The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/2398-5348.htm

Engaging elementary students in data science practices

Ibrahim Oluwajoba Adisa, Danielle Herro, Oluwadara Abimbade and Golnaz Arastoopour Irgens
Department of Education and Human Development/Learning Sciences, Clemson University, Clemson, South Carolina, USA

Received 12 June 2023
Revised 5 September 2023
12 October 2023
18 November 2023
Accepted 22 November 2023

Abstract

Purpose – This study is part of a participatory design research project and aims to develop and study pedagogical frameworks and tools for integrating computational thinking (CT) concepts and data science practices into elementary school classrooms.

Design/methodology/approach – This paper describes a pedagogical approach that uses a data science framework the research team developed to assist teachers in providing data science instruction to elementary-aged students. Using phenomenological case study methodology, the authors use classroom observations, student focus groups, video recordings and artifacts to detail ways learners engage in data science practices and understand how they perceive their engagement during activities and learning.

Findings – Findings suggest student engagement in data science is enhanced when data problems are contextualized and connected to students' lived experiences; data analysis and data-based decision-making is practiced in multiple ways; and students are given choices to communicate patterns, interpret graphs and tell data stories. The authors note challenges students experienced with data practices including conflict between inconsistencies in data patterns and lived experiences and focusing on data visualization appearances versus relationships between variables.

Originality/value – Data science instruction in elementary schools is an understudied, emerging and important area of data science education. Most elementary schools offer limited data science instruction; few elementary schools offer data science curriculum with embedded CT practices integrated across disciplines. This research assists elementary educators in fostering children's data science engagement and agency while developing their ability to reason, visualize and make decisions with data.

Keywords Data science practices, Data analysis, Elementary-school children, Student engagement, Qualitative research, Connected learning

Paper type Research paper

Introduction

Data and computing practices include generating, collecting, analyzing and reusing personal data. These practices are embedded in everyday applications on platforms such as Google, Facebook, Apple and Amazon. However, users rarely consider how and why these data are being used (Pangrazio and Selwyn, 2019). Society's reliance on these platforms, cloud-based storage, artificial intelligence, machine learning and other data-centric computing necessitates understanding how data are used to make personal and collective decisions (Weiland and Engledowl, 2022). Within these platforms and applications, data can

This research is supported by NSF Award 2031175: Inclusive Data Science Education for Rural Elementary Students: A Research Practice Partnership for Agile Learning.

2398-5348
DOI 10.1108/ILS-06-2023-0062

be used for good, such as Google's use of public data and artificial intelligence (AI)-based models to understand, predict, and stop the spread of infectious diseases (Sava, 2020). Data can also be used for harm such as propagating misinformation and manipulating public opinion (i.e. swaying political elections, inflating medical outcomes) through targeted social media that misrepresents data (Bradshaw and Howard, 2017). The modern digital world requires developing literate citizens capable of making informed data-based decisions in their daily lives and equipping them with essential data science skills for a rapidly evolving job market. These important skills must begin at an early age; however, current data science education for young learners, especially elementary and middle school children (aged 5–12), rarely prepares them for this societal need (Kjelvik and Schultheis, 2019).

Globally, elementary learners acquire very little experience with data science and usually arrive unprepared to deal with computational problems at higher levels of education (Martinez and LaLonde, 2020). Studies attribute this to students' infrequent engagement with data (Lee *et al.*, 2021) or limited supports for teachers' data science preparation hindering their ability to integrate data science practices into lessons (Bowen, 2021; LaMar and Boaler, 2021). Our research addresses this research gap by facilitating an understanding of how elementary students, specifically students in grades three-through-five, learn data science concepts and how they engage in data science practices.

As part of an NSF-funded research project spanning three years, we developed a research-practitioner partnership (Coburn and Penuel, 2016) focused on co-creating, with teachers, an integrated data science curriculum for elementary students. Similar to most elementary schools in the USA, data science was not integral to the formal curriculum (Weiland and Engledowl, 2022); instead, it was primarily addressed during mathematics instruction using limited data practices (e.g. graphing textbook data). The curriculum encompassed instructional units at each grade level referred to as "CT-STEM Pop-Ups" as they emphasized computational thinking (CT) and STEM skills. While CT is not new, it has primarily been associated with domain-specific disciplines like computer science and mathematics, yet it cuts across a wide-range of disciplines as critical CT components include reasoning practices (decomposing problems and recognizing patterns), problem solving (extracting and applying relevant information) and conceptual understanding (Wing, 2006). In data science education, CT is often used to frame and conceptualize how data is represented, collected, analyzed and applied to problem solving. CT is often integrated in other frameworks (see CT-STEM Taxonomy of Practices below) as a way to integrate disciplinary and problem-solving practices. In education, pop-ups are considered customizable, innovative curricula or units that are typically not covered in the everyday curriculum but can be taught in an integrated manner (pop-ups; Tranquillo and Matthew, 2015). Pop-ups typically offer hands-on discovery learning and are increasingly used in STEM-focused or design and engineering environments. In this paper, we refer to the units as CT-STEM pop-ups or pop-ups. Here, we report findings from the project's second year as we sought to understand students' data science practices when participating in the pop-ups. Our research question guiding this study is as follows:

RQ1. "How do elementary students engage in data science practices during CT-STEM Pop-Ups implementations?"

Data science education in K-12 settings

Data science education is considered important to K-12 students' development due to the abundance of data in today's world and the need to prepare students for disciplines and careers of the future that are heavily reliant on data (Weiland and Engledowl, 2022). The integration of data science education across K-12 curriculum allows students to become effective data users and creators while honing important data skills, such as data collection, visualization and analysis, which are needed to make sense of the data around them (Engel, 2017; Martinez and LaLonde, 2020). However, data science education in K-12 is still evolving [National Academies of Sciences, Engineering, and Medicine (NASEM), 2023] necessitating research on how data science can be effectively integrated across K-12 curricula in ways that promote inclusive access for students to understand and think critically about the role of data in society.

While data science education for elementary-aged students is relatively new, a few efforts offer and study data science experiences with children. For instance, initiatives that promote increasing CT and data science in K-8 classrooms include efforts by digital promise (Weaver, 2022) to provide downloadable resources and example computer science (CS) activities from teachers, and rubrics and assessments aligned with Next Generation Science Standards (NGSS; NGSS Lead States, 2013) to support data science practices in middle school classrooms. Their instructional model encourages using data sets from real-world problems and using the data practices of collecting, analyzing, evaluating and communicating data. Efforts to integrate data science into K-12 curriculum have resulted in initiatives such as Guidelines for Assessment and Instruction in Statistics Education II (GAISE II) report which promotes statistical problem-solving (Bargagliotti *et al.*, 2021) and the Data Science K-10 Big Ideas by Boaler and colleagues at the Standford Graduate School of Education (Stanford Graduate School of Education, 2023). The GAISE II report encourages engaging K-12 students in problem-solving activities and data/statistical practices that include formulating investigative questions, collecting and examining data, analyzing data and interpreting statistical results (Bargagliotti *et al.*, 2021). The Data Science K-10 Big Ideas build on the GAISE II report and centers on increasing data literacy at elementary level by connecting data science big ideas like questioning, data collection, analysis and interpretation to existing contents that teachers already teach at this level of education.

While these initiatives signify important steps towards advancing data science education in K-12, recent studies have emphasized the need for more research on how data science education can be effectively integrated in K-12 in ways that are inclusive and relevant to students' diverse environments (Kim, 2021; Liston *et al.*, 2022; NASEM, 2023; Weiland and Engledowl, 2022). For example, in a collaborative case study with researchers, teachers and elementary students, Liston *et al.* (2022) explored the impact of integrating data science and an internet of things (IoT) weather station on elementary students' (aged 11–12) STEAM (science, technology, engineering, arts and mathematics) practices. Using IoT-connected sensors and an Arduino board, students collected real-time environmental data (light, temperature, humidity and sound) from their surroundings. The data were sent to a Web application allowing students to visualize, extract meaning and make data-based decisions. The study found that allowing students to decide how they explore and work with data enabled them to develop crucial STEM skills including recognizing patterns, using models, creating and interpreting graphs, communicating results and developing reason-based arguments. Liston and colleagues reported students' mixed attitude toward the weather station project and advocated for future data science projects closely aligned with children's daily experiences. Similarly, Kim (2021) assessed the impact of a problem-based data science project on elementary students' CT and creativity. Using MIT's App Inventor program, students planned a school trip, followed problem-solving steps encompassing defining data problems, data collection, analysis, generalization, prediction and storytelling. The study revealed a significant connection between problem-based data science activities and heightened CT proficiencies like problem solving and data interpretation among students.

These projects inspired our project and point toward the increased emphasis on data science skill-building for elementary learners. However, our work differs in how we approached co-designing curriculum with elementary teachers. We purposely drew on students' interests, based our instruction on locally relevant issues, included digital and non-digital extension activities and incorporated a teacher-friendly instructional framework that empowers teachers to creatively design and implement data science activities across the curriculum (Arastoopour Irgens *et al.*, 2023).

STEM and data science practices in K-12 classrooms

Data science is the interdisciplinary practice of learning from and making decisions with data using modern computational tools (Ow-Yeong *et al.*, 2023). STEM and data science share a close relationship (Liston *et al.*, 2022), as data science leverages the analytical and problem-solving skills fostered by STEM education to make sense of big data, build models, optimize processes and drive innovation across industries. In turn, data science practices like collection, manipulation, analysis and visualization of data are integral for understanding the nature of problems and decision-making across many STEM professions (Roehrig *et al.*, 2021). The US Department of Homeland Security designated data science as a STEM field due to its interdisciplinary nature and focus on foundational components of statistics, mathematics, computing, data analysis and data management (Department of Homeland Security, 2022). In the USA, important disciplinary practices in STEM have been codified in learning standards. These standards define the core practices and ideas prioritized in the classroom. Reynante *et al.* (2020) analyzed three US K-12 STEM learning standards – the NGSS (NGSS Lead States, 2013), Common Core State Standards for Mathematics (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010) and American Society for Engineering Education (American Society for Engineering Education (ASEE), Corporate Member Council, 2008) standards – for K-12 engineering education and identified shared practices across the documents. These practices include understanding the nature of problems, planning and conducting investigations, developing and using models, using computational tools, working with data, evaluating ideas and communicating results.

