.gov/explore) allows scientists—and, perhaps more important, anyone, including children—to view an image of any location on Earth at any specified time (LandSatLook, 2021). This represents just one exciting area in science where the type, size, and accessibility of scientific data available are rapidly expanding.

Not limited to the Earth and space sciences, the type (images), size (very large) and the update frequency—of the data generated by sensors and other technologies are becoming commonplace across a range of scientific domains. For instance, the development of mRNA vaccines for COVID-19 was facilitated by the rapid sequencing and sharing of the genome of the SARS-CoV-2 virus (National Human Genome Research Institute, 2021). Large, collaborative physics experiments such as those that use particle accelerators generate *petabytes* (1000 of terabytes) of data per year (CERN, 2021). In addition, anyone with a smartphone can access or contribute to a collection of nearly one hundred million images of living organisms (California Academy of Sciences and National Geographic Association, 2021); and, smartphone applications make it possible to classify and log an unknown organism as one of more than one-third of a million living organisms in the *iNaturalist* database.

The availability of new types of data invites many scientific questions (Boyd & Crawford, 2012)—and questions for science teaching and learning, too (Krajcik & Mun, 2014; Lee & Wilkerson, 2018; National Academies of Sciences, Engineering, and Medicine, 2019). For one, to what extent do science teachers and learners have access to the kinds of data tools and methods that are increasingly used across varied scientific disciplines? How do science standards reflect the changes in the kinds of data used in contemporary scientific practice? And, what support do teachers need to enable them to engage students in science and engineering practices that are related to making sense out of data?

Motivated by the changing nature of scientific data, calls in recent science education reform documents, and advances in technology, this paper aims to describe which data sources and data analysis technologies teachers use with their students in their classrooms. To do so, we report on findings from a survey conducted in the United States of science teachers on the data sources, practices, and technologies common to their classroom.

## WHAT ADVANCES IN ANALYSING DATA SUGGEST FOR SCIENCE TEACHERS AT THE K-12 LEVEL

We were motivated by several advances related to the availability of data and digital tools available in science classes. We elaborate on two of these—science education standards and in technologies pertinent to analysing data—by specifying what these advances suggest for how teachers support their students to work with data in K-12 classrooms.

### Science education reform and improvement initiatives

New science education standards have been developed that call for students to engage in a range of data-related practices including analysing and interpreting data, planning and carrying out investigations, and using evidence to construct explanations and arguments (ACT, Inc., 2014; College Board, 2019; NGSS Lead States, 2013). These reforms seek to shift the focus from memorization of content and toward the acquisition and application of skills within meaningful contexts. For instance, in the United States, the Next Generation Science Standards (NGSS; NGSS Lead States, 2013), ACT College Framework (ACT, Inc., 2014), and AP Biology Curriculum Framework (College Board, 2019) focus on students' engagement with the aforementioned science and engineering practices—a shift from traditional content-based learning goals and from standards that consider practices (such as analysing or interpreting data; referred to broadly as “inquiry” in the past) within separate standards (National Research Council, 2012; Rudolph, 2019). Additionally, quantitative reasoning and scientific practices are explicitly linked (Kjelvik & Schultheis, 2019). For example, the NGSS states that the standards “aim to give middle school and high school science educators a clear road map to prepare their students for the quantitative demands of college and careers, where students need to apply quantitative tools in an applied or scientific context” (NGSS Lead States, 2013, p. 137).

Data-related practices are not only prominent in science education standards—and are not only prominent in the United States. The Common Core State Standards for mathematics and literacy (CCSS; National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010) focus on quantitative reasoning, using mathematical tools, and solving problems in applied—rather than decontextualized—contexts (Mayes & Koballa, 2012; National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). Other disciplines, such as computer science, have long-emphasized data-related knowledge and skills in standards (Computer Science Teachers

Association, 2017); some argue that analysing data should be the entry point into learning about computing (Krishnamurthi & Fisler, 2020). Furthermore, nations and organizations around the world are investing in what has come to be termed *data education* or *data science education*, such as in the European Union (see *the Digital Education Action Plan* (https://education.ec.europa.eu/focus-topics/digital-education/about/digital-education-action-plan) and its emphasis on data-intensive technologies), Scotland (see the *Data Education in Schools* program, https://dataschools.education/). Other nations, such as New Zealand, have long emphasized the importance of data and statistics in the pre-collegiate curriculum (Forbes, 2014), while other nations have—until recently—historically relegated such topics exclusively to college-preparatory courses. Finally, work with data has quietly entered existing disciplines in quiet, yet meaningful ways. For instance, citizen science initiatives have been prominent in the United Kingdom (see Tweddle et al., 2012) and through global networks used around the world (Sullivan et al., 2014). In this way, work with data represents a relatively uncommon point of overlap between science and mathematics standards and across international borders. Common across these calls is that the data students analyse can be relevant to students; analysing relevant data can call on students to use a range of capabilities to answer questions or solve problems, applying ideas from science, mathematics, or computer science in the process (Kjelvik & Schultheis, 2019; Lee, Wilkerson, et al., 2021; Rosenberg, Lawson, et al., 2020).

These are not *entirely* new calls to emphasize more meaningful engagements with data for learners (National Research Council, 2006), though they are different in their emphasis on digital tools and data sources that are larger and potentially more relevant to learners. Moreover, even in the context of the recent development of the NGSS, these calls are being renewed: A recent report that builds on both the National Research Council's (2006) call for better laboratory experiences for learners *and* the research undergirding the NGSS states simply that "working with data is at the heart of science investigation" (National Academies of Sciences, Engineering, and Medicine, 2019, p. 280). This report also calls for more research on how students work with data in science classrooms as well as how learning technologies can be adapted for science teachers and learners.

The NGSS calls for teachers to engage students in varied, ambitious practices related to data. However, the extent to which these are commonplace in science classrooms is unclear, with the *National Survey of Science and Mathematics Educators* (NSSME+; Banilower et al., 2018) a notable exception. This nationally representative survey included a question that asked teachers about how frequently they engaged their students in a range of practices. Banilower et al. (2018) reported that two practices were the most common: organizing or representing data using tables or graphs (enacted by 34–58% of teachers [depending on grade level taught] weekly or more frequently) and identifying patterns, trends or relationships in data (34%–47% of teachers). Notably, other practices were much less commonplace: using mathematical models (12%–26% of teachers), for instance, occurred with a frequency less than half that of organizing or representing data using tables or graphs. This information on what science teachers do to support the use of data in the classroom offers an initial portrait of the current use of data in classrooms, but they are also limited in several key ways.

While informative as to the practices teachers engage students in, the results of the NSSME+ do not speak to the *sources* of data that teachers bring to their students; which is becoming more important to understand because of the changing nature of scientific data and the potential challenges associated with accessing and analysing different types of data or data that is much larger than the data typically used by students. In addition to not speaking to the sources of data teachers use, the NSSME+ results do not interrogate the roles of technology, as summarized in the next section.

## Digital sources of data and digital tools for analysing data

Traditional use of data in the classroom relies on pen-and-paper, which has been an effective platform to help students develop their data literacy abilities, including the ability to organize and analyse data in tables, and create data visualizations using graphs. Yet, with datasets expanding in size and complexity, technology may become necessary to help students utilize these new types of data. The same practices done on paper can be performed using digital tools as well, however the learning opportunities using each modality may differ (Gardner et al., 2021; Kjelvik & Schultheis, 2019; Schultheis & Kjelvik, 2020). For example, a study on student graphing in paper and digital environments found that student graph quality was similar across platforms, however students found it easier to graph in digital environments and were better able to identify the appropriate variables for the graphs (Gardner et al., 2021). Compared to analysing data using pen-and-paper, an advantage of digital tools is that they allow learners to quickly analyse large datasets that are generated by a myriad of sources (Kahn & Jiang, 2021; Lee & Wilkerson, 2018; Rosenberg, Lawson, et al., 2020). Yet, the increased ease that some analysis and visualiations tools provide (eg, the charts and graphs functionality in Google Sheets that suggests suitable visualizations for users) may support students to quickly generate output without having thought through that kind of chart or graph they would like to create—and why.

Technology plays an integral role in how teachers can support their students to engage with data in more sophisticated ways (Krajcik & Mun, 2014; Lee & Wilkerson, 2018; National Academies of Sciences, Engineering, and Medicine, 2019). Within the NGSS, the science practice of analysing and interpreting data describes that “modern technology makes the collection of large data sets much easier, providing secondary sources for analysis” (NGSS Lead States, 2013, Appendix F, p. 23); by grades 9–12, students should “analyze data using tools, technologies, and/or models (eg, computational, mathematical) in order to make valid and reliable scientific claims” (p. 23). In their review on how secondary science teachers integrate data into their science teaching, Lee and Wilkerson (2018) suggest that new means of analysing data given advances in technology could involve students (in science classes) “manipulating moderately large sets of data (hundreds to thousands of data points), using algorithmic processes and instructions implemented through digital tools” (p. 31). A recent consensus report underscores the importance of technology to students' data analysis efforts. The National Academies of Sciences, Engineering and Medicine (2019) report articulates a key role for technology in contemporary science classrooms: “[Technology] can be used for data collection, as a source of data, for data analysis, for modeling, for visualization, for simulations, and for presentations” (p. 168).

There is far from one tool that all science teachers are likely to use. More broadly, there is not one tool ideal for teaching and learning to work with data (McNamara, 2018). Bhargava and D'Ignazio (2015) argue that digital technology that supports learning how to work with data is not necessarily a tool that maximizes ease of use, but rather one that makes the operations explicit so that learners understand exactly what is being done and why (Bhargava & D'Ignazio, 2015). In other words, the best tool for teachers and learners may not be the best tool for others, including professional data analysts and data scientists. Instead, the best tool may be one that highlights (and invites learners to explicitly work through) key steps in the data analysis process. An ideal data tool would view the audience as learners rather than (or in addition to) users (Bhargava & D'Ignazio, 2015). Moreover, what is not yet known is what digital tools science teachers use—or what would prevent them from using a new tool.

Apart from specific digital tools, such as the statistics education digital tools that around two-thirds of AP statistics teachers use (Lee & Harrison, 2021) as well as spreadsheets and calculators, we can consider the role of technology in a broader sense, as a *modality* that can be contrasted with a pen-and-paper approach to analysing data in K-12 educational contexts.

Research that highlights some of the differences between environments when learning to work with data in pen-and-paper and digital modalities indicates that there are specific learning benefits and weaknesses of each environment (Gardner et al., 2021). Making the transition from working with small, self-collected datasets to larger datasets typical of what is collected at scale may be a valuable skill for learners (Kjelvik & Schultheis, 2019; Wolff et al., 2019). Thus, in addition to knowing which specific tools for analysing data science teachers use, knowing about what modalities are common could assist those developing curricula, tools, or professional learning opportunities for teachers to know how extensive the use of digital tools is relative to the use of the more traditional pen-and-paper approach.

In addition to not speaking to the sources of data teachers use (as described in the previous section of the literature review), the NSSME+ results do not interrogate the specific roles of technology. More broadly, the NSSME+ does not speak to the modalities used by teachers. Some research has pointed to differences in the discussions students have while analysing data they collect (or, first-hand data) and data collected by others (second-hand data)—we know, for instance, that first-hand data may be especially generative for students' learning about measurement (Hug & McNeill, 2008). However, understanding how pen-and-paper and digital modalities may have benefits or drawbacks that influence teachers to use particular data sources may also be informative, particularly given the importance of the larger second-hand data sources increasingly available to teachers and students.

# THE PRESENT STUDY

Changes in science education standards and technologies for the integration of data in instruction suggest that documenting what data sources and technologies teachers use with students can be important at this time, especially given the absence of information on the particulars of what teachers do to support this central element of both science and science learning. To guide our investigation into the data sources and technologies common to K-12 classrooms, we ask the following three research questions (RQs):

- RQ #1: Which sources of data do science teachers select for their students to use?
- RQ #2: Which digital tools for analysing data do teachers use to facilitate their students' work with data?
- RQ #3: And, how does the modality (pen-and-paper or digital) of students' work with data differ based on the type of data source selected by teachers?

We work to answer these questions overall and based on the grade level (elementary, middle, and high) of teachers.

# METHOD

We describe how we used a survey research methodology. We begin with the instrument we developed, followed by details on the data collection, sample and data analysis strategy.

## Instrument

We developed a survey instrument that consisted of 11 questions on the use of data in the classroom and 15 demographic questions.[1] We used available questions (from the NSSME+; Banilower et al., 2018) when they were available. However, for many of the aspects of science teachers' use of data with their students, relevant questions from the NSSME+ or the wider literature we reviewed during the instrument development phase of the study did not exist, and so we developed new questions. Notably, we only analysed responses to questions we developed for this study—even though the broader survey we developed included NSSME+ questions). We undertook a pilot testing process with science educators and experts in survey research methods and quantitative methods to provide some initial validity evidence for these questions; this involved the pilot testers answering and providing input on each question, with a focus on their validity with respect to the sources, tools, and modalities of teachers' data use. Pilot testers' feedback helped us to revise the survey and to generate some initial validity evidence, though we note that we hope for future to further establish the validity and reliability properties of the questions we developed. The entire survey is made available on the Open Science Foundation website.[2]

This survey focused on teachers' self-reported gaps in currently available resources, challenges with pen-and-paper and digital activities related to classroom work with data, features that are conducive to classroom learning, and supports they think their students need to improve in data literacy outcomes. The subset of questions we used for this study and their source (NSSME+ or self-developed) is presented in Appendix A.

## Data collection

We administered the survey to a national (United States) sample of science educators in February 2020 in three ways. First, we contacted all of the science teachers who were a part of the mailing list for Data Nuggets (https://datanuggets.org) (Schultheis & Kjelvik, 2015, 2020), a National Science Foundation-funded program with ready-to-teach lessons on data literacy for science education. This mailing list consists of over 11,000 K-16 educators who voluntarily provided their email address through a pop-up on the Data Nuggets website. The second means was through users of DataClassroom (Mayes et al., 2020), an educational technology platform to assist teachers and students in analysing scientific data in classroom contexts. The mailing list consisted of approximately 8500 people—primarily educators. Finally, we shared the survey via social media posts on Twitter and in a Facebook group for Data Nuggets users. To incentivize participation, we communicated that the first 100 respondents would receive $10 for completing the survey. We compensated those first 100 respondents and promptly notified and thanked those who were not among the first 100 respondents, noting that we would contact them first in any subsequent, compensated survey research opportunities.

In total, we received 623 responses. For all 623 responses, participants consented to participate in the study, though many of these surveys were incomplete (see Figure S1). This response rate is relatively low, but in line with what others have found in research and marketing contexts. Specifically, we consulted both the peer-reviewed and “grey” literature (eg, white papers, blog posts and pre-prints) and found that a response rate of around 5%–30% is normative, with lower response rates expectable if those sending the request to complete the survey have interacted with those receiving the survey less; some business report response rates as low as 1%–3%. In total, 327 individuals from mailing lists and three from social media completed the entire survey; these 330 individuals and their responses comprised our sample and data set.