Educators use a variety of classroom strategies to engage students in these STEM and data science practices. Siverling *et al.* (2018) highlighted evidence-based reasoning as a means to enhance students' scientific and mathematical practices. For example, classroom activities may require students to provide and assess different explanations for observations based on evidence. Bybee (2011) emphasized that students engage more effectively in STEM practices when encouraged to question, conduct relevant investigations, analyze data using technology and communicate ideas through diverse mediums like diagrams, pictures, tables and graphs. Sulaeman *et al.* (2021) suggest the key to engaging students in STEM practices necessitates students' connecting classroom learning to real-world problems and relevant experiences. Similarly, engaging in hands-on experiences and flexible projects using relevant materials that meet students' individual needs and interests can increase engagement in a STEM classroom (English, 2017; Santana *et al.*, 2020). Carmona Reyes *et al.* (2021) posit that giving students choices and opportunities for self-directed exploration can increase their motivation and engagement by allowing them control over their learning. They note that students benefit from creating, interpreting and communicating meaning through multiple types of representations such as diagrams, graphs and tables.

Wu and Rau (2019) explained that drawing on activities and technology, such as animations and interactive learning tools, helps students visualize complex concepts and engage with science materials in new, exciting ways. They emphasized that these strategies depend on individual student's needs and interests, the class's specific context and the STEM concept being taught. As such, STEM teachers need to continually evaluate and adjust their engagement strategies to meet their students' needs and seek student feedback to determine the most effective strategies.

Theoretical and conceptual frameworks

Connected learning theory to envision designing curricular activities

Connected learning theory (Ito *et al.*, 2013) was used to theorize our work and assist in co-developing interest-based, potentially engaging curricular data science activities. Connected learning is grounded in social-constructivist theories (Vygotsky, 1973) and suggests effective learning environments draw on students' personal interests and social support to acknowledge and value individual contributions (Ito *et al.*, 2013). The theory proposes that learning and interest are linked, and when teachers or mentors promote students' pursuing their interests, there are positive outcomes, including academic achievement and increased engagement. When connected learning principles are used in education, they lessen the gap between in-school and out-of-school learning by recognizing diverse pathways for students to build and express knowledge with peers. We used this framework as it honors students' everyday experiences and aligns with how we and other educational researchers view designing effective learning environments and activities for students by connecting formal and informal learning (Kumpulainen and Sefton-Green, 2014). While we did not explicitly identify connected learning theory when collaborating with teachers in this project, we intentionally used the principles to guide co-creating the units. We encouraged teachers to use interest-based, locally relevant issues that included topics the children cared about such as using data to determine responsible pet ownership, choosing social media platforms and determining what makes music popular when solving data science problems. Thus, by offering data science curriculum linking student interests and learning when solving authentic data science problems, and encouraging group work through teacher facilitation, students may be more interested in problem-solving across STEM contexts (Arastoopour Irgens *et al.*, 2023).

Developing a data science framework based on CT-STEM taxonomy of practices

To create a teacher-friendly data science conceptual framework, we drew on Weintrop *et al.*'s (2016) widely used CT-STEM taxonomy of practices (Figure 1), which focuses on applying computational practices in STEM areas. The taxonomy consists of four strands: data practices, modeling/simulation practices, computational problem-solving practices and systems thinking practices. During this project, we aligned our data science unit creation with two of the strands: the data and computational problem-solving practices and included several of the sub-strands such as manipulating, visualizing and analyzing data, and choosing effective computational tools, troubleshooting and assessing different computational solutions to problems. We focused on these two strands as they were most appropriate for instruction with young children, and our funded research project was aimed at assessing gains in elementary students' knowledge in the two strands (Author, in review). Data practices are defined as gathering and/or generating data using multiple computational tools, using computational tools to analyze data and draw conclusions and being able to communicate and present data. Computational problem-solving includes decomposing and reframing problems, evaluating pros and cons of computational tools and distilling relevant information from a problem (Shute *et al.*, 2017).

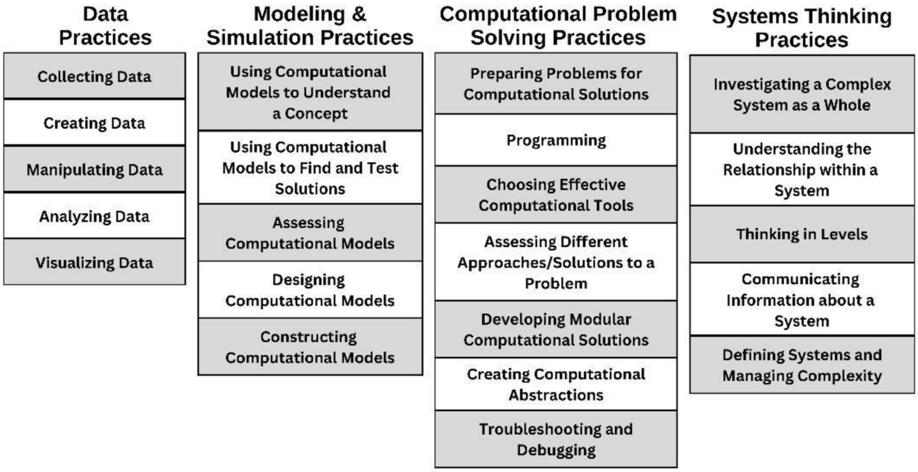


Figure 1.
CT-STEM taxonomy of practice

Source: Weintrop *et al.* (2016)

We then created a framework (Figure 2) guided by connected learning theory and based on the CT-STEM taxonomy of practices that guided teachers in their unit development and expanded on the fourth step, “Play with the Dataset,” using a simple data science cycle (Figure 3).

Research context

Highlands Elementary School (all names are pseudonyms) is a public STEM school in the rural Southeastern USA serving 458 pre-kindergartens to fifth-grade students, aged 3 to 12. Highlands is among the top 10 highest priority states regarding rural schools’ instructional and overall educational needs (Showalter *et al.*, 2019). The school district spans 497 square miles in a sparsely populated region near a mountain range. Nearly 20% of the surrounding population lives below the poverty level. There are 23 teachers, 2 administrators, 2 instructional coaches, 9 specialist teachers (e.g., art, music, physical education), 1 guidance counselor and 1 special educator. The school has a diverse

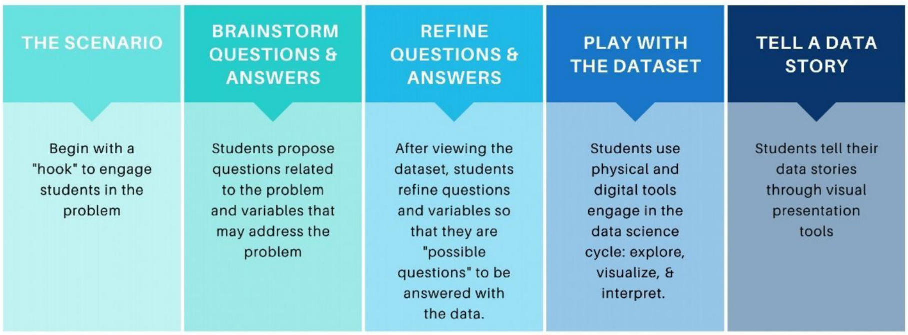
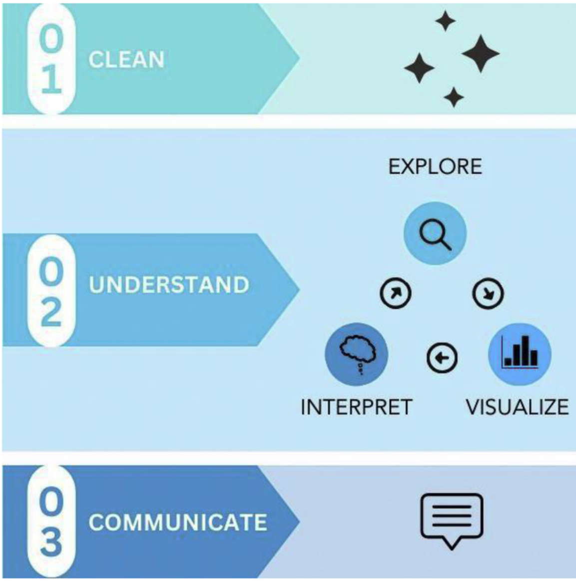


Figure 2.
CT-STEM pop-up framework

Source: Herro *et al.* (2022)



Source: Authors' own work

Figure 3. Simple data science cycle developed by our research team

student population, with 99% of students eligible for free lunch. The student–teacher ratio is 14.8 students per teacher.

To position our study's participants, we first describe the teachers' preparation and experiences as they co-developed the data science units (for more details regarding the unit development and co-design process, see Arastoopour Irgens *et al.*, 2023).

Teacher preparation to implement pop-ups

Our research team met with nine Highlands teachers several times and subsequently worked with them in summer to co-create grade-level data science units. Three teachers taught third grade, another three taught fourth grade, two taught fifth grade and a single teacher was responsible for fifth grade music instruction. Three teachers had five years or less of teaching experience, while the remaining six taught between 11 and 23 years. Despite being a STEM-focused school, none of the teachers had prior data science preparation. Except for the music teacher, all teachers taught general education classes covering several subjects (e.g. mathematics, science and ELA) and thus had flexibility to integrate data science across disciplines. During spring meetings, we administered a pre-survey to help understand their context, student population and instructional needs. Next, the teachers participated in an intensive 25-h professional development (PD) introducing them to data science software and activities for elementary children. During workshops, we assisted teachers in identifying students' interests, writing interest-based data science problem scenarios, and using a simple data science cycle (clean, understand and communicate the data) to facilitate instruction (see Arastoopour Irgens *et al.*, 2023, for details regarding development of the scenarios and units). Teachers engaged in embodied data science activities and used a Web-based data visualization tool, Tuva (www.tuvalabs.com/), to manipulate and explain data. They also participated in rotating workshops focused on designing comics, using 3D modeling and creating videos to interest children in further exploring data science problems. During the PD, each grade level team and the music teacher collaborated with our team to co-create data science units aligned with state standards and performance-based formative and summative assessments.