## Sample

This sample had a national presence, but was not strictly representative of the national population, largely because we relied on two mailing lists for most of our participants. For this reason, our sample was one of convenience (Fink, 2015). However, because of how widely used the tools associated with the mailing lists are, we note that the sample is far from homogeneous in terms of factors such as teachers' locations and years of teaching experience. Our sample of 330 responses consisted of responses from 260 (78.8%) women and eight (2.4%) who either did not respond or preferred not to answer, which very roughly corresponds with national statistics (Banilower et al., 2018): 94%, 71% and 57% women at the elementary, middle and high school levels (0% other).

Survey respondents came from 43 U.S. states, with the most teachers coming from California ($n = 29$) followed by Massachusetts ($n = 19$) and Wisconsin ($n = 19$). On average respondents had 18.4 ($SD = 8.51$) years of teaching experience. This is roughly comparable between grade levels, with 12% < 2 years, 14% 3–5 years, 17% 6–10 years, 37% with 11–20 years, and 20% with greater than 21 years teaching any subject at the K-12 level (these were roughly comparable for years of experience teaching science).

10.9% of respondents identified as under-represented in their work. Nine (2.7%) respondents identified as Black or African American, four (1.2%) as Hispanic, two (0.6%) as Asian, two (0.6%) as American Indian and one (0.3%) as Muslim. Nationally, 9%, 7% and 6% of science teachers identify as Hispanic or Latino, and 8%, 8% and 5% identify as Black or African American, 2%, 2% and 5% identify as Asian and 1%, 1% and 2% identify as American Indian (Banilower et al., 2018). As the NSSME+ notes, “Black, Hispanic, and Asian teachers continue to be underrepresented in the science teaching force” (p. 7).

At last, we compared our respondents to those in a nationally representative survey, the NSSME+, to gain insight into how our sample may or may not be nationally representative. In particular, we used the NSSME+ question, “How often do you have students do each of the following in your class?”, with the nine data analysis-related tasks, in our survey. In the 2018 report of the NSSME+, three tasks were reported to be the most common that teachers have students do: (1) Organize and/or represent data using tables, charts or graphs to facilitate analysis, (2) analyse data using grade-appropriate methods to identify patterns, trends or relationships and (3) determine which variables from a provided dataset are necessary to answer a scientific question. The NSSME+ summarizes how often teachers report they have their students do these tasks at least once per week—and how often they never have their students do this task by grade level (as this is how the NSSME+ report these results). Because these are not central to the results of this study, we report the percentages for our and the nationally representative NSSME+ respondents in Appendix C.

These statistics suggest that our sample is generally comparable to the NSSME+ along the lines of the frequency with which teachers engage their students in data analysis; for the first question, on organizing/representing data, our sample's middle and high school teachers reported having their students do this slightly less than teachers in the NSSME+ sample (45% and 51% for the middle and high school teachers in our sample compared to 49% and 58% for the NSSME+ sample respectively), whereas our sample's elementary teachers did this more (53% relative to 34%). The responses were generally comparable for the question on analysing data using grade-appropriate methods, whereas our sample's middle and high school teachers reported having their students determine necessary variables less than teachers in the NSSME+ (while our elementary teachers reported having their students do this at a rate comparable to those surveyed in the NSSME+).

The characteristics of the teachers in our sample, then, are not strictly nationally representative, though we have taken some exploratory steps to understand how the teachers we surveyed are similar to the population of science teachers across the United States with respect to the frequency with which they engage students in data analysis, though our sample of teachers is also different in several key ways. Namely, there was an underrepresentation of Hispanic teachers. Also, having drawn from the mailing lists of a data analysis-related resource for science teachers (DataClassroom) and a tool for teachers to use with their students (DataClassroom) may be biased toward teachers already inclined to supporting their students to analyse scientific data. Thus, any generalizations from our findings must be made with the characteristics of the teachers we surveyed in mind, as the sample was not designed or constructed in such a way so as to be nationally representative.

## Data analysis

### Quantitative data analysis

For quantitative data analysis, we used descriptive methods, as our goal was to provide a portrait of the state of analysing data with technology, rather than to make inferences about these levels or to test hypotheses about them. Specifically, for each of the questions with fixed response options, we calculated the sum of the responses overall and by grade level (elementary, middle, and high school). For questions with scale response options, we calculated the mean and standard deviation of the responses overall and by grade level.

### Qualitative data analysis

For qualitative data analysis, we used thematic analysis (Hatch, 2002). Specifically, we first open-coded the responses to the two open-ended questions on the survey (describing a typical data analysis activity and describing sources of data used with students). Then, we developed two initial coding frames corresponding to the responses for each of the two open-ended questions; these codes were not mutually exclusive, as multiple codes could—and often were—applied to the same responses.

Next, we engaged in multiple rounds of coding to establish inter-rater agreement. To do so, a random sample of 20 responses was selected for each round of inter-rater agreement. Then, both coders coded the responses using the coding frame after which the percentage of codes on which the two authors agreed was calculated. Then, the coders discussed disagreements, revising the coding frame to reflect their decisions and growing understanding of the particulars of each code. The authors conducted three rounds of such coding. After these rounds of coding, the coders agreed that further, substantial improvements in agreement—and further changes to the codes—were unlikely. The agreement between the two coders during the last rounds of coding was high: According to Landis and Koch (1977), for the eight codes, Cohen's Kappa values ranged from "perfect" or "almost perfect" (Curated, Primary, Simulation-based, Sensor-based and Other Data) to substantial (Student Collected and Raw data) and, in one case, moderate (Textbook/Curriculum Data). The percentage agreement statistics were all high, ranging from 0.80–1.00. These percentage agreement and Cohen's Kappa statistics are reported in Appendix B.

# RESULTS

We present the results by research question, starting with the sources used by teachers (RQ1) and proceeding to the technologies that support teachers use to support their students' work with data (RQ2) and how sources used, in general, differ by the (pen-and-paper or digital) modality (RQ3).

## RQ #1: Which sources of data do science teachers select for their students to use?

### Specific data sources

To understand the data sources used, we developed the coding frame represented in Table 1 through the process described in the qualitative analysis section. The most common type of data was student-collected (80.3%). As an example, an educator responded to the question about what data their students use as follows:

> Students gather data whenever we can through investigations … Sometimes the data might be change of a variable through time: sometimes it is how many seeds germinate. We generally have 10–15 data points. I love having them compare their data to each other.

We can consider this student-collected data to be *first-hand* data. The other data sources reported were primarily *second-hand* in nature—data collected by others and provided to or accessed by students.