Classroom instructional context

The pop-up units began by introducing a problem-solving scenario based on students' interests and the learning standards applicable to their grade levels. Each scenario (Figure 4) was designed around issues relevant to students' environment and posed a driving question asking students how they might use data to address the issue in the scenario. For example, in the third-grade class, students were presented with a scenario on determining the best dog for their families by considering dogs' traits and their family needs and/or preferences. The scenario ended with the driving question, "What dog would be best for your family based on data?" In the fourth-grade class, teachers challenged students to choose the most productive social media to help individuals meet personal or professional goals. The students answered the driving question, "How can an individual make data-based decisions when choosing their social media platforms?" The fifth-grade students completed two pop-up units in their general education and music classes. However, for this paper, we focused on the fifth-grade students' activities and data practices during their music pop-ups as the music teacher instructed two fifth-grade classes back-to-back, and we had more observational data. The other fifth general education classes were taught concurrently, making it difficult for our research team to observe several classrooms simultaneously. The students examined how music producers can decide what makes a song popular by exploring a Spotify data set of top-rated songs in the past 10 years. Students answered: "What elements need to be considered when writing and producing a popular song?"

The teachers instructed the students for 8-to-10 days over three weeks for approximately 35–50 min per lesson. They guided students to brainstorm data used in the scenario, conduct research and follow the data cycle (Figure 3) to address their respective driving questions.

Participants

Our participants included 202 students in grades three, four and five who participated in the 8–10-day data science unit co-developed by their teachers and our research team. Specifically, 86 participants were in third grade, 70 were in fourth grade and 46 were in fifth

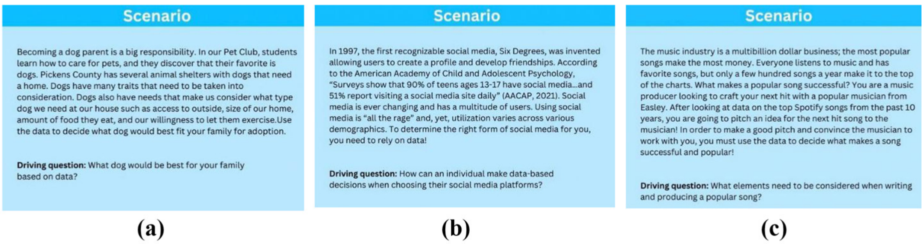


Figure 4. Data science scenarios addressed in the 3rd, 4th and 5th grade classes

Notes: (a) 3rd-grade data scenario and driving question; (b) 4th-grade data scenario and driving question; (c) 5th-grade data scenario and driving question

Source: Authors' own work

grade. In total, 25 of these students participated in focus group interviews after completing the units (Table 1).

Methodology

We used a phenomenological qualitative case study design to "understand the essence of the experience and lived phenomenon by several individuals" (Creswell and Poth, 2016, p. 78). Phenomenological research focuses specifically on the meanings individuals give to their experiences. Case study further guided the analysis of this phenomenon as we used multiple sources and aimed to develop an in-depth description of the case (Creswell and Poth, 2007). Our case was bound by 202 elementary students participating in data science curricular activities during one 8–10-day unit in five different classrooms and three grade levels at one school site (Table 2). This research design is appropriate as our goal was to understand the students' collective experiences (Creswell and Poth, 2007) while participating in the pop-ups and how they made sense of the activities. By focusing on children's lived experiences as they conceptualized data science, engaged in data practices and used data to answer questions that interest them, we deconstruct, reconstruct and describe the process of how elementary children develop an understanding of data science and connect it to personal experiences. In this case study, we explain how students were introduced to data science through activities that engaged students in investigating phenomena they encounter in their everyday environment and how they used data storytelling to demonstrate their learning. We acknowledge that although the case is defined by students across five different

No.	Name	Age	Gender	Race	Grade
1	Dane	9	Male	Two or more	3rd
2	Charlie	9	Male	White	3rd
3	Miranda	8	Female	Black	3rd
4	Lydia	7	Female	White	3rd
5	Brian	8	Male	Black	3rd
6	Maddie	7	Female	White	3rd
7	Anne	9	Female	Black	3rd
8	Johnnie	7	Male	White	3rd
9	Izzie	9	Female	Two or more	4th
10	George	9	Male	White	4th
11	Bianca	9	Female	White	4th
12	Howie	9	Male	White	4th
13	Hugo	11	Male	Two or more	4th
14	Heather	10	Female	White	4th
15	Lilianna	9	Female	White	4th
16	Candace	9	Female	White	4th
17	Renee	10	Female	White	5th
18	Augie	10	Male	White	5th
19	Deon	10	Male	Black	5th
20	Serena	10	Female	Hispanic	5th
21	Eddie	10	Male	Hispanic	5th
22	Heidi	10	Female	Two or more	5th
23	Ryan	10	Male	White	5th
24	Theresa	10	White	Male	5th
25	Sam	10	White	Male	5th

Source: Authors' own work

Table 1. Demographics of the interview participants

classrooms, not all students experienced the same classroom conditions, and there was no way to control for variation between teacher's implementation approaches. That said, each teacher used the same data science instructional framework to guide their unit co-creation and implementation.

Data sources

Data sources included classroom observations conducted during implementations, student focus group interviews, video recordings and photos of implementations, and student and teacher artifacts, including teachers' pop-up units and student data stories.

Observations We designed an observation protocol (Appendix) to collect data on how the students interacted with the activities. The observation protocol was revised from a similar protocol successfully used by Author Two to capture instruction and learning in STEM contexts (Quigley and Herro, 2016; Herro *et al.*, 2018) and designed to capture the physical setting, activities and interactions (Patton, 2002). The open-ended protocol noted the children's behavior, opinions, interests and challenges as they engaged with teachers, peers and the data science activities. We observed whether students expressed an understanding of data sciences concepts, how they engaged with tools and the data-based choices students made.

Focus group interviews After the unit implementations, we interviewed 25 children in focus groups to detail their experiences during the pop-ups. Focus group interviews were intended to encourage group members to hear one another's responses and make additional comments to get high quality data in a social context (Patton, 2002). We purposely chose, with teacher input, a diverse group of students (varied ages, races and reading abilities) from all three grades. The interviews were conducted per grade level; we interviewed nine fifth-grade students, eight fourth-grade students and eight third-grade students. Students responded to similar questions based on three themes: the children's experience with data science, including what they liked and what they found challenging; the CT-STEM pop-up data science framework (Figures 2 and 3), data concepts and practices. We asked questions like, "What do you think data science is? How would you explain data science to a friend? During the Pop-Up, what activities did you like and why? Was there anything that was hard for you?"

Video recordings and images Portions of each lesson were recorded for review during analysis; we focused on videoing key moments in each classroom to capture lesson introductions, student worktime and student presentations. The videos helped us understand how children engaged in data science practices, including how they collaborated and worked with data science tools. Several recordings included students' brainstorming sessions, group projects and data story presentations. Additionally, we took photos to document the research process. The recordings and images were secondary data sources for

Table 2.
Unit topics, grade levels and driving questions

Unit topic	Grade level	Driving question
The best family dog	3	What dog would be best for your family?
Social media platforms	4	How can an individual make data-based decisions when choosing their social media platforms?
Popular music	5 (music)	What elements need to be considered when writing and producing a popular song?

Source: Authors' own work

triangulating evidence of students' data science engagement and practices. Both were used as a "means of remembering and studying detail that might be overlooked" (Bogdan and Biklen, 2007, p. 151).

Artifacts We collected digital and nondigital artifacts created by students as during data activities. These varied based on the kind of activities completed in students' respective grade levels. Digital artifacts included graphs, data stories, slides and designs created by students using tools like Tuva (https://tuvalabs.com/), Canva (www.canva.com/) and Tinker CAD (www.tinkercad.com/). Nondigital artifacts were mostly paper graphs and post-it notes reflections created by students. We also included the pop-up units that guided the teachers' instruction; this assisted us in understanding the data problem and flow of the lessons. These artifacts were used as secondary data sources to document the students' data practices and help answer our research question of "how" students engage in data science practices during pop-up implementations.

Data analysis

Two research team members independently conducted an intensive reading of all qualitative data creating memos of participants' actions and statements (Charmaz, 2003) related to understanding and describing the data science implementation activities and students' responses to them. Initial broad data patterns were noted and discussed with a third research team member. Next, data were imported into MAXQDA (www.maxqda.com/) software for organization and analysis. A priori codes were generated from the CT-STEM Pop-up Framework, which included five steps (e.g. the scenario, brainstorm questions, refine questions, play with the data set and tell a data story), and helped us understand teachers' planned activities across the five classrooms. Next, we used open coding to code student data science practices (analyzing and visualizing data, interpreting graphs, using data science tools, exploring data and data problems). We did a second reading of the data and created categories. We began with the observations and transcribed interviews, using the videos and photographs to verify the observations. Finally, artifacts (curriculum and data stories) were analyzed as secondary sources to better understand how the teachers created and implemented the pop-ups units, and the students response to the activities. Data were triangulated across data sources until reaching saturation. In this case, we achieved saturation when we believed a depth of understanding was achieved through the number and content of our codes (i.e. teachers' planned activities and student data science practice across data sources; Saunders *et al.*, 2018). Our team met several times to compare and winnow codes, discuss and form categories and reach a consensus (Cascio *et al.*, 2019). Using an interpretive approach, the categories were analyzed and developed into themes (Creswell and Poth, 2016) to assist in explaining the students' experiences. Table 3 shows representative data sources, codes and categories.

Findings

To answer our research question, "How do elementary students engage in data science practices during CT-STEM Pop-up implementations?", we analyzed the teachers' presentation of the data science problem and the ways the children responded to it through brainstorming, reflection, discussion and roleplaying; the process the children followed to refine, analyze and visualize their data; and students' data stories. This analysis and reporting are based on the CT-STEM pop-up framework (Figures 2 and 3); this framework was grounded in the CT-STEM Taxonomy of Practices and connected learning theory. We describe this process and our analysis as three chronological phases of the pop-up lessons. Our first theme addresses students' engagement in the first three steps (scenario, brainstorm

Theme	Code	Definition	Example
Contextualizing data science	Explaining data science and data science concepts	Comments, observations and evidence exemplifying how students explain data science concepts using their own words	*Both have more informational data, but mostly numerical can have numbers categorical mostly have words, but category has some numbers.* Source: 5th grader, Focus Group Interview
	Connecting data science to previous knowledge or lived reality	Comments, observations and evidence exemplifying how students related the pop-up scenarios to their personal lives	*Three different kids told me personal stories about dogs and information they learned from the graph. "I didn't know my bulldog would only live 10 years and now I'm sad."* Source: Researcher Observation, 3rd grade class
	Associating data science with real-world problems	Comments, observations and evidence exemplifying how students associated data science with problem-solving or decision-making	*Data science you have to like, you have to use like data and graphs and stuff to like figure out science and make decisions.* Source: 3rd grader, Focus Group Interview
Practicing data analysis and making data-based decisions in multiple ways	Exploring data and data problems	Comments, observations and evidence exemplifying students' investigation of driving questions and data to answer questions	*My question was how many men use the most social like, was the most popular social media that made news. And so that's why I wanted to do it was because I wanted to know if it was the same as women. So, I put both men and women on the same graph so I could compare.* Source: 4th grader, Focus Group Interview
	Analyzing and visualizing data	Comments, observations and evidence demonstrating students' engagement in data analysis and visualization using graphs during pop-up	*One of the kids already grouped the dogs by whether they were good with children or not, and upon visualizing the graph he said, "Chihuahua are mean even though they are little."* Source: Researcher Observation, 3rd grader class
	Using data science tools	Comments, observations and evidence exemplifying students' use of tools like post-it notes, Tuva, Canva and TinkerCad	On Tuva, we are researching stuff about dogs and then we answered that question by using the data that we use on Tuva, and we link that to Canva. Source: 4th grader, Focus Group Interview

(*continued*)

Table 3.
Codebook showing themes, codes and code examples

Theme	Code	Definition	Example
Communicating patterns and interpreting graphs	Interpreting graphs	Comments, observations and evidence demonstrating students' ability to interpret and draw conclusions from the data graphs	The kids concluded "*If you want a song to be popular, make sure it has a duration of about 200 seconds and BPM of between 100 and 180.*" Source: Researcher Observation, 5th grade class
	Telling data stories	Comments, observations and evidence of students sharing their results and explaining how it answered their driving question	We did our chart and then we wrote in a piece of paper, and we shared it up on the board. Source: 3rd grader, Focus Group Interview
Challenges	Challenges	Comments, observations and evidence exemplifying data science concepts, activities and tools that students find challenging or difficult during the pop-up	*It's a challenging tool that gives you real good information, but it sometimes you wouldn't like it because it gives you too much information.* Source: 5th grader, Focus Group Interview

Source: Authors' own work

Table 3.

questions and answers and refine questions and answers; approximately Lessons 1–4) of the CT-STEM pop-up framework, our second theme addressed the fourth step of the framework (play with the data set; approximately Lessons 5–7), while the last theme addressed the fifth step (tell a data story; approximately Lessons 8–10) of the framework. See Table 4 for representative examples of themes across all grade levels.

Contextualizing data problems and connecting data science to lived experiences

The lesson began with the teachers introducing the data science framework to the students and explaining the data science cycle as collecting, analyzing and using data to make decisions. This first phase of the pop-up emphasized understanding data and brainstorming new questions. It also focused on whole class and group discussions aimed at contextualizing data science in the students' everyday lives.

In each classroom, the teachers presented students with the data-based problem scenario, discussing the challenge within it and ways to approach it. Through various prompts, students reflected on how the driving questions might be answered and information needed to answer them. Students brainstormed and shared opinions about the main ideas identified in the scenario. For example, in third-grade classes, the teacher facilitated a discussion about various traits and characteristics of dogs familiar to the children. The third-grade teachers used multiple methods to elicit students' responses during this activity, including asking students to think about dogs and write preferred dog characteristics on sticky notes (Figure 5). Many students expressed ideas such as "hyper," "messy," "legs," "aggressive," "fat," "helpful" and "protective." This example from the third-grade class demonstrates how the activities allowed the children to connect the data-based problem to their prior experience and externalize their knowledge about the topic.

Next, the third-grade students shared the traits they chose with their classmates and explained why such characteristics were important to them. One child, Freya, stated that she wanted a dog that was good with children because she has "a neighbor who is scared of

Table 4. Themes and representative examples of coded data across all grade levels

Theme	Source: Coded data Third-grade	Fourth-grade	Fifth-grade (Music Class)
Contextualizing data problems and connecting data science to lived experiences	Observation: Freya chose to plot dog breed vs good with children vs weight. She decided to go for this because she had a neighbor who was scared of dogs Group interview: *I think it [data science] helps you more because you can figure out how long it's like expected to live and how big it can get. And if it's good for indoors or if it's not good or good with children*	Observation: The teacher moved on to ask the students to provide examples of social media; they responded: TikTok, Snapchat, Instagram, Facebook and Twitter. Group interview: *On mine, my question was how many men use the most social like, was the most popular social media that made news. And so that's why I wanted to do it was because I wanted to know if it was the same as women. So, I put both men and women on the same graph so I could compare*	Observation: Kids were quite chatty and made observations while the music played, for example one kid said, "There is no country music on the list." Group interview: *I was interested in the popular songs, and oh, okay, that was it and then I like the writing and researching about all of those stuff i wrote down*
Practicing data analysis and making data-based decisions in multiple ways	Observation: Kids explored different graph options [in Tuva] and switched between pie, dot and bar graphs until they found the one they loved. Group interview: *I like it when we could use the graph and we got to, like, screenshot it and like, we got to like use the graph to see what dog was best for our family*	Observation: Students [are] in Tuva working on plots, screenshotting graphs. Writing sentences that summarize plots and data Group interview: *We actually use a bar graph for [choose] the right dog that she needs*	Observation: The class decided to plot highest BPM against Year and Ms. Fisher asked each group to choose a year and write their year and corresponding maximum bpm for that year on a post-it note. Observation: The students had to drag and drop different variables displayed on the left side of the screen into the dot plot. . . they . . . were able to visualize data once they dropped the variable into the chart area on their screens
Communicating patterns, interpreting graphs and telling data stories through student choice	Observation: One of the kids . . . grouped the dogs by whether they were good with children or not, and upon visualizing the graph he said "Chihuahua are mean even though they are little." Group interview: *We did our chart and then we umm we wrote in a piece of*	Observation: During the group sharing, one kid noted that "Siberian Husky has the most average life expectancy" but another kid corrected him saying "Chihuahua is the longest, 16 years" Group interview: *I was very surprised that Pinterest didn't have a lot of popularity for an age*	Observation: This group presented a graph of BPM vs Year. With their graph the group showed that the average duration for popular songs is 200 seconds. The kids concluded that "If you want a song to be popular, make sure it has a duration of about 200 seconds and BPM of

(*continued*)

Theme	Third-grade	Source: Coded data Fourth-grade	Fifth-grade (Music Class)
	paper, and we shared it up on the board	*group. Pinterest, there was one age group [that] didn't have a lot of Pinterest at all*	between 100 and 180." Observation: Another group of 2 females/2 males showed . . . their visualization in Tuva and explained how the highest bars showed the highest BPMs and then told me they learned that high BPMs in music relates to popularity

Source: Authors' own work

Table 4.

Source: Authors' own work

Figure 5.
Third-grade students sticky note showing characteristics of dogs

dogs," and she already had a guinea pig at home. As such, she wanted a dog that was friendly with her neighbors and her pet. While reflecting on how her environment might influence choices of dog attributes, another third-grade student commented, "If you have a small house, you can't take a big dog like a Great Dane because you will need to keep taking it outside and it can't stay outside forever." The class discussion revealed that many students had dogs at home and were eager to share about them. During discussions, the third-grade students connected the scenario to their homes and thought about how a dog might fit into that environment.

The third-grade teachers personalized the lesson by telling the students they would be choosing a dog for their family using data. By sharing their own pet stories, the teachers encouraged the children to share their personal experiences with dogs, thereby activating background knowledge about the dog adoption problem. For example, one third-grade student shared about having a Labrador that was "always bored because no one is home to keep him company." She suggested adopting another dog to play with the Labrador. The decision to connect the data-based problem to students' homes and environments excited the children. It allowed them to personalize the lesson and offer some expert knowledge about their environment and pet choices.

Similarly, in the fourth-grade class, many students discussed experiences with the social media scenario by indicating their family members have social media accounts like Facebook, Instagram and YouTube. When their fourth-grade teacher asked them what they knew about social media, students discussed how social media allows you to "see a lot of good things" and "keep up with the world." Others voiced, "social media can be harmful because people can say mean things" and "some things are actually bad" on social media.

To help fifth-grade students connect the music scenario to their everyday lives, the fifth-grade teacher played top 2021 billboard music that students were likely familiar with and asked them to "guess the artists." The fifth-grade students were excited hearing the music; they moved their bodies, sang along and raised their hands to answer questions about song titles and artists. Several students were familiar with the popular songs and even had debates about the genre of the music. For example, a student commented that "there is no country music on the list," while another student noted how one of the top songs was a "sad song." Next, the whole class moved into a brainstorming session where students worked in groups and were prompted to discuss music features and popular songs. The teacher got the students thinking by asking questions like "What do they [popular songs] have in common?" "What are the music elements?" The students discussed ideas in small groups, recording their thoughts on post-it notes that were later shared with their classmates [Figures 6(a) and 6(b)].

During brainstorming sessions in each grade level, the students had data sets to address the scenario. For example, in the fifth-grade class, students examined a music data set containing 50 songs and attributes like song title, genre, artist, duration, year and BPM printed on paper. The students identified and discussed what variables qualify as numerical or categorical data. Most students could correctly identify variables like artist and genre as

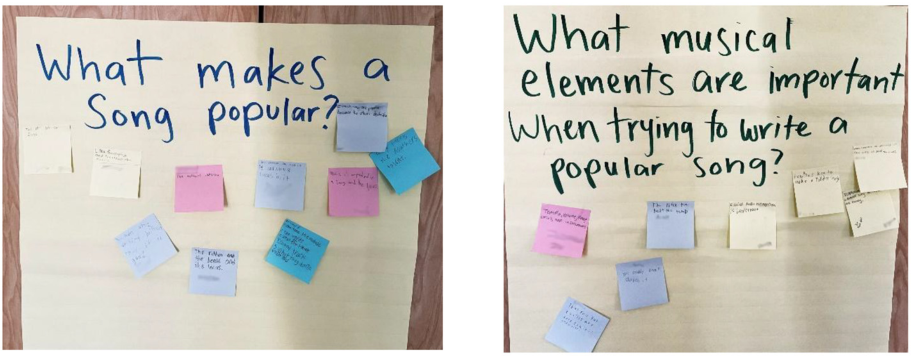


Figure 6.
Fifth-grade students' ideas about what makes a song popular

Source: Authors' own work

categorical data, while duration was determined to be numerical. However, several students were confused about the *year* variable and argued about whether it was a categorical or numerical variable. The teacher addressed the students' confusion by explaining how certain variables like *year* can be numbers and be considered categorical data for analysis. At the end of the brainstorming session, the music teacher showed a chart with examples of categorical and numerical data. Afterward, the students role-played a music producer using the data provided to determine what makes a song popular and how information might be used to produce profitable popular music. By using familiar songs and artists in the data set and asking students to roleplay a job desirable to many fifth graders, the teachers positioned the data problem and analysis in a context in which students had high interest and some experience.