Second-hand data were varied in nature. Curated data—data that were formatted and presented *by others* for teachers' and students' uses—was reportedly used by 36.6% of

**TABLE 1** Coding frame and findings for the sources and types of data

| Code | Description | Overall frequency % | Elementary % | Middle % | High % |
|---|---|---|---|---|---|
| Student-collected data | Data students create through a laboratory or other investigation | 80.3 | 81.3% | 74.6% | 82.9 |
| Curated data | Data formatted and presented specifically for use by educators or consumption by the general public (eg, Data Nuggets, Mystery Science) | 36.6 | 15.6% | 34.2% | 39.7 |
| Raw data | Full datasets are accessible online but without any analysis, interpretation, or questions to guide analyses or interpretation provided (Google Earth, long-term datasets, NOAA) | 21.2 | 31.3% | 19.3% | 16.2 |
| Textbook or curriculum-based data | Data from the textbook or curricular materials | 17.6 | 12.5% | 14.9% | 20.9 |
| Primary source data | Data published within the primary scientific literature | 15.8 | 3.1% | 5.2% | 11.5 |
| Simulation-based data | Data from online or computer-based simulations (eg, Gizmos, PhET simulations, NetLogo simulations) | 6.7 | 0.0% | 7.0% | 6.0 |
| Sensor-based data | Data from probeware | 3.3 | 6.2% | 1.8% | 2.6 |
| Other data | Data that do not fit into the other categories | 21.2 | 21.9% | 22.8% | 19.7 |

*Note*: The overall percentages are based on the 330 respondents. The elementary, middle, and high school teacher percentages are based on the number of teachers at those grade levels: 32, 114, and 232, respectively.

teachers with their students; raw data by 21.5%; textbook or curriculum-based data by 17.6%; and primary source data, such as data from the primary scientific literature, by 15.8%.

A few other data sources mentioned could be considered to overlap with student-collected data; these were also first-hand in nature: data from simulations (6.7%) such as PhET simulations) and probeware (3.3% eg, temperature or movement sensors that stream data to a mobile device or computer). Often, respondents mentioned how their students would graph the data from these sources.

These responses suggest that student-collected data are far and away the most predominant in science classrooms. Also, data intended to be used for educational purposes (curated and textbook/curricular) sources as well as sources that may reflect the changing nature of data (namely, raw and primary source) were also somewhat commonplace. We next consider the size of the data teachers supported their students to analyse.

### Size of data set

Based on an initial open coding procedure, we used three codes for the size of the data: small (fewer than 20 rows, 2 columns or fewer; or, in cases of ambiguity, 40 cells or fewer), medium (20–100 rows or fewer, three-four columns; or, 40–300 cells), and large (100–1000 or more, five or more columns; or, 301 or more cells). We found that of the 96 respondents who mentioned the size of the data their students analysed (in the context of answering the question about what data sources they used), 79.1% of teachers reported using small data sets, 40.6% medium and 18.8% large. These findings suggest that teachers are mostly supporting their students to analyse small data sets—data sets with 20 rows or fewer. Less than one-half used data that has between 20–100 rows; and, less than one-fifth analysed data with more than 100 rows.

## RQ #2: Which digital tools for analysing data do teachers use to facilitate their students' work with data?

First, we report on the extent to which teachers engaged their students in first-hand (student-collected) and second-hand (not student-collected) data using pen-and-paper and digital tools. We then describe the digital technologies teachers used and what barriers would prevent them from using new technology for analysing and interpreting data.

### Tools used

As reported in Table 2, we found that teachers used a variety of tools, but that the tools that are among the most commonplace are also that may be most familiar to teachers across grade levels: Google Sheets (used by 82.7% of teachers with their students) and calculators (used by 72.1% of teachers). These were generally comparable across grade levels. After these tools, Microsoft Excel (42.7%) was reported as widely used—and used more by high school than middle and elementary teachers. More specialized (to science education) tools, including DataClassroom (12.4%) and the Common Online Data Analysis Platform (4.2%) were less widely-used than many other tools.

**TABLE 2** Digital technologies used by students to analyse and interpret data

| Tools/Resources | Overall % | Elementary % | Middle % | High % |
|---|---|---|---|---|
| Google sheets | 82.7 | 81.2 | 78.1 | 85.8 |
| Calculator (not for graphing) | 72.1 | 75.0 | 71.9 | 75.0 |
| Microsoft excel | 42.7 | 34.4 | 30.7 | 50.0 |
| Graphing calculator | 28.8 | 18.8 | 17.5 | 38.4 |
| Desmos | 15.8 | 25.0 | 21.9 | 15.5 |
| DataClassroom | 12.4 (41) | 6.2 | 10.5 | 15.1 |
| Infogram (or similar online tool for creating an infographic) | 11.2 | 18.8 | 14.9 | 10.3 |
| Common Online Data Analysis Platform (CODAP) | 4.2 | 3.1 | 4.4 | 4.7 |
| TUVA | 3.3 | 3.1 | 6.1 | 2.6 |
| R | 1.8 | 3.1 | 0.9 | 1.7 |

*Note*: The overall percentages are based on the 330 respondents. The elementary, middle and high school teacher percentages are based on the number of teachers at those grade levels: 32, 114, and 232, respectively.

## Barriers to the use of digital tools to analyze data

We found that teachers reported that cost, far and away, was the strongest barrier to their use of a digital tool—85.5% of respondents indicated that this was a barrier and this high percentage was generally comparable across elementary, middle, and high school teachers (Table 3). After cost, the time it would take to re-develop pen-and-paper lessons (reported as a barrier by 52.1% of respondents); more middle and high school teachers (45.6% and 55.2%) than elementary teachers (28.1%) reported the time it would take to redevelop lessons. The difficulty of learning a new tool (46.4% of respondents) was another strong barrier. Notably, neither teachers' own (11.2%) nor teachers' perception of students' discomfort with computers were strong barriers (7.6%).

## RQ #3: How does the modality (pen-and-paper or digital) of students' work with data differ based on the type of data source selected by teachers?

Finally, we consider data sources by modality (Table 4). Overall, around four-fifths of teachers reported engaging their students in the analysis of first-hand data using pen-and-paper (80.0%) and digital tools (78.8%). The analysis of second-hand data was a bit less common overall: 68.5% of teachers reported analysing this type of data using pen-and-paper; 77.9% reported doing so using digital tools. Considered in terms of modality, engaging students in pen-and-paper-based analyses was slightly less commonly reported than doing so using digital tools. The most notable differences, however, can be found by considering modality and type of data source together. This reveals that analysing second-hand data using pen-and-paper was less common than the other data source by modality combinations--though, still, around two-thirds of teachers reported engaging their students in this type of data analysis.

Elementary teachers reported analysing first-hand data (90.6%) more than middle (81.6%) and high school (79.3%) teachers—and much more than they reported doing the same with digital tools. Elementary teachers used digital tools in both pen-and-paper and digital modalities less than middle and high school teachers; additionally,

**TABLE 3** Barriers identified by teachers to their adoption of digital tools to analyze data in their classrooms

| Barrier | Overall % | Elementary % | Middle % | High % |
|---|---|---|---|---|
| Cost | 85.5 | 90.6 | 86.8 | 85.8 |
| Time to develop new lessons that I previously did using pencil-and-paper | 52.1 | 28.1 | 45.6 | 55.2 |
| Difficulty of learning a new tool | 46.4 | 43.8 | 45.6 | 47.4 |
| Student information security | 31.2 | 43.8 | 40.4 | 28.0 |
| Availability of computers (includes tablets; laptops) | 28.5 | 21.9 | 22.8 | 30.6 |
| Student discomfort with computers | 11.2 | 0.0 | 9.6 | 12.9 |
| My own discomfort with computers | 7.6 | 3.1 | 8.8 | 6.5 |

*Note*: The overall percentages are based on the 330 respondents. The elementary, middle, and high school teacher percentages are based on the number of teachers at those grade levels: 32, 114, and 232, respectively.