Similarly, in the third- and fourth-grade classes, the students explored datasets related to their dog and social media scenarios. They discussed what each variable in the data represented or measured and whether it was numerical or categorical. Once the teachers established that the students understood the variables and their interrelationships, they prompted them to formulate a question that could be answered with the data. Students could combine and use any of the variables, but they could only address one question. At the end of the brainstorming session, the third- and fourth-grade students shared their questions with their peers and teachers, who then provided them with feedback to refine the questions. In this way, students were able to use data to pose familiar and interesting questions.

Practicing data analysis and making data-based decisions in multiple ways

To prepare for data science exploration, the teachers in each grade level engaged students in a series of practice activities to help students develop an understanding of data analysis and interpretation. The practice activities were intended to model the data science process before having students work on their own datasets or different topics.

In the third-grade classes, students engaged in an unplugged (nondigital) exercise involving choosing the best dog for their teacher's families. The teacher provided the students with a printed data set containing dog attributes like their breeds and sizes and explained the relationship between the data set rows and columns. Afterward, the teacher described her family scenario, stating that she had a big yard, a member of the family with asthma and a son who disliked big dogs. Based on this information, the students decided the variables they needed to keep and those they could "get rid of" to choose a dog that met the teacher's needs. Several students explored attributes like whether the dogs were suitable for indoor or outdoor and whether the dogs were good with children, and they crossed off attributes and dogs that did not meet these conditions. In the end, the students suggested a "Siberian Husky" would be the best dog for her family scenario. However, the teacher informed them that while Siberian Husky ticked all the boxes, the data did not show that "Huskies shed too much," which makes them problematic for asthmatic patients. Based on this realization, the students were encouraged to conduct additional research because data may not always give them all necessary information to answer questions. Here, the teacher offered students opportunities to connect selecting a dog to multiple, familiar experiences in the learning process (practice activities, print options, discussions and talking about families) fostering a sense of urgency to fully consider the data and potential solutions.

The fourth-grade teachers and fifth-grade music teacher scaffolded the students' experience with Tuva and data analysis by completing a practice activity using the dog data set in Tuva to make decisions about adopting a family dog. These teachers intentionally chose the dog scenario as a practice activity before presenting a new scenario (social media and popular song, respectively) to model the data science process. We observed that investigating the dog data set as a scaffold for using Tuva and interpreting data benefitted the students' data science understanding. Through the practice activities, students learned concepts of scaling, plotting axes and graphing elements like labels and legends. Afterwards, the teachers presented the grade-level scenarios they developed for their fourth- and fifth-grade students.

After these practice exercises, the fourth- and fifth-grade students completed another set of activities allowing them to explore their original social media and music data set, respectively. Students were given a data set and worked in groups to decide attributes they wished to explore. In the fourth-grade class, students drew bar graphs showing the timeline of popular social media applications using paper and pencils (Figure 7). In the fifth-grade class, students engaged in unplugged activities that involved using post-it notes to plot a graph of top songs' BPM against their year of release (Figure 8). This physical and embodied activity further facilitated their understanding of how graphs are constructed and prepared them to transition into using computational tools like Tuva for data visualization.

Across all three grade levels, the students used Tuva to create multiple plots that examined the relationship between the different attributes in their dataset. For example, the third-grade students explored relationships among the dog attributes (Figure 9), while the fourth-grade students created plots showing the popularity of social media platforms across different age groups (Figure 10). As they plotted their graphs, students also discussed

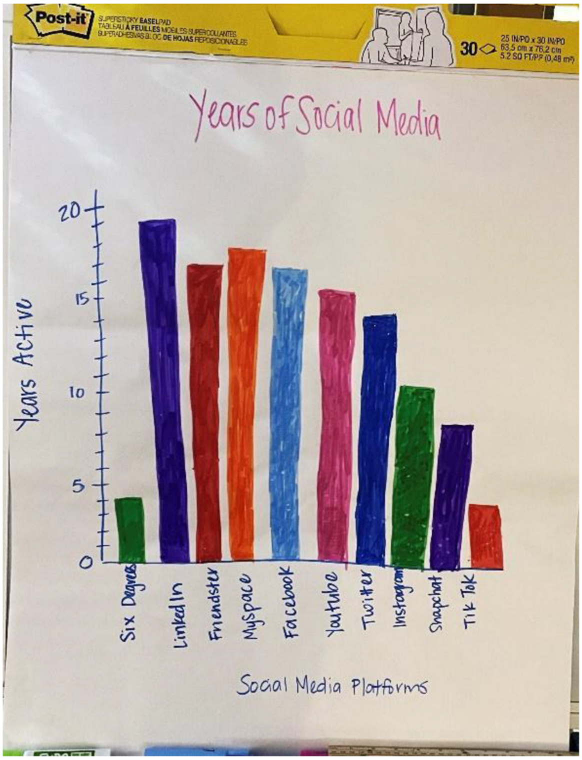
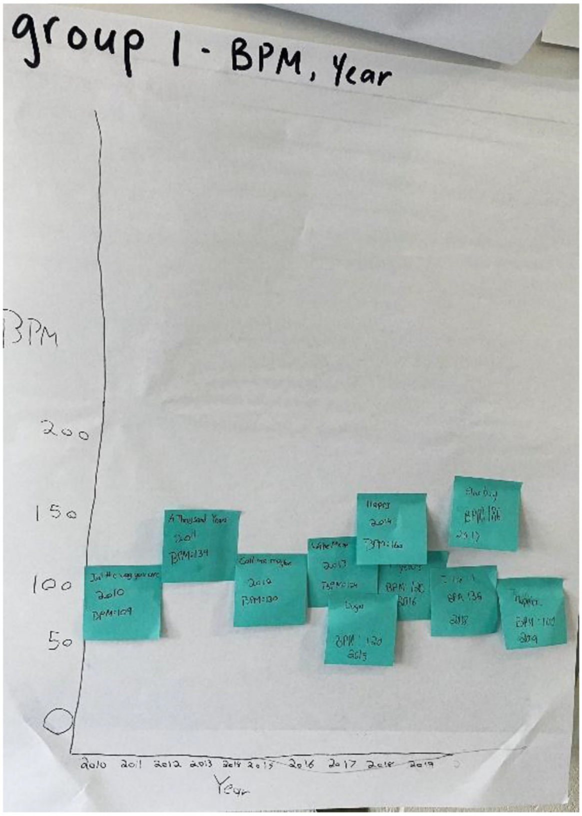


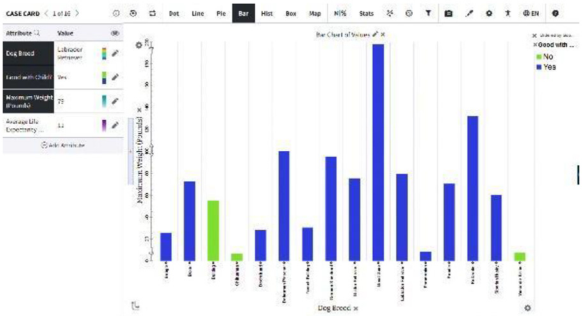
Figure 7.
Fourth-grade students' graph showing the adoption of various social media platforms

Source: Authors' own work



Source: Authors' own work

Figure 8.
Fifth-grade students' graph of BPM vs Year during unplugged activity



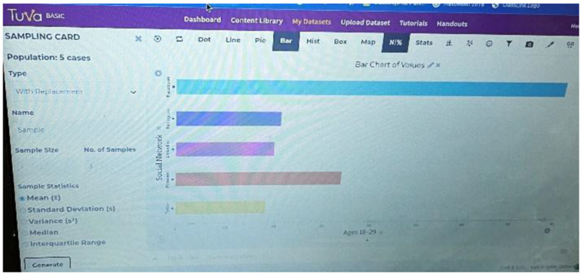
Source: Authors' own work

Figure 9.
Third-grade students' plot of "Dog Breed vs Maximum Weight vs Good with Children" in TUVA

interesting patterns they noticed in the visuals. In the third-grade class, a student saw a pattern between the dogs' weight and average life expectancy, commenting, "The small dogs live the longest while the big dogs live shorter." Similarly, one fourth-grade student commented on how "Facebook is the highest [most popular]" among people in the data set, while another commented that "it does kind of make sense when you think about it." This exchange suggests that even though students have data and evidence explaining a particular phenomenon, they often negotiate such evidence with their own background knowledge and experience with the phenomena. However, students' experiences sometimes conflicted with the data and ways they made data-based conclusions. For instance, during one observation, Zion, a third-grade student, noted that her graph indicated that "Chihuahuas are not supposed to be good with children," which contradicted her personal experience; she stated, "but my Chihuahua is good with children." Despite the data indicating otherwise, Zion maintained her position and concluded that "not all Chihuahua[s] are good with children," allowing her to consider potential bias in the data set. We also noted students reflecting on the implications of their analysis and graphs relative to their lives. For example, after visualizing the relationships between the dog breed and average life expectancy attributes, many students began to imagine how long their dogs might live. They made comments like "My dog has 9 more years until he is dead," "I didn't know my bulldog would only live 10 years, now I'm sad," and "I have a Chihuahua, and he will live to be 16, he is only five."

After practicing activities related to understanding, visualizing and analyzing the data, the students in each classroom worked in groups of three-to-five to solve their original scenarios using the dataset curated by the teachers. Although each scenario was accompanied by a driving question that guided the specific problem students were meant to address, the students had flexibility to choose how to answer the questions and what variables they wanted to investigate while developing solutions to the data-based problem. The questions or attributes students explored were primarily influenced by factors like students' homes, environment and interests. For example, in the fourth-grade class, a group of students narrated how they chose to investigate the popularity and usefulness of LinkedIn because they imagined themselves as "24-year-old trying to find a job." Another student was interested in the gender differences among users of popular social media sites and wanted to know which was popular among groups. He explained:

> On mine, my question was how many men use the most social like, [what] was the most popular social media that made news. And so that's why I wanted to do it, because I wanted to know if it was the same as women. I put both men and women on the same graph so I could compare. ~ Howie, grade 4.