**TABLE 4** The frequency with which teachers engage their students in first-hand (student-collected) and second-hand data (not student-collected) using pen-and-paper and digital tools

| Modality | Type of data source | Overall % | Elementary % | Middle % | High % |
|---|---|---|---|---|---|
| Pen-and-paper | First-hand | 80.0 | 90.6 | 81.6 | 79.3 |
| Digital tools | First-hand | 78.8 | 71.9 | 70.2 | 83.2 |
| Pen-and-paper | Second-hand | 68.5 | 62.5 | 63.2 | 70.3 |
| Digital tools | Second-hand | 77.9 | 65.6 | 75.4 | 81.5 |

*Note*: The overall percentages are based on the 330 respondents. The elementary, middle and high school teacher percentages are based on the number of teachers at those grade levels: 32, 114 and 232, respectively.

high school teachers reported using digital tools (83.2%) more than middle school teachers (70.2%).

# DISCUSSION

Our survey was an attempt to capture a moment in the field. We surveyed 330 educators from across the U.S. about the sources of data and technologies that support their and students' work with data that are common in their science classrooms. In this section, we discuss what we see as our key findings, followed by the implications of these findings for research, practice, and science education policy and some limitations of the study, and directions for future research.

## Student-collected (first-hand) data is predominant, but second-hand data is also used

Teachers primarily involve their students in the analysis of a particular kind of data: student-collected, or first-hand. This is notable as there are likely benefits as well as some drawbacks for student learning; it may be especially valuable for learners to consider the data-related practice of measuring and recording data (Hug & McNeill, 2008). Second-hand data, such as the raw data from sources such as climate data from the National Oceanic and Atmospheric Association (NOAA), can be more complex (used by 21.2% of teachers). Complexity presents both challenges and opportunities for students by inviting them to manipulate data, explore patterns and make inferences, and generate data-based conclusions to a greater extent than when they analyse first-hand data—practices that may be difficult for students to do without support (Hug & McNeill, 2008). We note that elementary teachers reported using this data source even more (31.3%) than middle and high school teachers (19.2% and 16.2%, respectively); this may be as first-hand data may be more easier to work with (and to find meaning in) for younger learners.

Also regarding data sources, we think the presence (if not the high proportion) of teachers reporting using raw and primary source data (used by 15.8% of teachers; and used more by high school than middle and especially elementary school teachers) is encouraging and in-line with calls for students to analyse these large, second-hand data sources to a greater extent (Lee & Wilkerson, 2018; Magnusson et al., 2004; National Academies of Sciences, Engineering, and Medicine, 2019; Rosenberg, Edwards, et al., 2020). We also found curated data to be a common second-hand data source (36.6%); the analysis of such data that is often messy and complex but is formatted and presented for teachers and students may serve as a lever for teachers and students looking to transition from first-hand to the often more difficult to analyse second-hand data sources (Kjelvik & Schultheis, 2019; Schultheis & Kjelvik, 2020). Such data sources may also present a context to analyse large data sets; our findings show that the analysis of data sets with as few as 20 rows (79.1% of teachers) is more than four times as common as the analysis of large data sets (18.8%)—those with greater than 100 rows. This suggests there is still much room for teachers to integrate the analysis of larger, more complex data sources into their teaching.

## Familiar and available, spreadsheets and calculators are widely-used digital tools

Regarding digital tools, we found that teachers predominantly use technologies that may be familiar to them, such as Google Sheets (used by more than four-fifths of teachers: 82.7%), and not-for-graphic calculators (used by almost three-quarters; 72.1%). Following these was Microsoft Excel (42.7%), graphing calculators (28.8%), and the web-based graphic calculator Desmos (15.8%). There is little research to compare these findings to; the NSSME+ reports the availability of microscopes, balances, and probes (the last of which is reportedly available in 81% of high school teachers' classrooms—which we think is a very high percentage given the relative rarity with which teachers mentioned using data from probes; Banilower et al., 2018). The NSSME+ also reports the availability of general instructional technologies, such as the projection devices available in 99% of classes, and the availability of foundational technologies/resources, including electrical outlets. Saying this is to highlight that we do not have evidence on how extensive the use of spreadsheets (Google Sheets and Microsoft Excel) and calculators (non-graphing and graphing, including Desmos) are.

That teachers use the data analytics tools that are accessible and familiar to them is a novel finding in our (science education) context, but it is not in the wider educational technology literature. Scholars have found that the introduction of new technologies into schools and classrooms does not necessarily (or even likely) lead to changes in instructional practice (Cuban, 2009; Cuban et al., 2001). The reasons for this are manifold and include organizational vestiges (ie, how schools are structured; Cohen, 1987; Cuban et al., 2001) and the (not always recognized) imperative for teachers to adapt technologies not originally designed for teaching and learning (Mishra et al., 2009). This prior research suggests that the reasons teachers use familiar technologies may be due to historical, contextual, and technological reasons; it would probably be incorrect to attribute teachers' use of familiar technologies for analysing data to their lack of interest, motivation or knowledge. We think that this prior research helps to explain why we found that a technology that is ostensibly well-suited to the collection of data in science classrooms—like probeware—remains relatively unused.

Our findings suggest that spreadsheets and calculators are the digital tools that are the most predominant concerning what students use to analyse data. They are domain-general tools, likely widely-available and familiar to teachers and students alike. Similar to Lee and Harrison (2021) found for statistics educators, tools specific to the discipline (in our case, science education—or even tools specifically designed for teaching and learning to analyse data) were relatively uncommon. DataClassroom (12.4%) was used by around one-eighth of educators; CODAP (4.2%), TUVA (3.3%) and R (1.8%) were rarely used. These suggest that there is more room for such tools to be used to support more involved or sophisticated forms of data analysis; they may be necessary especially for the cleaning and "wrangling" steps of data analysis (National Academies of Sciences, Engineering, and Medicine, 2019); they may also be useful for modelling data (Rosenberg, Edwards, et al., 2020).

The barriers that teachers see regarding the use of a new digital tool are not technology-driven, but instead are closely related to resources: cost and time. Such findings can be considered in light of educational technology research, which suggests that both teachers' understandings and knowledge as well as features of *their context* (including the cost to purchase digital technologies or access to them and how their planning and teaching schedules and structure afford the time to explore the use of new tools) matter in terms of the adoption of new tools (Mishra & Koehler, 2006; Porras-Hernández & Salinas-Amescua, 2013). Are cost and time truly barriers, or are these those that teachers find to be most salient? We think it is reasonable to infer that teachers' perception of cost likely relates to their ability as individuals to purchase educational technology tools, though teachers may have access to school or district funds that could-potentially—at least partially—ameliorate this concern. Even still, in many cases, teachers do not have direct control over how these funds are spent, and so their perception of cost may be salient whether they are considering their own or their institutions' funds. Furthermore, we note that free tools may present another challenge: they may take time to learn to use. R, for instance, is likely to be unfamiliar to anyone without a programming background and Desmos may be more familiar to mathematics teachers with experience using graphing calculators. Thus, we consider teachers reports of time and cost as key barriers as valid. Notably, the barriers reported are minimally related to teachers' knowledge and beliefs; teachers' own (11.2%) and students' discomfort (7.6%) with computers were rarely reported. Notably, in research that examines teachers' knowledge related to technology, the context-related barriers we identified in the present study have not been the focus of very much past research (Rosenberg & Koehler, 2015).

## Pen-and-paper and digital modalities using first- and second-hand data are common

Finally, regarding how modality (pen-and-paper and digital) and data source (first-hand and second-hand) intersect, we found that the state of students' work with data in science classrooms is complex; students are analysing highly-different types of data using both pen-and-paper and digital tools; two-thirds or more of teachers reported engaging students in data-related tasks using different modalities and types of data. This suggests that the target of curricular or technological design and development efforts must account for how teachers and students do not analyse data in one way, as may have been the case in the past when digital tools were less commonplace or accessible.

Though this finding of teachers engaging students within pen-and-paper and digital modalities with first- and second-hand data was generally the case across grade levels, the use of digital tools is more common among teachers at the upper-grade levels relative to lower-grade levels. That elementary teachers have fewer science-specific tools at their disposal that they use aligns with the results of the NSSME+ showing that tools such as probes were less-available in elementary classrooms (Banilower et al., 2018). Also, while all four of the percentages of modality by data source combinations were moderately high, these percentages were also not near 100%, indicating that some teachers. The responses of many teachers indicate that they *never* engage their students in recording data using pen-and-paper or digital tools. We think this raises provocative questions about whether all teachers can or should involve their students in the analysis of different sources of data using different modalities. How essential is it that students use digital tools to analyse second-hand data? We consider this question in the next section.

## Implications for research, practice and policy

These findings present the first descriptive portrait of the data sources and tools used by science teachers in the United States; accordingly, the implications of this study are delimited to this context. In this way, these findings have some implications for the kinds of research called for in the recent report by the National Academies of Sciences, Engineering, and Medicine (2019). Namely, this report called for research on “the appropriate roles within particular science investigation and design environments for student use of technology to collect, analyze, interpret, and communicate data” (p. 280). This study provides baseline information on the technologies used for data analysis that can inform which technologies researchers adapt or develop. For example, the predominance of spreadsheets and calculators means that researchers could consider what is missing from these tools and what can be gained by adding functionality (namely, for cleaning or “wrangling” data) in add-ons to these tools or—more likely—the development of new tools. Aligning with past research on how there is not one ideal tool for learning to analyse data (McNamara, 2018), we think that no single tool will be able to address all these challenges. We should also not assume teacher familiarity or comfort with things that may be assumed (or found) to be common, like probeware or simulations or using specialized software.

These findings also have some implications for research on the sources of data used. Some past research has documented the affordances of first- and second-hand data (Hug & McNeill, 2008; Lee, Drake, et al., 2021). Other research has called for the analysis of more second-hand data to overcome some of the inherent limitations of first-hand data (Magnusson et al., 2004; Palincsar & Magnusson, 2001); these calls have been amplified in light of new kinds of data—often large, complex data types that can afford particular types of student work with data. Our finding that student-collected, small data sets were common can inform research on starting points for teachers and students to make inferences about multivariate patterns and relationships—something scholars have called for future research to do (Lee & Wilkerson, 2018). For instance, researchers may consider developing a lesson sequence that involves students measuring one dozen plants prior to using observations from a large data set on the heights and soil composition of 1000s of plants. Beginning with the analysis of a smaller data set may help more students to start to make sense of this kind of data; furthermore, knowing how the data was collected may support students to understand how the larger data set was collected. Even still, we think our findings suggest researchers should consider working very carefully through the challenges that the analysis of large, multivariate data entails. Research from the statistics education literature may help inform how to do this. For example, research in statistics education has emphasized the pedagogical importance of modelling variability in real-world data a variety of ways, such as by emphasizing the range of mathematical and visual ways variation and covariation can be represented (eg, Kazak et al., 2021; Lehrer & English, 2018). Another area of statistics education research that we think may be especially fruitful for science education researchers and educators looking to support students' work with data is on informal statistical inference (eg, Makar & Rubin, 2018)—an approach to statistical inference related to advanced Bayesian methods that may (surprisingly) be helpful to learners (Rosenberg et al., 2021).

Set within this challenging time in classrooms across the country, teachers are still making impressive strides to infuse data and research into their classrooms. These moves are a step toward answering the call put out by the NGSS and other education reform efforts to help students deal with future lives and careers that will be rich in data experiences and where literacy is needed to make informed decisions about all aspects of life. To do this work effectively, scientists, curriculum developers and technology developers need to provide educators with the tools and professional development they need.

We think many possible efforts could address some of the gaps between what is common in K-12 classrooms and what is possible in terms of supporting students to engage data in ways called for in recent reform documents (eg, NGSS Lead States, 2013). Additionally, simply making new data sources or tools available to teachers is unlikely—alone—to be effective in supporting changes. Professional learning opportunities and other systems-level changes are necessary (National Research Council, 2015; Penuel et al., 2015; Stiles et al., 2017). Given the prominence of curated data in the classrooms of the teachers we surveyed, we think professional learning opportunities could focus on these data sources as an entry way to the analysis of more complex data sources, including raw or primary sources—which may be even more rich than curated sources. Furthermore, professional learning opportunities could introduce digital tools—

such as CODAP—that may suit the distinctive needs of science teachers, especially those that do not require ample time for teachers (and students) to learn to begin to use—and that have reasonable and affordable cost structures.

Though not the focus of the present study, we note that only 3.9% of the teachers in this study reported that they were not interested in professional development related to students' data collection and analysis; 75.4% reported they were interested and 20.0% reported they were unsure. We think PD on the analysis of second-hand data and teachers′ curation of raw, primary source or other large data sets may be especially useful.

## Limitations

Two limitations of this study merit mention here. First, we took steps to compare the respondents to our survey to those in the United States (by using statistics from the nationally-representative NSSME+). Nevertheless, there were notable differences between the teachers in our sample and those in the NSSME+—particularly striking for how we did not collect sufficient responses from teachers of colour—African American and Hispanic science teachers, particularly. More broadly, this sample was not intended to be nationally representative and the (more) deliberate steps that researchers would take to do so—including developing a sampling procedure and using sampling weights to assist with the process of making inferences from even an unbalanced sample to a population—were not taken in this study. Moreover, drawing from the mailing lists of Data Nuggets and DataClassroom to recruit participants to complete this study may lead to sources of bias particular to these respondents. Namely—as implicated in the method section—users of Data Nuggets and DataClassroom may analyse data with their students more frequently than the average teacher in the United States. Furthermore, that the teachers in our sample reported engaging their students in data-related practices at rates generally comparable to those teachers in the national sample gives us confidence that our teachers are not wholly different from those nationwide with respect to how their students work with data. Because our sample is generally comparable to a national sample (though the sampling strategy we used is different from the approach used by the developers of the NSSME+ to establish a nationally representative sample), this study represents a first step toward understanding the data sources and technologies science teachers use to support their students to work with data, but subsequent work may explore these topics more systematically or deeply.

A second key limitation of this study concerns how we analysed the data. We took a descriptive approach instead of hypothesis tests of the statistical significance of the statistics we presented. In some cases, such hypothesis tests may be reasonable—especially for comparisons between teachers of different grade levels. However, our focus was on describing the data sources and data analysis-related technologies teachers use. At the same time, we took care to only highlight differences when they were highly apparent. We did not, for example, claim that there were differences when percentages differed by a few points. Finally, readers may anticipate measures of spread, such as the standard deviation for continuous variables. However, our use of dichotomous and categorical data meant we were unable to calculate descriptive measures of the spread as would be possible for continuous data. For these reasons, we believe the focus on strictly describing our sample is merited but recognize that readers may reasonably be curious as to the statistical significance of some of the differences we explored and presented. Our intent (and recommendation) is for future research to study some of the key differences of interest in greater qualitative or quantitative depth.