Source: Authors' own work

Figure 10.
Fourth-grade students' TUVA plot of "Social Media Type by Age Groups"

As students prepared their final projects, they plotted multiple graphs and visualized the relationship between the variables in their data set. They continued to express excitement when seeing patterns that interested them in the graphs. The teachers primarily facilitated the activities, walking around to discuss students' claims or help with technology-related issues. Some students found using the visualization tool, Tuva, challenging and needed the teachers' support to choose appropriate visualization options. For example, some third-grade students complained about Tuva having "too many [graphing] options," making visualization more "complicated" and "challenging."

Additionally, some students felt that the driving questions limited their ability to analyze the data problem. Allison, a fifth-grade student, shared: "I don't like driving questions because it's that one thing you have to go by." The driving question was intended to give students clear direction and help focus their analysis on solving a specific problem. However, Allison felt this was limiting because it is the "one thing you have to go by" and limited her from pursuing other interesting ideas as she analyzed and worked with the data set.

Communicating patterns, interpreting graphs and telling data stories through student choice

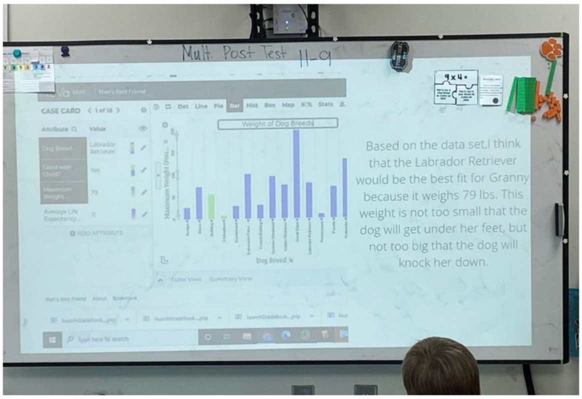
The final phase of the pop-up lesson started with children presenting their data stories and providing peer feedback based on their data inferences and conclusions. In all grades, students created their data stories using graphs they created in Tuva, adding their own interpretation of the graph and relying on information gathered from their research about the scenario. Students were given the choice to tell their stories using various digital and non-digital mediums. For example, students could tell their stories using Google Slides – a presentation program, Tinker Cad – a 3D modeling platform or Canva – an online graphic design tool. They also had the option of telling their stories by creating posters, using colored paper or post-it notes.

The teachers introduced students to data storytelling by modeling how to tell a data story, and presenting students with a rubric detailing important ideas a good data story should contain. For example, in the third-grade class, the teacher modeled a data story by examining the driving question, "How can we choose what dog is best for Granny?" The teacher and students brainstormed the problem's peculiarities and considered how attributes like the average life expectancy and dogs' disposition were less important in this context as Granny was not a child. Next, the class plotted a graph of dog weight vs. dog breed and later used the graph to tell a story. In her data story (Figure 11), the teacher included the justifications for not using certain variables in the data and why the dog's weight was important. The teacher concluded:

> I think the Labrador will be the best for granny because it weighs 79 pounds, and the weight of the dog is not too small such that the dog will get under granny's feet neither is it not too big that it will knock her off.

Similarly, in the fifth-grade class, students were provided with a storytelling template introducing the choice of attributes and potential interpretation for their graphs, "We chose [. . .] and [. . .] to be our variables [. . .] one thing we saw in our graph that was interesting is [. . .]."

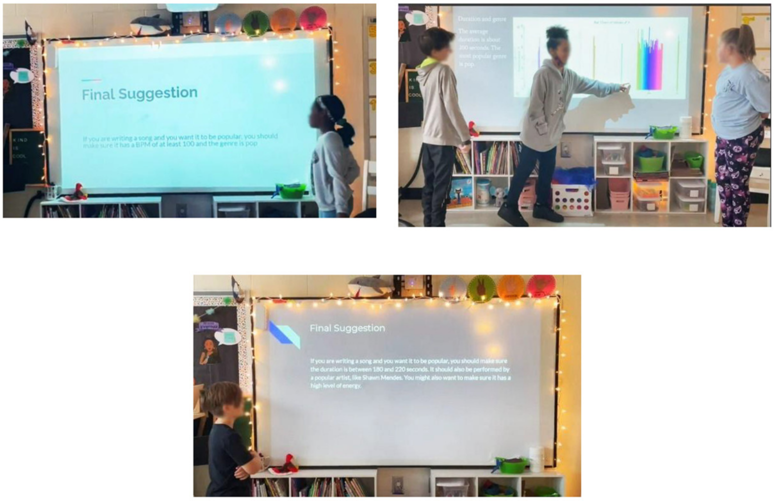
After determining their preferred data story format, students developed their stories using the rubric and engaging in several peer reviews and revisions of their data stories. The data stories were developed around their chosen driving questions, attributes they chose to examine in their final project, and graphs they plotted in Tuva. The students



Source: Authors' own work

Figure 11. A third-grade teacher's example showing student how to develop a data story

selected varying attributes while exploring and visualizing their data. At the end of their analysis, each group took screenshots of their graphs and exported them into their chosen data storytelling programs. For instance, within Canva, fourth-grade students combined screenshots of graphs with text, animation and sound and provided context about the scenario and made claims based on their data analysis (Figure 12). The students shared "aha" moments as they interpreted their graphs and developed their data stories making comments such as "According to data [. . .] YouTube has [the] highest usage, used for videos." The fifth-grade students used Google Slides to communicate their stories and shared responsibilities in their group as they prepared the narrative. In the final interview, one of the fifth-grade students explained how one group member worked on introducing the problem and providing information about the musical vocabulary, another worked on



Source: Authors' own work

Figure 12. Students presenting their data stories

taking screenshots of their plots in Tuva and interpreting the data the graph conveyed, while another developed the slides into a story of how BPM makes music more energetic, danceable and thus more popular.

In each classroom, the students shared the task of presenting their data stories to peers and their teachers. As they presented, the teachers asked probing questions compelling the students to use data to explain how the data and graph supported their claims and stories. Questions included, "How did you know that [the claim]?" "Does the graph tell us that?" Depending on the scenario and the variables students explored, they came up with a primary claim or statement addressing the driving question and used this claim as the primary talking point for their data stories.

While presenting their stories in each classroom, the students typically had a slide or scene describing their scenario or problem, information about their data set, graphs they plotted during analysis and their claim based on the graph/data. For example, while addressing the question of what makes a song popular, one group of fifth-grade students claimed: "BPM should be at least 100 and genre should be pop if you want to make a popular music." Whereas another group of students noted, "the most important elements to consider when writing a popular song is genre and duration." The group suggested that for a song to be popular, it should have "a duration of about 200 seconds." Several other students concluded the "genre [of the song] should be pop and the energy level should be at least 60" for it to be popular. Students further checked the relationship between the artists and energy variables to look for popular music. They suggested a duration of about 200 s and a high-energy performance by a famous artist was important for a song to be popular. During the final interview, some fifth-grade students expressed displeasure at not being able to use all the information from their analysis and research. One student said, "It's just sad that that one part gets left out [. . .] the other information you got, but you can't put it in there." This student was unhappy they only used the most significant data about the topic when telling their data stories, while the other information about the phenomena "gets left out."

The teachers also modeled how to provide constructive feedback that recognized the strength and areas of improvement in their peers' presentations. Typically, the student feedback focused on how well the data story answered the driving question and the reliability of their conclusion based on the data. Students provided feedback by giving stars or writing on sticky notes.

In summary, key findings revealed that contextualizing data science problems to align with students' background knowledge and interests enhanced their grasp of data problems. This approach encouraged students to formulate questions, establish connections with data variables and interpret these variables in personally meaningful ways. Teachers' modeling of the CT-STEM data science framework and interest-based practice activities empowered students to actively engage in data science practices, including data filtering, analysis, visualization and evidence-based conclusions. Ultimately, aligning data science activities with the principles of connected learning fostered a personalized approach to students' analysis and interpretation of data.

Discussion

Based on our findings, we offer suggestions for ways researchers and educators might promote and develop elementary school children's data practices by integrating engaging data science activities in classrooms.

The study highlighted the value of leveraging children's real-life experiences and data storytelling to make data science relevant. The teacher-friendly CT-STEM pop-up framework (Figure 2) offered guidance for teachers providing a roadmap for them to develop and implement activities situating students' experiences and promoting their engagement in data practices. As they moved through activities guided by the framework, students examined relevant data problems, worked with data using computational tools like Tuva and communicated results through data storytelling (Reynante *et al.*, 2020). We relied on connected learning theory to help teachers contextualize the data science lessons to topics students cared about. The theory provided a foundation for teachers to design activities to assist students in connecting their in-school experiences to out-of-school experiences, serving to increase their engagement (Ito *et al.*, 2013). The locally relevant nature of the data science problem facilitated students' ability to interpret their graphs and connect them to their own lives. Students were able to reflect on the personal relevance of the data problem, and they saw how their analyses "made sense" in the real world. Placing the problem within the context of children's experiences and interests also allowed them to be familiar with the data's attributes and variables. For example, children knew that dogs have weight and that some dogs can be "mean" while others can be friendly. This familiarity facilitated students' ability to analyze the data in terms of knowing what relationships to explore between two or more variables. It enhanced their ability to interpret the data in personable ways (Santana *et al.*, 2020). By connecting data science concepts to students' lives, educators can create a more engaging and relatable learning environment (Wu and Rau, 2019). This may enhance students' understanding of data analysis and interpretation while fostering a sense of personal relevance and ownership over their learning (Counsell and Wright, 2016; Slater and Sanchez, 2021). Thus, by leveraging children's lived experiences, educators can empower students to become active in using data in ways that equip them to recognize trends, relationships and correlations, which are crucial skills for developing analytical thinking (Franco and Patel, 2017).