## Recommendations for future research

How these findings might have changed given the COVID-19 pandemic is a question that we would like to see pursued in future research. Our survey was administered in February 2020—before the interruptions of the pandemic were in full effect in the United States. We think it will be important to re-administer this survey to capture how the pandemic has influenced data literacy instruction and the use of technology in the classroom. Though our respondents did not cite comfort with technology as a large barrier, it may be that confidence with and access to digital technologies have increased over the past year—and so future research may profitably document how the nature of teachers' (and students') data has changed since we carried out this study.

We began this manuscript with a nod to how the kind of data available across scientific disciplines to scientists and the public alike has changed and is continuously changing. Science teachers have long adapted and changed their teaching in response to external and internal (to the educational system) preferences, values, and even demands (Rudolph, 2019). We think that the ways in which teachers and students approach data in creative, ambitious ways can serve as a lever for positive changes much like other science practices such as arguing from evidence has helped to support changes in science teaching and learning (Ke & Schwarz, 2021). Our intent with this work was to document the state of teachers' efforts to engage their students in work with data so that others—researchers and teachers alike—may best serve and bolster teachers' work going ahead.

### ACKNOWLEDGEMENTS
Authors Schultheis and Kjelvik have funding for Data Nuggets through the W. K. Kellogg Biological Station (KBS) Long-Term Ecological Research program (NSF DEB 1832042) and NSF IUSE 2012014. This is KBS Contribution #2320.

### CONFLICT OF INTEREST
JR and OS disclose no conflicts of interest. ES and MK are co-creators of the Data Nuggets resources. AR is Co-founder and CEO of the data analysis tool DataClassroom.

## DATA AVAILABILITY STATEMENT

The analytic code necessary to reproduce all of the analyses presented in this manuscript is available at: https://rpubs.com/jmichaelrosenberg/data-in-science. The data we collected and analysed for this study is is made openly available at: https://osf.io/g5zbu/.

## ETHICS APPROVAL

This study is approved by the University of Tennessee, Knoxville Institutional Review Board (UTK IRB-20-05925-XM).

## ORCID

*Joshua M. Rosenberg* https://orcid.org/0000-0003-2170-0447
*Elizabeth H. Schultheis* https://orcid.org/0000-0003-3148-9833
*Melissa K. Kjelvik* https://orcid.org/0000-0003-2780-0131
*Aaron Reedy* https://orcid.org/0000-0001-8835-0811

## ENDNOTES

[1] The complete revised survey, which we constructed using the Qualtrics platform, is available in an anonymous form at https://tiny.utk.edu/data-in-science

[2] The complete which we constructed using the Qualtrics platform, is available in an anonymous form at https://tiny.utk.edu/data-in-science.

## REFERENCES

ACT, Inc. (2014). *ACT college and career readiness standards: Science*. www.act.org/standard/planact/science

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). *Report of the 2018 NSSME+*. Horizon Research, Inc.

Bhargava, R., & D'Ignazio, C. (2015). *Designing tools and activities for data literacy learners*. Wed Science: Data Literacy Workshop. https://www.media.mit.edu/publications/designing-tools-and-activities-for-data-literacy-learners/

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679.

California Academic of Sciences and National Geographic Association. (2021). *iNaturalist*. https://www.inaturalist.org/

CERN. (2021). *Computing storage*. https://home.cern/science/computing/storage#:~:text=The%20LHC%20experiments%20produce%20about,an%20essential%20function%20at%20CERN

Cohen, D. K. (1987). Educational technology, policy, and practice. *Educational Evaluation and Policy Analysis*, *9*(2), 153–170.

College Board. (2019). *AP Biology: Course and exam description*. https://apstudents.collegeboard.org/ap/2019-05/ap-biology-course-and-exam-description.pdf

Computer Science Teachers Association. (2017). *K-12 Computer Science Standards*. https://drive.google.com/file/d/1-dPTAI1yk2HYPKUWZ6DqaM6aVUDa9iby/view

Cuban, L. (2009). *Oversold and underused*. Harvard University Press.

Cuban, L., Kirkpatrick, H., & Peck, C. (2001). High access and low use of technologies in high school classrooms: Explaining an apparent paradox. *American Educational Research Journal*, *38*(4), 813–834.

Fink, A. (2015). *How to conduct surveys: A step-by-step guide*. Sage Publications.

Forbes, S. (2014). The coming of age of statistics education in New Zealand, and its influence internationally. *Journal of Statistics Education*, *22*(2).

Gardner, S. M., Suazo-Flores, E., Maruc, S., Abraham, J. K., Karippadath, A., & Meir, E. (2021). Biology undergraduate students' graphing practice in digital versus pen and paper graphing environments. *Journal of Science Education and Technology*, *30*, 431–446.

Hatch, J. A. (2002). *Doing qualitative research in education settings*. SUNY Press.

Hug, B., & McNeill, K. L. (2008). Use of first-hand and second-hand data in science: Does data type influence classroom conversations? *International Journal of Science Education*, *30*(13), 1725–1751.

Kahn, J., & Jiang, S. (2021). Learning with large, complex data and visualizations: youth data wrangling in modeling family migration. *Learning, Media and Technology*, *46*(2), 128–143.

Kazak, S., Fujita, T., & Turmo, M. P. (2021). Students' informal statistical inferences through data modeling with a large multivariate dataset. *Mathematical Thinking and Learning*, 1–21. https://doi.org/10.1080/10986065.2021.1922857

Ke, L., & Schwarz, C. V. (2021). Supporting students' meaningful engagement in scientific modeling through epistemological messages: A case study of contrasting teaching approaches. *Journal of Research in Science Teaching*, *58*(3), 335–365.

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), 1–8.

Krajcik, J. S., & Mun, K. (2014). Promises and challenges of using learning technologies to promote student learning of science. In *Handbook of research on science education* (Vol. *II*, pp. 351–374). Routledge.

Krishnamurthi, S., & Fisler, K. (2020). Data-centricity: A challenge and opportunity for computing education. *Communications of the ACM*, *63*(8), 24–26.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

LandSatLook. (2021). *LandSat Look*. https://landsatlook.usgs.gov/explore

Lee, H. S., & Harrison, T. (2021). Trends in teaching advanced placement statistics: Results from a national survey. *Journal of Statistics and Data Science Education*, *29*(3), 317–227.

Lee, V. R., Drake, J., Cain, R., & Thayne, J. (2021). Remembering what produced the data: Individual and social reconstruction in the context of a quantified self elementary data and statistics unit. *Cognition and Instruction*, *39*(4), 367–408.

Lee, V. R., & Wilkerson, M. (2018). *Data use by middle and secondary students in the digital age: A status report and future prospects* (pp. 1–43). Commissioned Paper for the National Academies of Sciences, Engineering, and Medicine, Board on Science Education, Committee on Science Investigations and Engineering Design for Grades 6–12. https://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_189500.pdf

Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A call for a humanistic stance toward K–12 data science education. *Educational Researcher.*, *50*, 664–672.

Lehrer, R., & English, L. (2018). Introducing children to modeling variability. In *International handbook of research in statistics education* (pp. 229–260). Springer.

Magnusson, S. J., Palincsar, A. S., Hapgood, S., & Lomangino, A. (2004). How should learning be structured in inquiry-based science instruction? Investigating the interplay of 1st- and 2nd-hand investigations. In Y. Kafai, W. Sandoval, N. Enyedy, A. Nixon, & F. Herrera (Eds.), *Proceedings of the Sixth International Conference of the Learning Sciences* (pp. 310–317). Lawrence Erlbaum Associates, Inc.

Makar, K., & Rubin, A. (2018). Learning about statistical inference. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 261–294). Springer International Publishing. https://doi.org/10.1007/978-3-319-66195-7_8

Mayes, R., & Koballa, T. R., Jr. (2012). Exploring the science framework. *The Science Teacher*, *79*(9), 27.

Mayes, R., Long, T., Huffling, L., Reedy, A., & Williamson, B. (2020). Undergraduate quantitative biology impact on biology preservice teachers. *Bulletin of Mathematical Biology*, *82*, 1–28.

McNamara, A. (2018). Key attributes of a modern statistical computing tool. *The American Statistician*, *73*(4), 375–384.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054.

Mishra, P., Koehler, M. J., & Kereluik, K. (2009). Looking back to the future of educational technology. *TechTrends*, *53*(5), 49.

NASA. (2021). *Landsat science*. https://landsat.gsfc.nasa.gov/landsat-9/landsat-9-overview

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 Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. National Governors Association Center for Best Practices and the Council of Chief State School Officers.

National Human Genome Research Institute. (2021). *COVID-19 mRNA vaccine production*. https://www.genome.gov/about-genomics/fact-sheets/COVID-19-mRNA-Vaccine-Production

National Research Council (2006). America's lab report: Investigations in high school science. committee on high school science laboratories: Role and vision. In S. R. Singer, M. L. Hilton, & H. A. Schweingruber (Eds.), *Board on science education, center for education. Division of behavioral and social sciences and education*. The National Academies Press.

National Research Council. (2012). *A framework for k-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

National Research Council. (2015). *Guide to implementing the next generation science standards*. National Academies Press. https://doi.org/10.17226/18802

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. The National Academies Press.

Palincsar, A. S., & Magnusson, S. J. (2001). The interplay of first-hand and second-hand investigations to model and support the development of scientific knowledge and reasoning. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 151–193). Lawrence Erlbaum Associates Publishers.

Penuel, W. R., Harris, C. J., & DeBarger, A. H. (2015). Implementing the next generation science standards. *Phi Delta Kappan*, *96*(6), 45–49. https://doi.org/10.1177/0031721715575299

Porras-Hernández, L. H., & Salinas-Amescua, B. (2013). Strengthening TPACK: A broader notion of context and the use of teacher's narratives to reveal knowledge construction. *Journal of Educational Computing Research*, *48*(2), 223–244.

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://learningcenter.nsta.org/resource/?id=10.2505/4/tst20_087_05_30

Rosenberg, J. M., & Koehler, M. J. (2015). Context and technological pedagogical content knowledge (TPACK): A systematic review. *Journal of Research on Technology in Education*, *47*(3), 186–210.

Rosenberg, J. M., Kubsch, M., Wagenmakers, E., & Dogucu, M. (2021). *Making sense of uncertainty in the science classroom: A Bayesian approach*. https://doi.org/10.31219/osf.io/aznyq [OSF pre-print]

Rosenberg, J. M., Lawson, M. A., Anderson, D. J., Jones, R. S., & Rutherford, T. (2020). Making data science count in and for education. In E. Romero-Hall (Ed.), *Research methods in learning design & technology* (pp. 94–110). Routledge.

Rudolph, J. L. (2019). *How we teach science: What's changed, and why it matters*. Harvard University Press.

Schultheis, E. H., & Kjelvik, M. K. (2015). Data nuggets: Bringing real data into the classroom to unearth students' quantitative & inquiry skills. *The American Biology Teacher*, *77*(1), 19–29.

Schultheis, E. H., & Kjelvik, M. K. (2020). Using messy, authentic data to promote data literacy & reveal the nature of science. *The American Biology Teacher*, *82*(7), 439–446.

Stiles, K. E., Mundry, S. E., & DiRanna, K. (2017). *Framework for leading next generation science standards implementation*. WestEd.

Sullivan, B. L., Aycrigg, J. L., Barry, J. H., Bonney, R. E., Bruns, N., Cooper, C. B., Damoulas, T., Dhondt, A. A., Dietterich, T., Farnsworth, A., Fink, D., Fitzpatrick, J. W., Fredericks, T., Gerbracht, J., Gomes, C., Hochachka, W. M., Iliff, M. J., Lagoze, C., La Sorte, F. A., … Kelling, S. (2014). The eBird enterprise: An integrated approach to development and application of citizen science. *Biological Conservation*, *169*, 31–40.

Tweddle, J. C., Robinson, L. D., Pocock, M. J. O., & Roy, H. E. (2012). *Guide to citizen science: developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK*. NERC/Centre for Ecology & Hydrology.

Wolff, A., Wermelinger, M., & Petre, M. (2019). Exploring design principles for data literacy activities to support children's inquiries from complex data. *International Journal of Human-Computer Studies*, *129*, 41–54.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Rosenberg, J. M., Schultheis, E. H., Kjelvik, 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*, 1179–1201. https://doi.org/10.1111/bjet.13245

# APPENDIX A

## SURVEY QUESTIONS

- What software, tools, or resources related to analysing and interpreting data have students used while taking your class(es)? (11 fixed-response options for specific tools plus an "other" option)
- You indicated that your students use the following software, tools, or resources. Approximately how often do students use each of the following? (four fixed response options: daily, weekly, monthly, or rarely)
- What sources of data do you use with your students (eg, student-collected data, datasets found online or through other sources, etc.)? (open-ended)
- In which of the following activities do you involve your students? (dichotomous items; options follow)
  - Analysing first-hand data using pen-and-paper
  - Analysing second-hand data using pen-and-paper
  - Analysing first-hand data using digital tools
  - Analysing second-hand data using digital tools
- Which of the following might prevent you from using a digital tool? (seven fixed-response options plus an "other" option; options follow)
  - Cost
  - Availability of computers (tablets, laptops, etc.)
  - Student information security
  - Time to develop new lessons that I previously did using pencil-and-paper
  - Difficulty of learning a new tool
  - Student discomfort with computers
  - My own discomfort with computers
  - Other (please describe)

# APPENDIX B

## AGREEMENT AND COHEN'S KAPPA STATISTICS FOR THE CODES FOR THE SOURCES OF DATA TEACHERS USED

| | % agreement | CoK |
|---|---|---|
| Student-collected data | 0.95 | 0.64 |
| Raw data | 0.9 | 0.6 |
| Textbook or curriculum-based data | 0.8 | 0.41 |
| Curated data | 0.95 | 0.89 |
| Primary source data | 0.95 | 0.82 |
| Simulation-based data | 1 | 1 |
| Sensor-based data | 1 | N/A |
| Other data | 0.95 | 0.89 |

N/A, not applicable.

## APPENDIX C

### COMPARISON OF TEACHERS IN OUR SAMPLE TO A NATIONALLY-REPRESENTATIVE (NSSME+) SAMPLE IN TERMS OF WHETHER TEACHERS ENGAGE THEIR STUDENTS WEEKLY OR MORE FREQUENTLY ON SELECT DATA-RELATED PRACTICES

| | Elementary (Our sample, %) | Elementary (NSSME+, %) | Middle (Our sample, %) | Middle (NSSME+, %) | High (Our sample, %) | High (NSSME+, %) |
|---|---|---|---|---|---|---|
| Organize and/or represent data using tables, charts, or graphs in order to facilitate analysis | 53 | 34 | 45 | 49 | 51 | 58 |
| Analyse data using grade-appropriate methods in order to identify patterns, trends, or relationships | 50 | 27 | 45 | 43 | 54 | 47 |
| Determine which data would need to be collected in order to answer a scientific question | 28 | 29 | 21 | 39 | 20 | 39 |