Both connected learning (Ito *et al.*, 2013) and results of recent studies (Liston *et al.*, 2022; Sulaeman *et al.*, 2021) emphasize the importance of designing data science instruction congruent with students' context and experiences. However, our study's findings also illuminate a nuanced dynamic. While students' familiarity with and immersion in real-life scenarios fostered a meaningful engagement with data problems and variables within the data set, this familiarity, at times, introduced distractions during the data analysis and interpretation phase. Students negotiated tensions between empirical evidence and their personal realities. The resonance between the data problem and students' lives instilled expectations of patterns or outcomes within the data, and confirmation of these expectations often led to enthusiasm. In contrast, instances where results of the data analysis deviated from students' preconceptions prompted surprise and occasional conflict. This was evidenced in experiences like Zion's as she asserted that her own "chihuahua[s] are good with children" in contrast to what the data suggested. This finding underscores the intricate relationship between students' personal experiences and the objective analysis of data, implying that effective data science pedagogical practices necessitate teachers acknowledging and discussing these "data incongruencies" with students.

Our findings extend what is known about how elementary teachers might acknowledge and respond to ways student negotiate personal meaning while analyzing and interpreting patterns in data. Teachers in this study assisted students to make evidence and data-based conclusions (Siverling *et al.*, 2018) from their analysis by asking questions that compel students to refer to the data and justify how it supports their conclusion. However, we posit teachers can also help students resolve inconsistencies and make effective connections between in-school and out-of-school experiences with data by acknowledging that most data sets are incomplete and assisting students in understanding that data have embedded assumptions and biases. Teachers can ask clarifying questions helping students think critically about the history and context of the data they use for analysis. This may foster a more nuanced understanding of data analysis and interpretation by creating a space for dialogue and negotiation to provide students with a more holistic understanding of data and potential biases.

Our findings also highlight the critical role that practice activities (to model learning), unplugged learning activities and visualization tools like Tuva play in enhancing students' data practices (English, 2017; Santana *et al.*, 2020; Wu and Rau, 2019). The initial hands-on exploration of the data problem and data sets through teacher modeling and whole group practices with common data sets, using unplugged activities such as categorizing variables and using post-it notes to plot graphs, facilitated students' understanding of data concepts (i.e. types of data) and fostered engagement in practices like data filtering, and visualization. Bybee (2011) emphasized that practices like data filtering and understanding the nature of variables are critical to planning and conducting systematic investigation in science and engineering as scientists and engineers often need to clarify what counts as data and variables before conducting experiments or designs. In our case, students engaged in data filtering by "get[ting] rid" of variables irrelevant to problems they were interested in solving. They also externalized their understanding of the data by explaining the properties of the variables they chose to investigate and why it was important to them. Besides providing opportunities to better understand the data concepts, the unplugged hands-on activities also served as an effective segue for introducing students to data analysis and visualization using Tuva. Within Tuva, students were able to explore their data using multiple visualization options, including line graphs, bar graphs and dot plots. This self-directed exploration and meaning making with Tuva promoted students' ability to interpret and summarize relationships in data (Carmona Reyes *et al.*, 2021). Unlike the physical activities where students were limited to one graph that the teacher suggested, Tuva provided more choices and gave students flexibility and freedom to easily explore the aggregate patterns in the data and how variables in the data vary among individual units or cases. For example, a single plot in Tuva allowed students in the third grade to notice the pattern between the sizes and average life expectancies of dogs.

Generally, students were able to represent relationships between data variables in different ways and were able to produce graphs that conveyed the same idea. We believe the efficiency and flexibility of tools like Tuva made it easier for students to visualize and interpret data (Carmona Reyes *et al.*, 2021), make meaningful interpretations as they interact with the graphs and draw inferences about their own environments based on the aggregate patterns revealed in the data. We observed how students were sometimes focused on the look of their graphs (e.g. color and layout) which tended to distract them from the meaning the graphs conveyed. As such, teachers need to support students to move beyond focusing on the graph's appearance to examining the relationship or pattern it represents. That said, most students did not have any challenges interpreting the general patterns revealed in their graphs. However, familiarity with data/topics sometimes made students question aggregate patterns in the data. Consequently, teachers should encourage students to think about whole and individual patterns in datasets as well as outliers that do not support general trends in data.

Conclusion

With limited research on ways elementary teachers might incorporate engaging data science instruction in their everyday practices, we observed and listened to children's perspectives to address this important research gap. Grounded in theoretical (Ito *et al.*, 2013) and conceptual frameworks (Weintrop *et al.*, 2016) that support using problem-based activities and heightening CT instruction (Kim, 2021), and literature promoting inclusive and relevant data science instruction (Liston *et al.*, 2022; NASEM, 2023), this work provides a roadmap for elementary educators to consider pedagogical approaches towards data science instruction for young children. We assist in answering the call to build a body of research and practical models for data science pedagogies; research that is still in its infancy (Pangrazio and Sefton-Green, 2020). In this study, relying on a CT-STEM framework guided teachers' pedagogical practices, which in turn impacted the students' practices. Our research found that students effectively engaged in data science practices when data problems and activities were connected to their everyday experiences, when they practiced data analysis and making data-based decisions in multiple ways, and when they were encouraged to communicate inferences and conclusions through presentation choices when telling data stories. Similar to Liston *et al.*'s (2022) study, our findings underscore the need to use relevant data and real-world problems to engage elementary children in data science practices. Importantly, our research also points to the need to help students negotiate tensions between their lived experiences and the data they are immersed in. We also acknowledge several limitations to our research. Our sample size was relatively small and inclusive of rural students and may not represent the context of all elementary classrooms. Some findings are limited by self-reported data from students and reflects what participants choose to share. We acknowledge the possibility of confirmation bias (Nickerson, 1998) as interest in this study originated from our existing beliefs about how student might engage with data science; this may have influenced how we theorized our hypothesis. Although our work was conducted in a rural setting, we posit drawing on students' lived experiences, supporting multiple ways to analyze and make decisions about data and providing choices for children to tell data stories is a fruitful way for *all* elementary school children to hone data practices in an agentic manner. Future research directions include expanding students' problem-solving to encompass culturally relevant data problems and studying ways to offer similar data science PD to teachers in remote areas through virtual, asynchronous instruction.

References

American Society for Engineering Education (ASEE), Corporate Member Council (2008), "K-12 engineering guidelines for all Americans", American Society for Engineering Education.

Arastoopour Irgens, G., Herro, D., Fisher, A., Adisa, I. and Abimbade, O. (2023), "Bop or flop?: Integrating music and data science in an elementary classroom", *The Journal of Experimental Education*, pp. 1-25, doi: 10.1080/00220973.2023.2201570.

Bargagliotti, A., Arnold, P. and Franklin, C. (2021), "GAISE II (guidelines for assessment and instruction in statistics education): bringing data into classrooms", *Mathematics Teacher: Learning and Teaching PK-12*, Vol. 114 No. 6, pp. 424-435.

Bogdan, R.C. and Biklen, S.K. (2007), *Qualitative Research for Education: An Introduction to Theories and Methods*, 5th ed., Pearson Education.

Bowen, J. (2021), "Why is it important for K-12 students to understand data and statistics? 'Understanding how data is used, how it's collected and why it's collected helps you understand that you can be empowered by it or you can be manipulated by it,' says professor Hollylynne Lee | college of education news", *College of Education News*, available at: https://ced.ncsu.edu/news/2021/09/21/why-is-it-important-for-k-12-students-to-understand-data-and-statistics-understanding-how-data-is-used-how-its-collected-and-why-its-collected-helps-you-understand-that-yo/

Bradshaw, S. and Howard, P. (2017), "Troops, trolls and troublemakers: a global inventory of organized social media manipulation", (Computational Propaganda Research Project Working Paper No. 2017.12), available at: https://ora.ox.ac.uk/objects/uuid:cef7e8d9-27bf-4ea5-9fd6-855209b3e1f6

Bybee, R.W. (2011), "Scientific and engineering practices in K-12 classrooms: understanding a framework for K-12 science education", *Science and Children*, Vol. 49 No. 4, p. 10.

Carmona Reyes, J.A., Attai, S.L., York, J., Ranney, K. and Hyde, T.W. (2021), "Student engagement as a mediator to increase interest in STEM areas", 1144.

Cascio, M.A., Lee, E., Vaudrin, N. and Freedman, D.A. (2019), "A team-based approach to open coding: considerations for creating intercoder consensus", *Field Methods*, Vol. 31 No. 2, pp. 116-130, doi: 10.1177/1525822X19838237.

Charmaz, K. (2003), "Grounded theory", *Qualitative Psychology: A Practical Guide to Research Methods*, Sage Publications, pp. 81-110.

Coburn, C.E. and Penuel, W.R. (2016), "Research–practice partnerships in education: outcomes, dynamics, and open questions", *Educational Researcher*, Vol. 45 No. 1, pp. 48-54, doi: 10.3102/0013189X16631750.

Counsell, S.L. and Wright, B.L. (2016), "Science learning for ALL young scientists: exploring, investigating, learning, and growing together with ramps and pathways in diverse settings", *Childhood Education*, Vol. 92 No. 5, pp. 365-372, doi: 10.1080/00094056.2016.1226110.

Creswell, J.W. and Poth, C.N. (2007), *Qualitative Inquiry and Research Method: Choosing among Five Approaches*, Sage, Thousand Oaks, CA.

Creswell, J.W. and Poth, C.N. (2016), *Qualitative Inquiry and Research Design: Choosing among Five Approaches*, Sage publications, Thousand Oaks, CA.

Department of Homeland Security (2022), "Update to the department of homeland security STEM designated degree program list [notice]", available at: www.federalregister.gov/d/2022-01188

Engel, J. (2017), "Statistical literacy for active citizenship: a call for data science education", *Statistics Education Research Journal*, Vol. 16 No. 1, pp. 44-49.

English, L.D. (2017), "Advancing elementary and Middle school STEM education", *International Journal of Science and Mathematics Education*, Vol. 15 No. S1, pp. 5-24, doi: 10.1007/s10763-017-9802-x.

Franco, M.S. and Patel, N.H. (2017), "Exploring student engagement in STEM education: an examination of STEM schools, STEM programs, and traditional schools", *Research in the Schools*, Vol. 24 No. 1, pp. 10-30.

Herro, D., Quigley, C. and Cian, H. (2018), "The challenges of STEAM instruction: lessons from the field", *Action in Teacher Education*, Vol. 41 No. 2, pp. 172-190, doi: 10.1080/01626620.2018.1551159.

Herro, D., Madison, M., Irgens, G.A., Hirsch, S., Abimbade, O. and Adisa, O. (2022), "Exploring elementary teachers' perceptions of data science and curriculum co-design through professional development", *Journal of Technology and Teacher Education*, Vol. 30 No. 4, pp. 493-525, available at: www.learntechlib.org/primary/p/220700/

Ito, M., Gutiérrez, K., Livingstone, S., Penuel, B., Rhodes, J., Salen, K., Schor, J., Sefton-Green, J. and Watkins, S.C. (2013), "Connected learning: an agenda for research and design", *Digital Media and Learning Research Hub*, available at: http://dmlhub.net/

Kim, Y.M. (2021), "The effects of PBL-based data science education program using app inventor on elementary students' computational thinking and creativity improvement", *Ilköğretim Online*, Vol. 20 No. 1, pp. 551-562, doi: 10.17051/ilkonline.2021.01.124.

Kjelvik, M.K. and 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*, Vol. 18 No. 2, p. es2, doi: 10.1187/cbe.18-02-0023.

Kumpulainen, K. and Sefton-Green, J. (2014), "What is connected learning and how to research it?", *International Journal of Learning and Media*, Vol. 4 No. 2, pp. 7-18.

LaMar, T. and Boaler, J. (2021), "The importance and emergence of K-12 data science", *Phi Delta Kappan*, Vol. 103 No. 1, pp. 49-53, doi: 10.1177/00317217211043627.

Lee, V.R., Wilkerson, M.H. and Lanouette, K. (2021), "A call for a humanistic stance toward K–12 data science education", *Educational Researcher*, Vol. 50 No. 9, pp. 664-672, doi: 10.3102/0013189X211048810.

Liston, M., Morrin, A.M., Furlong, T. and Griffin, L. (2022), "Integrating data science and the internet of things into science, technology, engineering, arts, and mathematics education through the use of new and emerging technologies", *Frontiers in Education*, Vol. 7, p. 757866, doi: 10.3389/feduc.2022.757866.

Martinez, W. and LaLonde, D. (2020), "Data science for everyone starts in kindergarten: strategies and initiatives from the American statistical association", *Harvard Data Science Review*, Vol. 2 No. 3, doi: 10.1162/99608f92.7a9f2f4d.

National Academies of Sciences, Engineering, and Medicine (NASEM) (2023), *Foundations of Data Science for Students in Grades K-12: Proceedings of a Workshop*, National Academies Press, doi: 10.17226/26852.

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, Council of Chief State School Officers.

NGSS Lead States (2013), *Next Generation Science Standards: For States, by States*, National Academies Press, available at: https://nap.nationalacademies.org/catalog/18290/next-generation-science-standards-for-states-by-states

Nickerson, R.S. (1998), "Confirmation bias: a ubiquitous phenomenon in many guises", *Review of General Psychology*, Vol. 2 No. 2, pp. 175-220.

Ow-Yeong, Y.K., Yeter, I.H. and Ali, F. (2023), "Learning data science in elementary school mathematics: a comparative curriculum analysis", *International Journal of STEM Education*, Vol. 10 No. 1, p. 8, doi: 10.1186/s40594-023-00397-9.

Pangrazio, L. and Selwyn, N. (2019), "Personal data literacies': a critical literacies approach to enhancing understandings of personal digital data", *New Media and Society*, Vol. 21 No. 2, pp. 419-437, doi: 10.1177/1461444818799523.

Pangrazio, L. and Sefton-Green, J. (2020), "The social utility of 'data literacy'", *Learning, Media and Technology*, Vol. 45 No. 2, pp, 208-220, doi: 10.1080/17439884.2020.1707223.

Patton, M.Q. (2002), *Qualitative Research and Evaluation Methods*, 3rd ed., Sage.

Quigley, C. and Herro, D. (2016), "Finding the joy in the unknown: implementation of STEAM teaching practices in middle school science and math classrooms", *Journal of Science Education and Technology (JOST)*, Vol. 25 No. 3, pp. 410-426, doi: 10.1007/s10956-016-9602-z.

Reynante, B.M., Selbach-Allen, M.E. and Pimentel, D.R. (2020), "Exploring the promises and perils of integrated STEM through disciplinary practices and epistemologies", *Science and Education*, Vol. 29 No. 4, pp. 785-803, doi: 10.1007/s11191-020-00121-x.

Roehrig, G.H., Dare, E.A., Ellis, J.A. and Ring-Whalen, E. (2021), "Beyond the basics: a detailed conceptual framework of integrated STEM", *Disciplinary and Interdisciplinary Science Education Research*, Vol. 3 No. 1, p. 11, doi: 10.1186/s43031-021-00041-y.

Santana, O.A., Braga, G.D., Braga, J.O.D. and Carvalho, H. (2020), "Interactive model tool about center of mass during COVID-19 pandemic: a new learning path in STEM for K-12 education", *2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, pp. 503-508, doi: 10.1109/TALE48869.2020.9368486.

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H. and Jinks, C. (2018), "Saturation in qualitative research: exploring its conceptualization and operationalization", *Quality and Quantity*, Vol. 52 No. 4, pp. 1893-1907, doi: 10.1007/s11135-017-0574-8.

Sava, D. (2020), "COVID-19 forecasting model", *Google Cloud Blog*, available at: https://cloud.google.com/blog/products/ai-machine-learning/google-cloud-is-releasing-the-covid-19-public-forecasts

Showalter, D., Hartman, S.L., Johnson, J. and Klein, B. (2019), "Why rural matters 2018-2019: the time is now", A Report of the Rural School and Community Trust. Rural School and Community Trust, available at: https://files.eric.ed.gov/fulltext/ED604580.pdf

Shute, V.J., Sun, C. and Asbell-Clarke, J. (2017), "Demystifying computational thinking", *Educational Research Review*, Vol. 22, pp. 142-158.

Siverling, E., Suazo-Flores, E. and Moore, T. (2018), "STEM content in elementary school students' evidence-based reasoning discussions (fundamental)", *2018 ASEE Annual Conference and Exposition Proceedings*, p. 30986.

Slater, T.F. and Sanchez, R.L. (2021), "Evaluating K-16 student engagement in STEM-based drone racing", *Journal of Astronomy and Earth Sciences Education (JAESE)*, Vol. 8 No. 2, pp. 81-90, doi: 10.19030/jaese.v8i2.10405.

Stanford Graduate School of Education (2023), "Data science | K-10 BIG IDEAS", *Youcubed*, available at: www.youcubed.org/data-big-ideas/

Sulaeman, N.F., Putra, P.D.A., Mineta, I., Hakamada, H., Takahashi, M., Ide, Y. and Kumano, Y. (2021), "Exploring student engagement in STEM education through the engineering design process", *Jurnal Penelitian Dan Pembelajaran IPA*, Vol. 7 No. 1, pp. 1-16.

Tranquillo, J. and Matthew, V. (2015), "So you want to offer a pop-up class?", *Epicenter*, available at: http://epicenter.stanford.edu/resource/so-you-want-to-offer-a-pop-up-class.html

Vygotsky, L. (1973), *Thought and Language*, in Kozulin, A. (Trans. and Ed.), Harvard University Press, (Original Work Published 1934).

Weaver, D.J. (2022), "Delivering on the promise of digital equity", Digital Promise, doi: 10.51388/20.500.12265/166.

Weiland, T. and Engledowl, C. (2022), "Transforming curriculum and building capacity in K–12 data science education", *Harvard Data Science Review*, Vol. 4 No. 4, doi: 10.1162/99608f92.7fea779a.

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L. and Wilensky, U. (2016), "Defining computational thinking for mathematics and science classrooms", *Journal of Science Education and Technology*, Vol. 25 No. 1, pp. 127-147, doi: 10.1007/s10956-015-9581-5.

Wing, J.M. (2006), "Computational thinking", *Communications of the ACM*, Vol. 49 No. 3, pp. 33-35, doi: 10.1145/1118178.1118215.

Wu, S.P.W. and Rau, M.A. (2019), "How students learn content in science, technology, engineering, and mathematics (STEM) through drawing activities", *Educational Psychology Review*, Vol. 31 No. 1, pp. 87-120, doi: 10.1007/s10648-019-09467-3.

Corresponding author
Ibrahim Oluwajoba Adisa can be contacted at: iadisa@g.clemson.edu

Appendix

Inside the Classroom Observation/Analytic Protocol of CT-STEM Pop-Up

Researcher/Observer: ___________________________ Date of Observation: ____________
List of Other Teachers in the Room: _______________
Time: Start: _________ End: _________
School: ______________________________________ District: ________________________

Basic Descriptive Information

1. Name of lesson/activity: ____
2. Grade Level(s): ____
3. Students: Number of females: ______________ Number of males: _______________

Purpose of the Lesson:

Indicate how the purpose of the lesson relates to data science and provide the teacher's stated purpose of the lesson.

Classroom, Lesson Arrangements and Activities

In question 1 of this section, please divide the total duration of the lesson into instructional and non-instructional time. In question 2 make your estimates based only on the instructional time of the lesson.

1. Approximately how many minutes during the lesson were spent:

(a) On Instruction: _________________ (b) On Activities: ________________________

(c) On housekeeping unrelated to the lesson/interruptions/other non-instructional activities: __________

(d) Describe any major interruptions (e.g., fire drill, shortened class period, assembly): ______________

2. How was the classroom environment arranged?

Describe:

(*conituned*)

3. Considering the instructional time of the lesson (listed in 1a above), approximately what percent was spent in each of the following arrangements?

a. Whole class: ________________________
b. Pairs/small groups: ___________________
c. Individual: __________________________

4. Narrative

Describe what happened in this lesson (What materials and resources were used? Type of data science activities students were involved in; how did the teacher organize the lesson in relation to data science? How did the teacher maintain students' interest during the lesson?).

5. Lesson alignment to CT-STEM

Check alignment to data practices and computational problem-solving practices noted during observation.

Data Practices	☐	Computational problem-solving practices	☐
Analyzing Data	☐	Preparing problems for computational solution	☐
Collecting Data	☐	Programming	☐
Creating Data	☐	Choosing effective computational tools	☐
Manipulating Data	☐	Assessing different approaches/solutions to a problem	☐
Visualizing Data	☐	Developing modular computational solutions	☐
Troubleshooting and debugging	☐	Creating computational abstractions	☐

(*conituned*)

6. Assessment

N/A ☐

(How were the learners assessed? What type of formative and summative assessments were used? Was the evidence from assessments used to drive instructional strategies?)

Noted successes and challenges:

Strengths/successes:

Challenges:

Note if any curriculum artifacts were collected or images taken with naming convention.